Universiteit Utrecht

Centrum Wiskunde & Informatica

Graduate School of Natural Sciences

Analysis of rare events for AC power flow models

MASTER THESIS

Bas Verseveldt
Mathematical Sciences

Supervisors:
Prof.dr. A. P. Zwart, CWI/Eindhoven University of Technology
Prof.dr. S. M. Verduyn Lunel, Utrecht University

August 2022

Contents

1 Introduction

The electrical grid is essential for the functioning of a modern society, and covers a broad array of engineering, physical and mathematical aspects[1]. Electricity is transported on power lines with a certain capacity. Exceeding this capacity on a line can lead to the line breaking down, or being shut down by the grid operator in advance over safety concerns. In both cases the line is effectively (temporarily) removed from the grid. Without this line, the power then has to take a different route, changing the power flow on the remaining lines in the grid. Consequentially, some other lines that were previously safe may now suddenly be overloaded, resulting in these lines being removed well. As soon as power cannot be supplied to all customers any more, we have a *blackout*, with a part of the grid or possibly even the whole grid going down[2]. The 2003 North American blackout was one of the most widespread blackouts in recent history and as such is often used as an example in studies[3]. Blackouts can be very damaging to a country and its economy, and also negatively impact the life of civilians, and therefore it is of great importance to prevent blackouts. A grid operator is tasked with providing all consumers with the desired power, while limiting the risk of a blackout as much as possible. The difficulty is in the fact that power demand is not constant, and therefore neither is the power flow. A network can be safe at one moment, but it may not be safe anymore when a consumer increases or decreases the power demand. This is made more difficult by the transition to renewable energy that depends on weather conditions. As a result power grid operation is currently an important research topic[4]. Another factor for a grid operator is the *robustness* of the network: it is desirable that a single line failure does not immediately lead to a blackout. In this thesis we will only focus on preventing overloads, and not on how overloads can cascade into blackouts.

It is important to note that since the capacity of the network will always be limited, it is not possible to make overloads impossible while also allowing consumers complete freedom to use as much power as desired. This means that if we do not constrain consumers, the probability of an overload is always non-zero[1]. Conversely, if we can make an estimate of the power demand, then a grid operator will certainly design a network in such a way that an overload will not occur under the expected power demand, and make sure that there is a significant margin for perturbations. This means that the probability of an overload will be low, but still non-zero. This way, an overload becomes a *rare event*. Rare events is a research topic concerning events are highly unlikely, but do have an enormous impact if they do happen. This field of research is not limited to power grids, but also concerns events such as extreme weather, heavy earthquakes, or stock market crashes[5]. Since we are dealing with very small probabilities, observing how often such an event happens (or, in case of a computer simulation, brute-force sampling) will likely not be able to give us a proper estimate of the probability. Instead we will be using several methods in *rare event analysis*. This way, the grid operator can decide if the chance of an overload is acceptably low given the network and the behaviour[2] of its customers. We will also consider the limiting case in which the typical deviation from the expected power demand becomes extremely small,

[1]Since this thesis focuses on the mathematical analysis on a given network, we will ignore any external factors that can also lead to an overload, such as a power line being destroyed.

[2]In this context, 'behaviour' refers to the stochastic power demand.

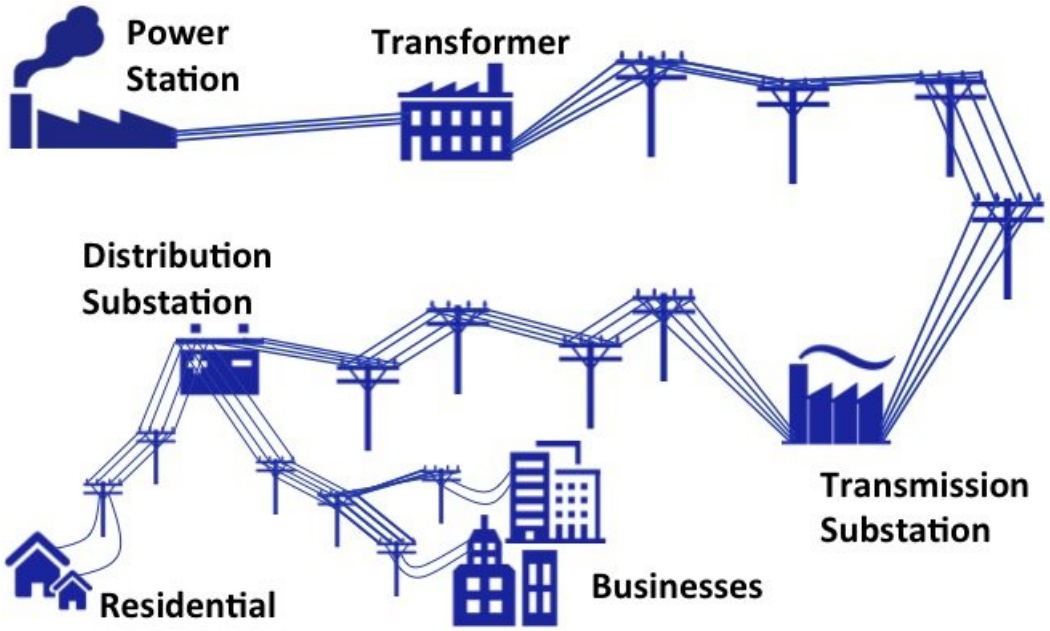

Figure 1: The structure of a power grid, simplified. For image credit, see [6].

and hence so does the probability of an overload. To evaluate how the overload probability evolves in this case we will be using *large deviations theory*.

1.1 Grid structure

A power grid generally consists of three different levels. Transportation over large distances is done via the *transmission grid*. This grid is used to transport power from generators to large substations. To minimize losses, voltages are kept high (>100 kV) in order to limit losses due to resistance. The transmission grid spans a very large area, so as to connect regions with lots of power generators with regions with lots of power demand. The second level is the *distribution grid*, which covers a smaller geographic area (like a city or neighborhood). The voltages here are lower than in the transmission grid (about 10 kV). Finally, the power is delivered to houses and businesses. A schematic view of a power grid is shown in Figure 1.

1.2 Power-flow study

In an electricity grid, the field of power flow study concerns the analysis of the flow of electric power. The most frequently used model for a power grid is the *alternating current power-flow model*, which will be explained in section 3.1. It consists of a set of non-linear equations that together describe the power flow through each transmission line. The line losses due to resistance, as well as the amount of reactive power in lines, are also included. Since power networks are usually too complex for solving the power flow equations analytically, this is usually done numerically. An analytic solution is however sometimes possible for either a

very simple model, or with making various simplifying assumptions (such as assuming that there is no resistance in the lines).

Since we are studying the mathematical aspect of power networks, we view a power network as a graph $G = (V, E)$, with vertices (most often called *buses* or *nodes*) V and edges (usually called *lines*) E. We assume that we have complete information about the network. That is, we know where all the lines are, we know the resistance and capacity of all the lines, and also know the total power input at every node. This is not all that natural in practice: for example, lines may be faulty and therefore have a different resistance or capacity, or a company may be consuming more electricity than documented, but we ignore such more social problems in this thesis. With all information, we are tasked to compute the line power flows.

1.3 Thesis overview

We will first be discussing large deviations theory in section 2. Next, in section 3.1 we will derive the AC power flow equations and solve them exactly for a two-node model. We will also explore the DC approximation, which is a list of often made simplifications that are used to solve the power flow equations exactly for larger networks. In section 4, we will be looking at a grid shaped as a three-cycle, and solve it analytically without making the assumption of small angles, and compare the results to the solution of the DC approximation. Next, in section 5 we discuss a line graph, which represents a street of houses that gets its power from a single source, and find the most likely realization of an overload. In section 6 we see how much less (or more) likely an overload is if we make an additional link to transform our line into a circle. We will look at approximations assuming small perturbations, an upper and lower bound, and compare the bounds with the actual numerical results. We also present a semi-definite program that can be used to solve the model exactly if the bounds are not considered accurate enough.

2 Rare events analysis and large deviations theory

If we are studying events (such as an overload), the field of probability theory concerns finding a probability between 0 and 1 for such an event. If we are lucky, such a probability can be computed exactly, ideally as a function of several parameters (such as line limits, behaviour of customers, etc.). In practice, however, power networks are usually too complicated to make such computations exactly. For this reason, we try to estimate the probability. A method often used to estimate probabilities for complicated models is *brute-force sampling*. For example, if we run a simulation 10^4 times and count the number of times an event occurs, we can estimate the probability based on this. However, this approach fails when concerning events that are extremely infrequent (or at least supposed to be), such as overloads. The field of *rare event analysis* concerns estimating such probabilities.

2.1 Importance of rare events

Rare events do not just occur in the field of electricity grids, and can be caused by both anthropogenic and natural factors (or a combination of both). A few other examples of rare events are:

- Major 8.0+ Richter magnitude earthquakes (natural)
- A nuclear accident resulting in a major disaster that leads to an area becoming uninhabitable for thousands of years (anthropogenic)
- Dam bursts caused by extreme rainfall (natural and anthropogenic)

The metric to analyze the danger of a rare event is often computed as the frequency of an event multiplied with its damage. Although these rare events are (as the same states) uncommon, they do often have a massive impact. For this reason, they are still important to analyze. One branch of rare events analysis is estimating the probability of such an event, which is difficult since they do not occur frequently. Another branch is assessing the impact of such an event. This is probably even more difficult, since the last such event may have been decades or even centuries ago, meaning that any data on the cost of such a disaster is likely to be obsolete and hence rather useless. In this paper, we will only be looking at the probability of overloads happening in several grid models. The analysis of the impact of such an overload is a totally different problem that cannot be solved with the same methods used in this paper, which is why they are not analyzed here. Research on this topic can be found in [8].

2.2 Large deviations theory

Due to the infrequency of rare events, it is difficult to estimate their frequency. An important tool for doing this is the *rate function*.

Definition 2.1 (Rate function) *Let X be a Hausdorff topological space. An extended real-valued function $I : X \to [0, +\infty]$ is a rate function if it is not identically $+\infty$ and lower semi-continuous, meaning that the set $I^{-1}([0, c])$ is closed for any $c \geq 0$. If this set is compact for any finite c, then I is a good rate function.*

In this research project, X is always a subset of $\mathbb{R}^n$ for some n, usually the solution of a set of polynomial equations, with the induced topology. This guarantees it to be Hausdorff. The next important definition is that of a large deviations principle.

Definition 2.2 (Large deviations principle) *Let X be a Hausdorff topological space. A family of probability measures $\{\mathbb{P}_\epsilon\}_{\epsilon>0}$ on X satisfies a large deviations principle with rate function $I : X \to [0, +\infty]$ if for any Borel measurable set $E \subseteq X$ we have*

$$\limsup_{\epsilon \to 0} \epsilon \log(\mathbb{P}_\epsilon(E)) \leq - \inf_{x \in \overline{E}} I(x)$$
$$\liminf_{\epsilon \to 0} \epsilon \log(\mathbb{P}_\epsilon(E)) \geq - \inf_{x \in \mathring{E}} I(x).$$

*Here, $\overline{E}$ and $\mathring{E}$ denote the closure and interior of E, respectively. The set E is often referred to as the **event** that we are analyzing.*

In practice, the function I is often not just lower semi-continuous, but actually continuous. In this case, we know from real analysis that

$$\inf_{x \in \overline{E}} I(x) = \inf_{x \in \mathring{E}} I(x) \tag{1}$$

and hence we find the much simpler relation

$$\lim_{\epsilon \to 0} \epsilon \log(\mathbb{P}_\epsilon(E)) = - \inf_{x \in E} I(x). \tag{2}$$

Note that this means that a small $I(x)$ corresponds to a higher probability density at x. If we assume to be in the limiting case in which ϵ is very small, then we obtain the relation

$$\epsilon \log(\mathbb{P}_\epsilon(E)) \simeq - \inf_{x \in E} I(x) \tag{3}$$

where $\simeq$ denotes asymptotic equality. This implies that

$$\mathbb{P}_\epsilon(E) \approx \exp\big[- \inf_{x \in E} I(x)/\epsilon \big]. \tag{4}$$

In this thesis, we are working with a power injection that has a mean μ and a noise with variance proportional to ϵ: the probability measures generated by this will be our $\mathbb{P}_\epsilon$. We will later see that these measures indeed satisfy a large deviations principle. If we wish to analyse the probability of an overload, we let E be the event of an overload. The set $E \subset X$ corresponds to all configurations that lead to an overload. Now there are two possible situations:

1. If $\inf_{x\in E} I(x) = 0$ then we find that $\mathbb{P}_\epsilon(E) \simeq 1$. In practice, this happens if the mean power injection μ already leads to an overload. So in that case an overload would not be a rare event. Hence, this case is usually excluded. From a practical viewpoint, it is obvious that no grid operator will allow the average power injection to lead to an overload, because overloads need to be avoided.

2. If $\inf_{x\in E} I(x) = c > 0$, then we find that $\mathbb{P}_\epsilon(E) \simeq e^{-c/\epsilon}$. This happens if μ does not lead to an overload. In this case, we obtain a direct relation between the variance ϵ of the noise and the probability of an overload. If the variance ϵ is replaced with ϵ/a, then the probability of an overload is replaced by $e^{-ac/\epsilon} = (e^{-c/\epsilon})^a$.

With all this we can see why the rate function is such a fundamental concept in large deviations theory, because it allows us to compute the probability of a rare event $\mathbb{P}_\epsilon(E)$ in terms of ϵ. The number $\inf_{x\in E} I(x)$ is known as the *decay rate* and the subset $\arg\inf_{x\in E} I(x)$ of E is the *most likely realization*. Note that this implies that in the limit $\epsilon \to 0$, only the most likely realization (i.e. the $x \in \overline{E}$ with the lowest value $I(x)$) is important for determining $\mathbb{P}_\epsilon(E)$. This means that if we consider E to be an overload event, then for small ϵ the overload will almost certainly happen at a point in or close to $\arg\inf_{x\in E} I(x)$

2.3 The contraction principle

If we have a transformation from one space to another, we may wonder how a large deviations principle evolves under this transformation. This can for example be used when we have a large deviation principle for power inputs, and wish to obtain a large deviations principle for power flows or voltages. This allows us to estimate the probability of a line overflow if we only have information about the power inputs.

Theorem 2.1 (Contraction principle) *Let X and Y be Hausdorff topological spaces and let $(\mu_\epsilon)_{\epsilon>0}$ be a family of probability measures on X that satisfies the large deviations principle with rate function $I : X \to [0,\infty]$. Let $T : X \to Y$ be a continuous function, and let ν_ϵ be the push-forward measure of μ_ϵ by T, defined as $\nu_\epsilon(E) = \mu_\epsilon(T^{-1}(E))$ for every event $E \subseteq Y$. Let*

$$J(y) = \inf\{I(x) \mid x \in X \text{ and } T(x) = y\} \tag{5}$$

with the convention that $\inf(\emptyset) = \infty$. Then:

- *$J : Y \to [0,\infty]$ is a rate function on Y*
- *J is a good rate function on Y if I is a good rate function on X*
- *$(\nu_\epsilon)_{\epsilon>0}$ satisfies the large deviation principle on Y with rate function J.*

For a proof and further reading on the contraction principle and related concepts, see [9] or [10]. Effectively, the contraction principle states that a rate function on the space Y can

be defined pointwise for any $y \in Y$. Equation (5) can intuitively be understood as follows. Recall that a low rate function corresponds to a higher probability density, and that in the limit $\epsilon \to 0$ the most likely realization of an event becomes dominant. So it roughly states that in this limit the likelihood of a value y is determined by[3] the most likely value x that T maps on y, i.e. the most likely realization of the point y.

[3]Note that I avoid writing 'equal to', since that is unfortunately not necessarily the case. For instance, if the probability density of the point y is twice as high as that of x, then the rate functions can still be the same, since the factor 2 drops out in equation (2) when taking the limit.

3 The AC power flow equations and their simplifications

A power grid can be seen as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ is the set of nodes (also referred to as *buses*), and $\mathcal{E}$ is the set of edges (also referred to as lines). We assume $\mathcal{G}$ to be connected, otherwise we can just consider its connected components. Also, since power can flow in any direction, the graph $\mathcal{G}$ is undirected. The line between k and ℓ is hence denoted as $\{k, \ell\}$. If their orientation is relevant for a specific situation, then we denote (k, ℓ) for the line from the perspective of k. For the line (k, ℓ), we denote the power flow from node k to node ℓ by $S_{k\ell}$. It is important to remember that in general $S_{k\ell} \neq -S_{\ell k}$ since there can be power losses in the line due to resistance. Let $n = |\mathcal{N}|$ and $m = \mathcal{E}$ be the number of buses and lines.

3.1 The AC model

Power is most often generated and transmitted using Alternating Current (AC), of which we will give a description in this section, closely following [7]. Let j be the imaginary unit[4]. Let $S_k(t)$ be the power injected at node k at time t. Similarly $S_{k\ell}$ denotes the power flow from node k to ℓ. Write $S_k(t) = P_k(t) + jQ_k(t)$, with $P_k(t)$ the active power injected, and $Q_k(t)$ the reactive power injected. If node k consumes more power than it produces, then $P_k(t)$ is negative. We similarly define $S_{k\ell}(t) = P_{k\ell}(t) + jQ_{k\ell}(t)$.

If $\{k, \ell\} \in \mathcal{E}$, then the complex current injected by node k into line (k, ℓ) is defined as $i_{k\ell}(t)$, at voltage $v_k(t)$. Since we are dealing with alternating current, the voltage leads the current with a phase δ_k, so that we have $v_k(t) = |v_k(t)|e^{j\delta_k}$. Here $|v_k(t)|$ is the voltage magnitude, and δ_k is the voltage phase angle. Usually δ_k is chosen to be in $(-\pi, \pi]$ by convention. To save notation, we no longer write the time dependence of power, current and voltage from now on.

The AC power flow equations follow from two basic equations: Ohm's law and Kirchoff's law. Ohm's law states that

$$z_{k\ell} i_{k\ell} = v_k - v_\ell \tag{6}$$

which means that power flow is proportional to voltage drop and the line impedance $z_{k\ell}$. The line impedance is a complex generalization of resistance $r_{k\ell}$ and the two are related by

$$z_{k\ell} = r_{k\ell} + jx_{k\ell} = \frac{1}{y_{k\ell}} = \frac{1}{g_{k\ell} + jb_{k\ell}}. \tag{7}$$

Here, $r_{k\ell}, x_{k\ell}, y_{k\ell}, g_{k\ell}, b_{k\ell}$ are respectively called *resistance, reactance, admittance, conductance, susceptance*. Since we will not be using these a lot in this thesis a detailed explanation or physical interpretation is omitted. It is best to just treat (7) as a definition, and not think too much about the true meaning of these numbers. If you are interested in learning more, see for example [11].

[4]The letter i is used for current, and hence j is used to denote the complex unit.

Kirchoff's law states that there is power balance at every node:

$$i_k = \sum_{\ell=1}^{n} i_{k\ell}. \tag{8}$$

That is, the current injection at bus k is equal to the sum of currents flowing out of bus k. For combining Ohm's law with Kirchoff's law, we would like to substitute $i_{k\ell}$ using equation (9). For this, we note that (9) is equivalent to

$$i_{k\ell} = y_{k\ell}(v_k - v_\ell) \tag{9}$$

and now we can define the following matrix

$$Y_{k\ell} = \begin{cases} -y_{k\ell} & \text{if } k \neq \ell \\ \sum_{l=1}^{n} y_{kl} & \text{if } k = \ell \end{cases}. \tag{10}$$

If $\{k, \ell\} \notin \mathcal{E}$, then we take $y_{k\ell} = 0$. Now, we get that

$$i_k = \sum_{\ell=1}^{n} i_{k\ell} = \sum_{\ell=1}^{n} y_{k\ell}(v_k - v_\ell) = \sum_{\ell=1}^{n} Y_{k\ell} v_\ell. \tag{11}$$

We can write this in matrix form

$$\mathbf{i} = \mathbf{Y}\mathbf{v} \tag{12}$$

where $\mathbf{i} = (i_1, \ldots, i_n)^T$ and $\mathbf{v} = (v_1, \ldots, v_n)^T$ are real vectors. Now the power injected by bus k into the line (k, ℓ) is known to be equal to

$$S_{k\ell} = v_k i_{k\ell}^\star \tag{13}$$

where the star denotes complex conjugation. This implies that

$$\begin{aligned} S_k &= v_k i_k^\star \\ &= v_k \sum_{\ell=1}^{n} Y_{k\ell}^\star v_\ell^\star \\ &= \sum_{\ell=1}^{n} |v_k||v_\ell| Y_{k\ell}^\star e^{j(\delta_k - \delta_\ell)}. \end{aligned}$$

We can split $\mathbf{Y}$ into a real and complex part by writing $\mathbf{Y} = \mathbf{G} + j\mathbf{B}$. Then, we obtain the AC equations

$$P_k = \sum_{\ell=1}^{n} |v_k||v_\ell|(B_{k\ell}\sin(\delta_k - \delta_\ell) + G_{k\ell}\cos(\delta_k - \delta_\ell)) \tag{14}$$

$$Q_k = \sum_{\ell=1}^{n} |v_k||v_\ell|(-B_{k\ell}\cos(\delta_k - \delta_\ell) + G_{k\ell}\sin(\delta_k - \delta_\ell)). \tag{15}$$

From this, we can derive the AC power flow equations

$$P_{k\ell} = -|v_k|^2 G_{k\ell} + |v_k||v_\ell|(B_{k\ell}\sin(\delta_k - \delta_\ell) + G_{k\ell}\cos(\delta_k - \delta_\ell)) \tag{16}$$

$$Q_{k\ell} = |v_k|^2 B_{k\ell} + |v_k||v_\ell|(-B_{k\ell}\cos(\delta_k - \delta_\ell) + G_{k\ell}\sin(\delta_k - \delta_\ell)). \tag{17}$$

Now the classical AC power flow problem is computing the line power flows in equation (16) and (17) given the power injections and voltages at every node as well as the grid constants $B_{k\ell}$ and $G_{k\ell}$. However, the AC power flow equations are non-linear and except for very simple models it is usually not possible to solve them exactly. It seems logical to apply numerical analysis instead, but due to the complexity of the equations this may take a lot of time. This is a problem, since a grid operator usually only has limited time to solve the power flow problem.

3.2 The DC approximation

Since the AC power flow equations are so difficult to solve, the equations are often simplified. For this, there are four assumptions that are often made:

1. We only consider active power to describe power flow, ignoring reactive power. This means that equations (15) and (17) are simply removed.
2. The resistance of a transmission line is significantly less than the reactance. This means that $r_{k\ell} \ll x_{k\ell}$. Hence $g_{k\ell} \approx 0$ and $b_{k\ell} \approx -1/x_{k\ell}$. We can now define the weight of an edge to be $w_{k\ell} = 1/x_{k\ell}$, with the convention that $w_{k\ell} = 0$ if $\{k, \ell\} \notin \mathcal{E}$.
3. The voltage magnitude $|v_k|$ is almost the same for all nodes, i.e. $|v_k| \approx v$ for all k. By using a per-unit system, we may just as well set v to 1. This means that $|v_k| \approx 1$ for all k.
4. We assume the voltage angle differences to be small. This means that $\sin(\delta_k - \delta_\ell) \approx \delta_k - \delta_\ell$ and $\cos(\delta_k - \delta_\ell) \approx 1$.

This set of assumptions is called the *DC approximation*. If we make all these assumptions, the AC equations become linear, and can be solved exactly for any network. An analysis of this can be found in [7]. This approximation is frequently used due to its simplicity. What we may be interested in is what happens if we make some of these assumptions, but not all. This is certainly important, as not all assumptions will always be valid. It also allows us to make a more accurate approximation of the actual solution of the AC power flow equations. In this thesis, we will be looking at several models that make some, but not all, assumptions of the DC approximation.

3.3 The two-node model

The simplest model (and one of the few that can be solved exactly without making any of the DC assumptions) is the two-node model with one line between the two nodes. Hence we have $n = 2$ and $m = 1$. Although rather trivial, this model is useful to demonstrate some properties of an AC network.

The power inputs are known, but since there is only one line the summations have only one term:

$$P_1 = |v_1||v_2|(B_{12}\sin(\delta_1 - \delta_2) + G_{12}\cos(\delta_1 - \delta_2))$$
$$P_2 = |v_1||v_2|(-B_{12}\sin(\delta_1 - \delta_2) + G_{12}\cos(\delta_1 - \delta_2)).$$

The power flow is hence equal to

$$P_{12} = P_1 - |v_1|^2 G_{12} = |v_1||v_2|(B_{12}\sin(\delta_1 - \delta_2) + G_{12}\cos(\delta_1 - \delta_2)) - |v_1|^2 G_{12}$$
$$P_{21} = P_2 - |v_2|^2 G_{12} = |v_1||v_2|(-B_{12}\sin(\delta_1 - \delta_2) + G_{12}\cos(\delta_1 - \delta_2)) - |v_2|^2 G_{12}.$$

This essentially solves the power flow equations for this case, but let's look at some more results. Note that

$$\begin{aligned} P_{12} + P_{21} &= 2G_{12}|v_1||v_2|\cos(\delta_1 - \delta_2)) - (|v_1|^2 + |v_2|^2)G_{12} \\ &= G_{12}\big(2|v_1||v_2|\cos(\delta_1 - \delta_2) - |v_1|^2 - |v_2|^2\big) \end{aligned}$$

which is generally non-zero. So let's see if we can find a case when it is zero. We have the following series of inequalities:

$$\begin{aligned} &(|v_1| - |v_2|)^2 \geq 0 \\ \iff &|v_1|^2 - 2|v_1||v_2| + |v_2|^2 \geq 0 \\ \iff &2|v_1||v_2| - |v_1|^2 - |v_2|^2 \leq 0 \\ \implies &2|v_1||v_2|\cos(\delta_1 - \delta_2) - |v_1|^2 - |v_2|^2 \leq 0 \end{aligned}$$

where the last inequality holds since $|v_1||v_2| > 0$ (since we are taking absolute values) and $\cos(\delta_1 - \delta_2) \leq 1$. Note that we have equality in the last equation only if we both have $\cos(\delta_1 - \delta_2) = 1$ and $|v_1| = |v_2|$.

Since $G_{12} \geq 0$ we conclude that $P_{12} + P_{21} \leq 0$, which effectively means that we can only lose power due to resistance, and never gain it (which makes sense). A next question is when it occurs that $P_{12} + P_{21} = 0$. One possibility is that $G_{12} = 0$, which means that $R_{12} = 0$ and hence there is no resistance. The other possibility is that $\cos(\delta_1 - \delta_2) = 1$ which means $\delta_1 = \delta_2$ (i.e. there is no phase difference between both nodes) and $|v_1| = |v_2|$. This implies $P_{12} = P_{21} = 0$, so there is no power flow, which explains why there is no power loss in this case.

We can also compute

$$\cos(\delta_1 - \delta_2) = \frac{P_1 + P_2}{2|v_1||v_2|G_{12}}$$
$$\sin(\delta_1 - \delta_2) = \frac{P_1 - P_2}{2|v_1||v_2|B_{12}}.$$

These two equations combined allow us to compute $\delta_1 - \delta_2$. Note that only the phase angle difference is relevant, so we always have one degree of freedom left for determining

the phase angles. It is possible to assume that $\delta_1 = 0$ and then compute δ_2. Note that if P_1 and P_2 are chosen, then there is only one possible value for $|v_1||v_2|$, since we must have $\cos(\delta_1 - \delta_2)^2 + \sin(\delta_1 - \delta_2)^2 = 1$. It makes sense that there is such a requirement: since there is resistance in the system, there must be a certain relation between power injection and voltages so as to match the resistance. Also note that if the voltage increases by a factor c then the power increases by a factor c^2, which matches the relation $P = V^2/R$ that we know from grids with direct current.

4 Analysis of a system without small angles

In this section, we will be analyzing the system in which we make all of the DC approximations, except the small angle approximation. One reason that this model is relevant is because we may expect it to show more dynamics than the full DC approximation, as the AC power flow equations without this assumption are not linear or even polynomial (in their initial form). At first glance it may seem very difficult to work with due to all the geometric functions. It turns out that the problem can actually be reduced to a polynomial equation in a single variable. Although the degree of this polynomial is usually much higher than 4 (and hence the solutions cannot be computed exactly), we can actually approximate the solutions very well with the concept of *near-solutions*, which are simple expressions that are very close to the actual solutions (and in the limit of deviations δ going to zero, they become exact).

For the most part, we will be investigating the 3-cycle, the simplest non-trivial model. We will first solve it by full linearization, and then without assuming small angles. We then move on to approximate the probability of an overflow in the system, assuming given line limits.

4.1 Solving the three-node cycle

As a first non-trivial model, we take the 3-cycle, so that we have $m = n = 3$. We also assume that all lines are identical for simplicity. See Figure 2. Although not very realistic with all these assumptions, it is still highly useful as an example.

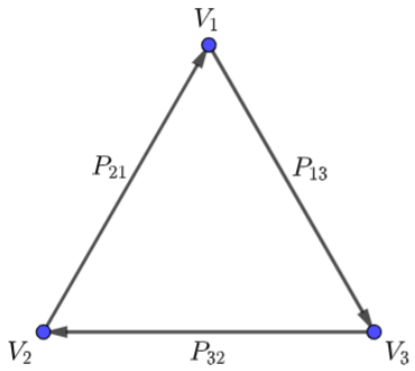

Figure 2: The three-cycle model shown in a diagram. We assume that the lines have no resistance and have identical reactance. The vertices are labeled V_1, V_2, V_3. At node V_i the voltage phase angle is δ_i and the power injection is P_i. The power injections are known, and due to power balance we must have $P_1 + P_2 + P_3 = 0$. The problem we try to solve is computing the power flows P_{21}, P_{13}, P_{32}.

One reason that this model is interesting is because it is the simplest non-tree model. Since the AC power flow problem is simpler for tree graphs, most research focuses on these instead.

We will first simplify the power flow equations with the DC approximations that we do still make. Then we will solve them with the additional assumption that $\sin(\delta_k - \delta_\ell) \approx \delta_k - \delta_\ell$, so that we can check against this model later. After that, we move on to solving them without this assumption.

We make the DC approximations stated in section 3.2, except the small angle approximation. Since we are ignoring reactive power, the only relevant equation is (16), which states that

$$P_{k\ell} = -|v_k|^2 G_{k\ell} + |v_k||v_\ell|(B_{k\ell}\sin(\delta_k - \delta_\ell) + G_{k\ell}\cos(\delta_k - \delta_\ell)).$$

We assume that all voltages are approximately equal to 1. Additionally we assume resistance to be much less than reactance, so we have $G_{k\ell} \approx 0$. So this simplifies into

$$P_{k\ell} = B_{k\ell}\sin(\delta_k - \delta_\ell).$$

Since $Y_{k\ell} = -y_{k\ell}$ if $k \neq \ell$, we have that $B_{k\ell} = -b_{k\ell} \approx 1/x_{k\ell} = w_{k\ell}$. Since we assume that all lines are identical, all weights are equal. We may just as well work in a per-unit system and set this weight to 1. So we get

$$P_{k\ell} = \sin(\delta_k - \delta_\ell) \tag{18}$$

and hence we find the power flow equations

$$\begin{aligned} \sin(\delta_1 - \delta_2) + \sin(\delta_1 - \delta_3) &= P_1 \\ \sin(\delta_2 - \delta_1) + \sin(\delta_2 - \delta_3) &= P_2 \\ \sin(\delta_3 - \delta_2) + \sin(\delta_3 - \delta_2) &= P_3 \end{aligned} \tag{19}$$

that we need to solve, given the power injections P_1, P_2, P_3.

4.1.1 The full DC approximation

In the full DC approximation, we also assume that $\sin(\delta_k - \delta_\ell) \approx \delta_k - \delta_\ell$. With this assumption, the system becomes

$$\begin{aligned} 2\delta_1 - \delta_2 - \delta_3 &= P_1 \\ -\delta_1 + 2\delta_2 - \delta_3 &= P_2 \\ -\delta_1 - \delta_2 + 2\delta_3 &= P_3 \end{aligned} \tag{20}$$

which can also be written as

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix}.$$

In shorthand, we can write this as

$$\mathbf{L}\delta = \mathbf{p}. \tag{21}$$

Since the matrix $\mathbf{L}$ is singular, there is not just one solution to this system. The kernel of $\mathbf{L}$ is spanned by $(1,1,1)^T$, which means that if we increase all δ_i by the same amount, the power injections do not change. This makes sense since only the phase angle difference

is relevant for the line power flows. But it does mean that we have to remove a degree of freedom to solve the system. There are several common choices for this, one of them is to set $\delta_1 = 0$. This deletes the first column of the matrix, and removes δ_1 from the vector δ. We can also remove the third row[5] without losing information, since the third equation directly follows from the first two. This way we get the matrix equation

$$\begin{pmatrix} -1 & -1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}. \tag{22}$$

By taking the inverse, we find the solution

$$\begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -1 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -P_1 + P_2 \\ -2P_1 - P_2 \end{pmatrix}. \tag{23}$$

Since $P_{k\ell} = \delta_k - \delta_\ell$ and $P_3 = -P_1 - P_2$ (the total injected power must add up to zero, as there are no losses in the lines in this model), we find

$$\begin{aligned} P_{13} &= \frac{1}{3}(P_1 - P_3) \\ P_{32} &= \frac{1}{3}(P_3 - P_2) \\ P_{21} &= \frac{1}{3}(P_2 - P_1) \end{aligned} \tag{24}$$

as the solution to the power flow equations in the full DC approximation. Note that with the reintroduction of P_3 this becomes cyclic. The point of working this out is that we can use it as a sanity check. If we take the solutions we find without the full DC approximation and then apply the approximation on those solutions, we should recover (24). With this in mind, we are ready to solve the model without the small angles.

4.1.2 The model reduced to a single equation

We return to the model that we originally intended to solve, without the small angle approximation. Recall that the system then becomes

$$\begin{aligned} \sin(\delta_1 - \delta_2) + \sin(\delta_1 - \delta_3) &= P_1 \\ \sin(\delta_2 - \delta_1) + \sin(\delta_2 - \delta_3) &= P_2 \\ \sin(\delta_3 - \delta_1) + \sin(\delta_3 - \delta_2) &= P_3. \end{aligned}$$

First of all, we notice that the three equations sum to zero (since $P_1 + P_2 + P_3 = 0$) and hence we can omit the last equation. Secondly, we can set $\delta_1 = 0$ without loss of generality, since only the phase angle differences are relevant. This simplifies the system (19) into

$$\begin{aligned} \sin(-\delta_2) + \sin(-\delta_3) &= P_1 \\ \sin(\delta_2) + \sin(\delta_2 - \delta_3) &= P_2. \end{aligned} \tag{25}$$

[5]You may wonder why I did not remove the first row. That is also possible, but in the next computation we get rid of the third row and it is nice to run parallel to this computation.

The first equation can be written as[6]

$$\sin(\delta_3) = -P_1 - \sin(\delta_2) \tag{26}$$

which also means that[7]

$$\cos(\delta_3) = \sqrt{1 - (P_1 + \sin(\delta_2))^2} \tag{27}$$

and with an addition formula the second equation of (19) can be written as

$$\sin(\delta_2) + \sin(\delta_2)\cos(\delta_3) - \cos(\delta_2)\sin(\delta_3) = P_2. \tag{28}$$

We can now write this as a function of δ_2 only, as

$$\sin(\delta_2) + \sin(\delta_2)\sqrt{1 - (P_1 + \sin(\delta_2))^2} - \cos(\delta_2)(-P_1 - \sin(\delta_2)) = P_2. \tag{29}$$

We can define $z = \sin(\delta_2)$ so simplify this into

$$z + z\sqrt{1 - (P_1 + z)^2} + \sqrt{1 - z^2}(P_1 + z) = P_2. \tag{30}$$

All geometric functions have been removed, but now we have square roots in the equation. So a logical next step in solving the equation is getting rid of the square roots, so as to obtain a polynomial equation. First, turn the equation into

$$z\sqrt{1 - (P_1 + z)^2} = P_2 - z - \sqrt{1 - z^2}(P_1 + z)) \tag{31}$$

and take the square[8]

$$\begin{aligned}
z^2(1 - (P_1 + z)^2) &= (P_2 - z - \sqrt{1 - z^2}(P_1 + z))^2 \\
z^2(1 - (P_1 + z)^2) &= (P_2 - z)^2 + (1 - z^2)(P_1 + z)^2 - 2(P_2 - z)\sqrt{1 - z^2}(P_1 + z) \\
2(P_2 - z)\sqrt{1 - z^2}(P_1 + z) &= (P_2 - z)^2 + (1 - z^2)(P_1 + z)^2 - z^2(1 - (P_1 + z)^2) \\
2(P_2 - z)\sqrt{1 - z^2}(P_1 + z) &= (P_2 - z)^2 + (1 - z^2)(P_1 + z)^2 - z^2 + z^2(P_1 + z)^2 \\
2(P_2 - z)\sqrt{1 - z^2}(P_1 + z) &= (P_2 - z)^2 + (P_1 + z)^2 - z^2 \\
4(P_2 - z)^2(1 - z^2)(P_1 + z)^2 &= ((P_2 - z)^2 + (P_1 + z)^2 - z^2)^2
\end{aligned}$$

to obtain the equation

$$S(z) := 4(P_2 - z)^2(1 - z^2)(P_1 + z)^2 - ((P_2 - z)^2 + (P_1 + z)^2 - z^2)^2 = 0 \tag{32}$$

which is a polynomial equation of degree 6 in terms of z. Written out, it becomes

$$\begin{aligned}
&4z^6 + 8(P_1 - P_2)z^5 + (4P_1^2 - 16P_1P_2 + 4P_2^2 - 3)z^4 + 4(P_2 - P_1)(1 + 2P_1P_2)z^3 \\
&+ (2P_1^2 + 2P_2^2 + 4P_1^2P_2^2 + 8P_1P_2)z^2 + 4(P_1 - P_2)(P_1 + P_2)^2z + (P_1 - P_2)^2(P_1 + P_2)^2 = 0.
\end{aligned} \tag{33}$$

[6] I avoid writing $\delta_3 = \arcsin(-P_1 - \sin(\delta_2))$, due to the fact that $\sin(\arcsin(x))$ is not always x but depends on the chosen boundary conditions.

[7] Note that there could be a minus sign here, but as we will square out all of the square roots later this does not matter. It is possible to do the computation while avoiding this problem altogether, but the reason I chose not to is to make the process more intuitive.

[8] All these equations are equivalent. To save writing I do not write $\iff$ every line.

By solving this equation for z given values P_1, P_2, we can compute all line power flows: we have $P_{21} = \sin(\delta_2 - \delta_1) = \sin(\delta_2) = z$, then $P_{13} = P_{21} + P_1$ and $P_{32} = P_{13} + P_3$. This way we can define a function

$$F : \mathbb{R}^3 \to \mathbb{R}^3 : I = F(P) \tag{34}$$

where I is the vector of currents, and P the vector of power injections. The function F is now implicitly defined (using the solution z of an equation of degree 6), so it makes sense to look for an explicit version of F.

4.1.3 Solving the three-node model equation

After generating the values P_i, we know that z must be a solution of (33). Also, because $z = \sin(\delta_2)$, we have the constraint that $-1 \leq z \leq 1$. However, this leads us to a new problem: we can find more than one solution for z between -1 and 1. In fact, all six solutions are normally real numbers between -1 and 1. The fact that the solutions are in this interval can be explained as follows[9]. Since all P_i are much smaller than 1 (as perturbations are small), we can approximate equation (33) by setting $P_1 = P_2 = 0$ which leaves

$$4z^6 - 3z^4 = 0 \tag{35}$$

with solutions $z = \pm\sqrt{3/4} \approx \pm 0.866$ and four solutions $z = 0$. Indeed, in practice we find one solution that is about 0.866, one that is about -0.866 and four that are very close to 0. Note that the solutions $z = \pm\sqrt{3/4}$ correspond to δ_2 and δ_3 being approximately $2\pi/3$ and $4\pi/3$.

We have six solutions, and since $P_{12} = \sin(\delta_1 - \delta_2) = -\sin(\delta_2) = -z$, all six lead to different power flows. However, given fixed power injections, we expect the system to have only one solution. The two solutions with $z \approx \pm\sqrt{3/4}$ are clearly not physically relevant, since they imply that there is a power flow when there are no power injections, so we can eliminate these. However, the other four solutions all satisfy that $z \to 0$ when $P_1, P_2 \to 0$, and hence it seems that we have four realistic solutions to our system rather than one. And indeed, if we numerically solve the equation for small values P_1, P_2, we indeed get four different power flows, and all of them are in fact physically possible. We will get to how this can be in a moment: first we need to know more about these four solutions.

To get more information on the location of the zeroes, we may try to interpolate them. If we find $z_1, z_2, \ldots, z_7$ such that $z_1, z_3, z_5, z_7 < 0$ and $z_2, z_4, z_6 > 0$, then we know that the interval (z_i, z_{i+1}) contains a zero for every $1 \leq i \leq 6$, and hence we (approximately) located all zeroes.

[9]The fact that they are all real under normal conditions will be shown later.

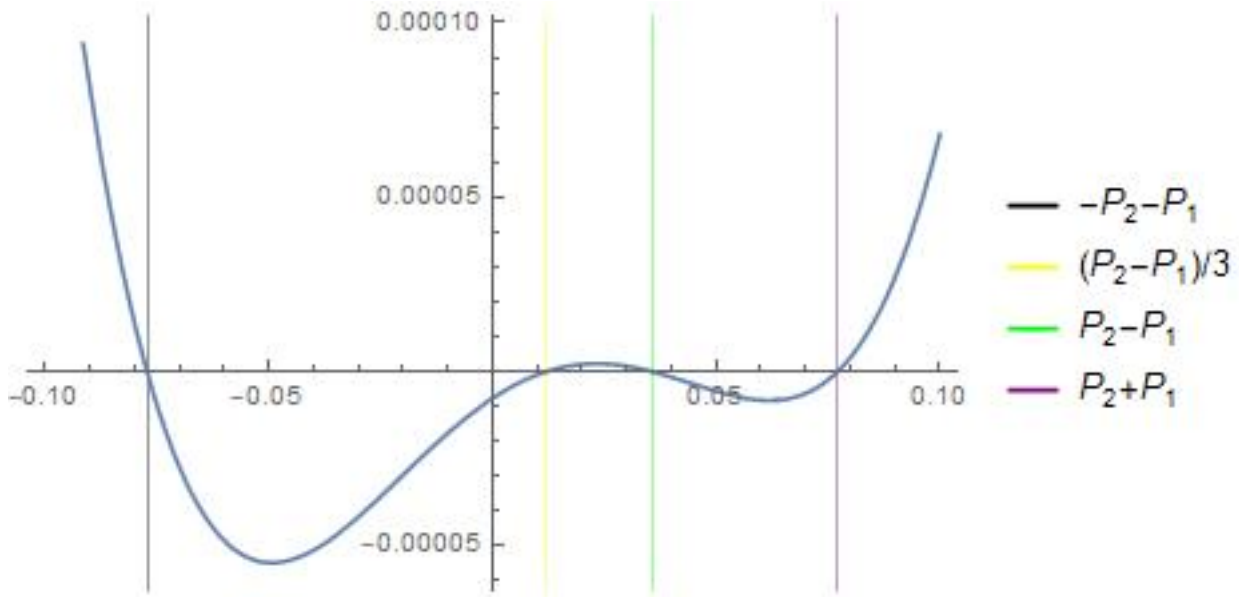

Figure 3: Zooming in on the polynomial $S(z)$ near $z = 0$. Zeroes of this polynomial correspond with solutions to the power flow equations, where z is the power flow from node 2 to node 1. For small perturbations $\mathbf{p}$, the polynomial $S(z)$ always has four zeroes with z close to 0. They can very accurately (though not exactly) be expressed in terms of the power injections P_1 and P_2.

If we plug in several linear combinations of P_1 and P_2, we get the following outcomes:

$$\begin{aligned}
S(0) &= -(P_1+P_2)^2(P_1-P_2)^2 \\
S(-P_1) &= -P_2^2(2P_1+P_2)^2 \\
S(P_2) &= -P_1^2(2P_2+P_1)^2 \\
S(P_2+P_1) &= -4P_1^2(P_1+P_2)^2(2P_1+P_2)^2 \\
S(P_2-P_1) &= -4P_1^2(P_1-P_2)^2P_2^2 \\
S(-P_2-P_1) &= -4P_2^2(P_1+P_2)^2(P_1+2P_2)^2 \\
S(\tfrac{1}{3}(P_2-P_1)) &= -\frac{4}{729}(P_1-P_2)^2(2P_1+P_2)^2(P_1+2P_2)^2.
\end{aligned}$$

The difficulty here is that all of these numbers are always negative. In fact, there is no simple linear combination of P_1 and P_2 for which S always returns a positive value. So we have to take a different approach in order to find these values. Is there any difference between the seven negative outcomes for S that we got? Yes: in the bottom four equations, the polynomial on the right is of degree 6 in terms of P_1, P_2. In the top three, it is of degree 4. This means that (assuming that P_1, P_2 are small) the right hand sides on the bottom four equations are (in absolute value) a lot smaller than the top three. So, we take the bottom four equations as our 'guesses' for the zeroes of S. They are plotted in Figure 3. The 'near-solutions' yield remarkable accuracy to the actual solutions (we will compute how accurate later on). Now it is possible to find our interpolating values z_2, z_4, z_6, using the following steps:

1. Order the four near-solutions $P_2 + P_1, P_2 - P_1, \frac{1}{3}(P_2 - P_1), -P_2 - P_1$ from smallest to largest (not taking absolute value!): name them a, b, c, d in this order
2. We can take $z_2 = -1/2$ and $z_6 = 1/2$: for small P_1, P_2 this is always positive
3. We take $z_4 = (b + c)/2$: now using the fact that we know $d > c > b > a$ we can prove that $S(z_4) > 0$

With this interpolation, we find that there are indeed always six real solutions, as was claimed before. We will not prove these statements here since it mostly involves a lot of not very interesting writing. Note that z_4 can always be written as a linear combination of P_1 and P_2, however which linear combination depends on P_1 and P_2 since it depends on the ordering of the near-solutions. This also explains why we could not find this linear combination previously.

Using the near-solutions we obtained, we can easily compute the power flows. We distinguish between the different cases. As an example, we take $z = P_2 + P_1$.
In this case, we have $P_{21} = \sin(\delta_2 - \delta_1) = \sin(\delta_2) = z = P_1 + P_2 = -P_3$ (recall that we already chose $\delta_1 = 0$). By load balance, we have $P_{13} = P_{21} + P_1 = 2P_1 + P_2 = P_1 - P_3$. Also $P_{32} = P_{13} + P_3 = P_1$.
I will not work out the other ones here (it is fairly straightforward) but the results are in the table below.

	P_{13}	P_{32}	P_{21}
$z = P_2 + P_1$	$P_1 - P_3$	P_1	$-P_3$
$z = P_2 - P_1$	P_2	$-P_1$	$P_2 - P_1$
$z = -P_2 - P_1$	$-P_2$	$P_3 - P_2$	P_3
$z = \frac{1}{3}(P_2 - P_1)$	$\frac{1}{3}(P_1 - P_3)$	$\frac{1}{3}(P_3 - P_2)$	$\frac{1}{3}(P_2 - P_1)$

Note that with $z = \frac{1}{3}(P_2 - P_1)$, we recover (24), so we do indeed find this solution again. However, we also find three new solutions, that do not occur in the DC approximation. The reason for this is that if the power flows are small, the phase angle differences need not to be. If $z = \sin(\delta_1 - \delta_2)$ is small, then either $\delta_1 - \delta_2$ is small, or $\delta_1 - \delta_2 \approx \pi$. So instead of $\delta_2 \approx 0$, we may also have $\delta_2 \approx \pi$, and the same goes for δ_3. This gives us four possible configurations: $(\delta_2, \delta_3) \approx (0, 0), (0, \pi), (\pi, 0)$ or (π, π). Each one of these configurations turns out to correspond to a solution, not just $(\delta_2, \delta_3) \approx (0, 0)$. This was not taken into account before, which is why we find these new solutions now. In fact, it is possible to directly derive all solutions using this idea. We will work this out in detail later in section 4.1.5

4.1.4 Accuracy of the approximations

We already saw that the four solutions near zero could be approximated with $P_2 + P_1, P_2 - P_1, -P_2 - P_1, \frac{1}{3}(P_2 - P_1)$. One may wonder how accurate this approximation is. Well, we

see that for all of them the value $S(z)$ is a polynomial of order 6. If we assume P_1, P_2 to be of order σ, then we find $S(z) = \mathcal{O}(\sigma^6)$. Also, we have that

$$S'(z) = 24z^5 + 40(P_1 - P_2)z^4 + 4(4P_1^2 - 16P_1P_2 + 4P_2^2 - 3)z^3 + 12(P_2 - P_1)(1 + 2P_1P_2)z^2 + 2(2P_1^2 + 2P_2^2 + 4P_1^2P_2^2 + 8P_1P_2)z + 4(P_1 - P_2)(P_1 + P_2)^2. \quad (36)$$

Note that $z = \mathcal{O}(\sigma)$, and therefore the lowest order terms are of order $\mathcal{O}(\sigma^3)$. Ignoring all higher order terms, we get

$$S'(z) \approx -12z^3 + 12(P_2 - P_1)z^2 + 16P_1P_2z + 4(P_1 - P_2)(P_1 + P_2)^2. \quad (37)$$

We can plug in all 4 solutions, but none of them reduces the polynomial to zero, and hence we have that the leading order in $S'(z)$ is exactly σ^3, or $S'(z) = \Theta(\sigma^3)$.
If z is a near-solution and z_0 is an exact solution, we find that

$$S'(z) \approx \frac{S(z) - S(z_0)}{z - z_0}$$
$$z - z_0 = \frac{\mathcal{O}(\sigma^6)}{\Theta(\sigma^3)} = \mathcal{O}(\sigma^3)$$

and hence the error in our estimate is $\mathcal{O}(\sigma^3)$, and the relative error is $(z - z_0)/z_0 = \mathcal{O}(\sigma^2)$. For such a simple approximation this can be considered very accurate. This has two major benefits:

1. We can couple any exact solution of $S(z) = 0$ to a near-solution that is a linear combination of components of $\mathbf{p}$. This means that if P_1, P_2 are no longer constant we can track every solution, without confusing it with a different solution, even if the z values may be the same at one point.

2. If we want to increase the precision of our guess using numerical methods, we already have a very good initial guess. This also means that there is very little risk of the numerical algorithm converging to a different solution than intended.

4.1.5 Directly deriving the linear approximations

Now that we know that the additional solutions arise due to δ_2, δ_3 being approximately π instead of 0, we can actually linearize the equations around those points instead, in order to get the near-solutions. We already know what happens if we linearize around the point $(\delta_2, \delta_3) = (0, 0)$ because we did that in section 3.2. We will first linearize around $(\delta_2, \delta_3) = (\pi, 0)$. In this case, we define $\delta_2' = \delta_2 + \pi$. Then equation (19) becomes

$$\begin{aligned} -\sin(\delta_1 - \delta_2') + \sin(\delta_1 - \delta_3) &= P_1 \\ -\sin(\delta_2' - \delta_1) - \sin(\delta_2' - \delta_3) &= P_2 \\ \sin(\delta_3 - \delta_1) - \sin(\delta_3 - \delta_2') &= P_3. \end{aligned}$$

If we make the small angle approximation now, we get

$$\begin{aligned} \delta_2' - \delta_3 &= P_1 \\ \delta_1 - 2\delta_2' + \delta_3 &= P_2 \\ \delta_2' - \delta_1 &= P_3 \end{aligned}$$

or, in matrix form

$$\begin{pmatrix} 0 & 1 & -1 \\ 1 & -2 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2' \\ \delta_3 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix}. \tag{38}$$

Once again, we can set $\delta_1 = 0$ and eliminate the third equation (since the sum of all three equations is zero), and simplify this to the system

$$\begin{pmatrix} 1 & -1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \delta_2' \\ \delta_3 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}. \tag{39}$$

We can invert the matrix to obtain the solution

$$\begin{pmatrix} \delta_2' \\ \delta_3 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} P_1 + P_2 \\ 2P_1 + P_2 \end{pmatrix}. \tag{40}$$

Hence, we find

$$\begin{aligned} P_{21} &= \sin(\delta_2 - \delta_1) \\ &= \sin(\delta_2) \\ &= -\sin(\delta_2') \\ &\approx -\delta_2' \\ &= -P_1 - P_2 \end{aligned}$$

which means that this case corresponds to the near-solution $z = -P_2 - P_1$, so the configuration $(P_{13}, P_{32}, P_{21}) = (-P_2, P_2 + P_3, -P_3)$

The computation for the other configurations is similar: we get that $\delta_2 \approx 0$ and $\delta_3 \approx \pi$ yields the near-solution $z = P_2 - P_1$ and $(P_{13}, P_{32}, P_{21}) = (P_2, -P_1, P_2 - P_1)$. Finally $\delta_2, \delta_3 \approx \pi$ yields $z = P_2 + P_1$ and $(P_{13}, P_{32}, P_{21}) = (P_1 - P_3, P_1, -P_3)$. So we exactly recover the four solutions that we found, without having to go through solving a polynomial equation of degree 6!

Another important benefit of this approach is that it answers an important question that you may or may not have asked yourself: is it possible that the system 'switches' between the four solutions? Suppose that δ_2 changes from 0 to π. Then since δ_2 is continuous in time, we at some point have $\delta_2 - \delta_1 = \pi/2$ and hence $\sin(\delta_2 - \delta_1) = 1$ (or $\delta_2 - \delta_1 = -\pi/2$, but the argument remains the same). We know that the sines are usually small, hence such an event is highly unlikely in practice, and can only occur when we at some point in time are at a boundary point of the domain of all $\mathbf{p}$.

We can actually apply this method to larger networks as well. A major benefit is that we still obtain a linear problem, just like in the full DC approximation, and still get all possible solutions rather than just one. For larger networks we expect the corresponding polynomial to be of a very high degree (if it is possible to compute it at all), which is why this approach can be useful. A downside, however, it that it does not grant us insight in how accurate the approximation is. In the polynomial we could see that the error in the approximation is of order $\mathcal{O}(\sigma^2)$, which is not directly clear now. Also it does not allow us to improve our approximation. Using the polynomial, we can start at a near-solution and use any numerical algorithm to find a more accurate approximation, which we cannot do now either. It would be interesting to find an algorithm that can improve the accuracy of the near-solutions without computing the entire polynomial, however a big question is if this would be any faster in a computer.

4.2 Probability of overflow

Using our near-solutions, we are returning to the original question: analyzing the probability of an overload in a power network.

4.2.1 The probability distribution

To compute such probabilities, we need a probability density function $f(P_1, P_2)$ (note that since $P_3 = -P_1 - P_2$ we do not need P_3 as a variable). It makes sense to choose a normal distribution for this. The only problem is that P_1, P_2 are not independent, since we have the constraint $P_1 + P_2 + P_3 = 0$. So the question is how we can choose the entries of $\mathbf{p}$ according to a normal distribution while still forcing them to add up to zero. This can be handled using a multivariate normal distribution. We define $\mathbf{r}$ as a vector with three components, such that R_1, R_2, R_3 are i.i.d. with $R_i \sim \mathcal{N}(0, \sigma^2)$. Then we set

$$\begin{aligned} P_1 &= \tfrac{2}{3}R_1 - \tfrac{1}{3}R_2 - \tfrac{1}{3}R_3 \\ P_2 &= -\tfrac{1}{3}R_1 + \tfrac{2}{3}R_2 - \tfrac{1}{3}R_3 \\ P_3 &= -\tfrac{1}{3}R_1 + \tfrac{2}{3}R_2 - \tfrac{1}{3}R_3. \end{aligned} \tag{41}$$

This way, we are guaranteed to have $P_1 + P_2 + P_3 = 0$. And since the sum of several normal distributions is still a normal distribution, the P_i still follow a normal distribution. If we were to continue the computation with P_3, then the constraint $-P_1 - P_2$ means that the distribution is only defined on a lower-dimensional subspace of $\mathbb{R}^3$, necessitating the usage of Lebesgue integrals to compute the probability of an event. To avoid this, one option is to perform the integration over R_1, R_2, R_3. But a better strategy at this point is to just forget about P_3 and work with P_1, P_2 only. Define

$$P' = A\mathbf{r}, \quad P' = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}, \quad A = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix}. \tag{42}$$

This means that P' is also a multivariate normal distribution, and hence we know that its probability density function is given by

$$f(P_1, P_2) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left[-\frac{1}{2} {P'}^T \Sigma^{-1} P' \right] \tag{43}$$

where

$$\Sigma = \sigma^2 AA^T = \sigma^2 \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}.$$

The determinant of this matrix is equal to $|\Sigma| = \sigma^4/3$. Note that Σ is positive definite with eigenvalues $\sigma^2/3$ and σ^2, so that the integral of f over $\mathbb{R}^2$ converges.
The definition of $f(P_1, P_2)$ requires the inverse of Σ: it is equal to

$$\Sigma^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{44}$$

and with this, we find

$$f(P_1, P_2) = \frac{\sqrt{3}}{2\pi\sigma^4} \exp\left[-\frac{1}{2\sigma^2} \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \right]. \tag{45}$$

It will be difficult to even make a proper estimate of such an integral as P_1 and P_2 are not independent, since the matrix Σ^{-1} has off-diagonal entries that are non-zero. To make it easier to handle, we can apply a coordinate transformation to decouple them.
We define[10]

$$T_1 = \frac{\sqrt{3}}{2\sigma}(R_1 + R_2), \quad T_2 = \frac{1}{2\sigma}(R_1 - R_2)$$

so that we have

$$\begin{aligned} T_1^2 + T_2^2 &= \frac{1}{4\sigma^2}\big(3(P_1 + P_2)^2 + (P_1 - P_2)^2\big) \\ &= \frac{1}{4\sigma^2}\big(4P_1^2 + 4P_2^2 + 4P_1P_2\big) \\ &= \frac{1}{2\sigma^2} \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}. \end{aligned}$$

This way we can write the probability density function as[11]

$$g(T_1, T_2) = \frac{1}{\pi} \exp\big[-(T_1^2 + T_2^2)\big]. \tag{46}$$

4.2.2 Computing the overload probabilities

Now that we have this distribution, we can compute the overload probabilities. Assume that all lines have the same capacity $a\sigma$. There are two different cases we need to consider.

1. In one case, we can force the system to take the 'safest' solution $z = \frac{1}{3}(P_2 - P_1)$

[10] To save writing I am simply posing this substitution as a result here, rather than explaining how I found it, since it is not really related to the subject of the thesis.

[11] The new forefactor can be computed using a Jacobian, or just using the fact that the integral over the entire space $\mathbb{R}^2$ must be 1 for a probability distribution.

2. In another case we cannot, and then all four physically relevant solutions need to be safe

In the first case, we have a large number of conditions that need to be satisfied in order for the system to be safe. Note that the safety of the three solutions $z = P_2 + P_1$, $z = P_2 - P_1$ and $z = -P_2 - P_1$ implies the safety of the fourth solution $z = \frac{1}{3}(P_2 - P_1)$, since the line flows in this solution are the average of those in the other three solutions. We can take all conditions from the table in section 4.1.3, and use $P_3 = -P_1 - P_2$.

$$\begin{aligned}
-a\sigma \leq P_1 \leq a\sigma \\
-a\sigma \leq P_2 \leq a\sigma \\
-a\sigma \leq P_1 + P_2 \leq a\sigma \\
-a\sigma \leq P_1 - P_2 \leq a\sigma \\
-a\sigma \leq 2P_1 + P_2 \leq a\sigma \\
-a\sigma \leq P_1 + 2P_1 \leq a\sigma
\end{aligned}$$

Note that the top three inequalities can be derived from the bottom three. Hence, we can reduce this system to the three boundaries

$$\begin{aligned}
-a\sigma \leq P_1 - P_2 \leq a\sigma \\
-a\sigma \leq 2P_1 + P_2 \leq a\sigma \\
-a\sigma \leq P_1 + 2P_1 \leq a\sigma.
\end{aligned}$$

We can use the same substitution as before to obtain

$$\begin{aligned}
-a \leq 2T_2 \leq a \\
-a \leq T_1\sqrt{3} + T_2 \leq a \\
-a \leq T_1\sqrt{3} - T_2 \leq a
\end{aligned}$$

as boundaries for the domain D. This domain is plotted in Figure 4. Note that these lines bound a regular hexagon, with side length $a/\sqrt{3}$. This means that we can bound the domain D by

$$B(0, a/2) \subset D \subset B(0, a/\sqrt{3}). \tag{47}$$

We can evaluate the integral

$$\begin{aligned}
\frac{1}{\pi}\iint_{B(0,a/2)} \exp\Big[-(T_1^2 + T_2^2)\Big] dT_1 dT_2 &= \frac{1}{\pi}\int_0^{2\pi} d\phi \int_0^{a/2} r \exp\Big[-(r^2)\Big] dr \\
&= -\frac{1}{\pi} 2\pi \frac{1}{2} e^{-r^2}\Big|_{r=0}^{r=a/2} \\
&= 1 - e^{-a^2/4}
\end{aligned}$$

and similarly

$$\frac{1}{\pi}\iint_{B(0,a/\sqrt{3})} \exp\Big[-(T_1^2 + T_2^2)\Big] dT_1 dT_2 = 1 - e^{-a^2/3}. \tag{48}$$

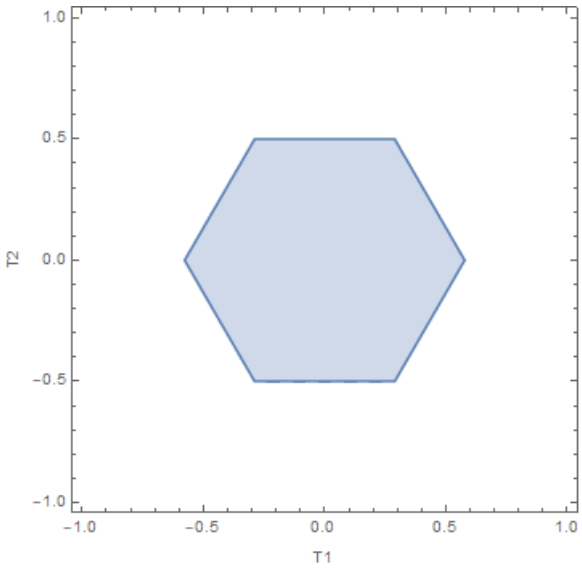

Figure 4: The area of 'safe' power flows over which we have to integrate, in case we want any power flow configuration to be safe. We can approximate this integral by drawing a circle inside of it, or outside of it. If the system only needs to be safe for the DC approximation, then the hexagon becomes three times larger, and hence the chance of an overflow is reduced significantly.

Therefore, we obtain

$$e^{-a^2/3} \leq \mathbb{P}(\text{overload is possible}) \leq e^{-a^2/4}. \tag{49}$$

Recall that the event 'overload is possible' means that at least one configuration would lead to an overflow given the power injections.
The problem with this approach is that the two bounds do not converge to each other in the case $a \to \infty$ and therefore it is not the most useful for large deviations. In order to get a good approximation in the case $a \to \infty$, we can make use of half-planes. If we consider the half-plane H defined by the equation $T_2 \geq a/2$, then we have

$$\iint_H \frac{1}{\pi} \exp\big[-T_1^2 - T_2^2\big] dT1 dT2 = \int_{a/2}^{\infty} \frac{1}{\sqrt{\pi}} \exp\big[-T_2^2\big] \approx \frac{1}{a\sqrt{\pi}} e^{-a^2/4} \tag{50}$$

using the error function approximation in the limit of large a. We can now view the complement D^C of D as the union of six half-planes, one half-plane on the outside of every side of the hexagon. Label the six half-planes $H_1, \ldots, H_6$. Then we have

$$\begin{aligned}
\iint_{D^C} \exp\big[-T_1^2 - T_2^2\big] dT1 dT2 &\leq \sum_{i=1}^{6} \iint_{H_i} \exp\big[-T_1^2 - T_2^2\big] dT1 dT2 \\
&\approx \sum_{i=1}^{6} \frac{1}{a\sqrt{\pi}} \exp\big[-a^2/4\big] \\
&= \frac{6}{a\sqrt{\pi}} \exp\big[-a^2/4\big].
\end{aligned}$$

Note that the $\leq$ is there due to overlap between hyperplanes, leading to some points being counted more than once. However, any point counted twice is at least a distance $a/\sqrt{3} > a/2$ away from the origin, and hence for large a these points become insignificant compared to points located only $a/2$ away from the origin for large a. All together we find the approximation

$$\mathbb{P}(\text{overload is possible}) \approx \frac{6}{a\sqrt{\pi}} \exp\big[-a^2/4\big] \tag{51}$$

for the overflow probability. We can compare this to numerical results[12].

a	exact integral	approximation
5	0.00118399	0.00130697
10	$4.61223 \cdot 10^{-12}$	$4.70126 \cdot 10^{-12}$
25	$1.86958 \cdot 10^{-69}$	$1.87553 \cdot 10^{-69}$
50	$2.49005 \cdot 10^{-273}$	$2.49204 \cdot 10^{-273}$

It is clear that the approximation gets better in the limit $a \to \infty$, as we expected since the overlap becomes less important for larger values a. Unfortunately it is quite hard to compute the exact probabilities for larger values of a, since the integral becomes so small. This is also exactly the reason that large deviations theory is useful: for estimating probabilities that cannot be computed otherwise.
Now we move on to the case in which we can force the system to take the (configuration corresponding to the) DC approximation. In this case the constraints in terms of P_1 and P_2 are

$$\begin{aligned} -a\sigma \leq \frac{1}{3}(P_1 - P_2) \leq a\sigma \\ -a\sigma \leq \frac{1}{3}(2P_1 + P_2) \leq a\sigma \\ -a\sigma \leq \frac{1}{3}(P_1 + 2P_1) \leq a\sigma \end{aligned}$$

or, equivalently

$$\begin{aligned} -3a\sigma \leq P_1 - P_2 \leq 3a\sigma \\ -3a\sigma \leq 2P_1 + P_2 \leq 3a\sigma \\ -3a\sigma \leq P_1 + 2P_1 \leq 3a\sigma. \end{aligned}$$

Note that this is the same system of equations we analyzed earlier, however a is replaced with $3a$. Hence we can get the desired outcomes by replacing a with $3a$. We hence get

$$e^{-3a^2} \leq \mathbb{P}(\text{overload}) \leq e^{-9a^2/4} \tag{52}$$

[12]Note that the exact integral is still not the exact overload probability: for this we would need to also incorporate the fact that the near-solutions are not exact. If we take the limit $\sigma \to 0$ then it does become the exact overload probability. The point of this is showing that the approximation converges to the integral in the limit $a \to \infty$.

and in the limiting case $a \to \infty$ we get

$$\mathbb{P}(\text{overload}) \approx \frac{2}{a\sqrt{\pi}} \exp\big[-9a^2/4\big]. \tag{53}$$

Note that this is significantly smaller than in (51). So if we can force the system to take this configuration, we know that the probability of an overload will be reduced significantly as compared to not taking any action. In practice, this is especially relevant when we restart a system, because that may lead to the system taking a different configuration that we wish. So a grid operator must beware that the angles are perfectly aligned when everything starts up, and not flipped!

4.3 An approach with the contraction principle

We can also try to tackle the problem of finding the evolution of the overload probability with the contraction principle. This is a way of checking our results, and also a good example to see how the contraction principle works in practice. To use the contraction principle, we first need a probability measure on the initial space, the space of power injections. Recall that the probability density for the power injections is equal to

$$f(P_1, P_2) = \frac{\sqrt{3}}{2\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}(P_1^2 + P_1P_2 + P_2^2)\right] \tag{54}$$

with σ the standard deviation in the power injections P_1, P_2. Hence, we can define the probability measure in terms of ϵ as

$$\mu_\epsilon(K, L) = \frac{\sqrt{3}}{2\pi\epsilon} \exp\big[-(K^2 + KL + L^2)/\epsilon\big]. \tag{55}$$

We let X be the space of all perturbations (K, L) (so $X = \mathbb{R}^2$). The rate function becomes

$$I(K, L) = K^2 + KL + L^2. \tag{56}$$

As the function T, we take the largest line flow in the system (in absolute value, since an overload can happen regardless of the direction of the current). Let a be the capacity of the lines. If we assume the system to be in the solution corresponding to the DC approximation, then the function T is (approximately) equal to

$$T(K, L) = \text{absmax}\left(\frac{1}{3}(L - K), \frac{1}{3}(2K + L), \frac{1}{3}(-K - 2L)\right). \tag{57}$$

The contraction principle applied to the transformation T states that

$$J(c) = \inf\{I(K, L) \mid (K, L) \in X \text{ and } T(K, L) = c\} \tag{58}$$

is a rate function for the largest line flow c. Computing this function J comes down to solving the optimization problem

$$\begin{aligned} &\text{minimize } I(K, L) \\ &\text{subject to } T(K, L) \geq c \end{aligned}$$

and we want to find the probability that $T(K, L) \geq c$ for some limit c. To solve this, we can apply Lagrange multipliers:

$$\begin{aligned} 2K^\star + L^\star &= \lambda^\star \frac{\partial T}{\partial K}\bigg|_{K=K^\star, L=L^\star} \\ K^\star + 2L^\star &= \lambda^\star \frac{\partial T}{\partial L}\bigg|_{K=K^\star, L=L^\star} \\ T(K^\star, L^\star) &= c. \end{aligned}$$

For notational simplicity, we omit the condition $K = K^\star, L = L^\star$ at the derivatives from now on. The derivative cannot be computed that easily, but we can distinguish between three cases:

- $\frac{1}{3}(L - K)$ is the largest (in absolute value). In that case, we have

  $$\frac{\partial T}{\partial K} = -\frac{1}{3}, \quad \frac{\partial T}{\partial L} = \frac{1}{3}.$$

  Note that if $\frac{1}{3}(K - L)$ were negative, this would only invert the sign of $\lambda^\star$ and not change $K^\star$ or $L^\star$. We get the system

  $$\begin{aligned} 2K^\star + L^\star &= -\frac{1}{3}\lambda^\star \\ K^\star + 2L^\star &= \frac{1}{3}\lambda^\star \\ \frac{1}{3}(L^\star - K^\star) &= c. \end{aligned}$$

  The first two equations added together yields $K^\star = -L^\star$. Then the final equation becomes $\frac{2}{3}L^\star = c$, and hence $L^\star = \frac{3}{2}c$ and $K^\star = -\frac{3}{2}c$. Hence, we get

  $$I(K^\star, L^\star) = \frac{9}{4}c^2$$

- $\frac{1}{3}(2K + L)$ is the largest (in absolute value). In that case, we have

  $$\frac{\partial T}{\partial K} = \frac{2}{3}, \quad \frac{\partial T}{\partial L} = \frac{1}{3}.$$

  We get the system

  $$\begin{aligned} 2K^\star + L^\star &= \frac{2}{3}\lambda^\star \\ K^\star + 2L^\star &= \frac{1}{3}\lambda^\star \\ \frac{1}{3}(2K^\star + L^\star) &= c. \end{aligned}$$

  Subtract the second equation twice from the first one: this yields $-3L^\star = 0$, hence $L^\star = 0$. Then the final equation becomes $\frac{2}{3}K^\star = c$, and hence $K^\star = \frac{3}{2}c$ and $L^\star = 0$. Hence, we get

  $$I(K^\star, L^\star) = \frac{9}{4}c^2$$

- $\frac{1}{3}(-K-2L)$ is the largest (in absolute value). In that case, we have

$$\frac{\partial T}{\partial K} = -\frac{1}{3}, \quad \frac{\partial T}{\partial L} = -\frac{2}{3}.$$

We get the system

$$\begin{aligned} 2K^\star + L^\star &= -\frac{1}{3}\lambda^\star \\ K^\star + 2L^\star &= -\frac{2}{3}\lambda^\star \\ \frac{1}{3}(-K^\star - 2L^\star) &= c. \end{aligned}$$

Subtract the first equation twice from the second one: this yields $-3K^\star = 0$, hence $K^\star = 0$. Then the final equation becomes $-\frac{2}{3}L^\star = c$, and hence $L^\star = -\frac{3}{2}c$ and $K^\star = 0$. Hence, we get

$$I(K^\star, L^\star) = \frac{9}{4}c^2.$$

Unsurprisingly, every line returns the same answer: this makes sense as the system is symmetric, and hence every line has the same probability of exceeding flow c. In every case, we get $I(K^\star, L^\star) = \frac{9}{4}c^2$. Hence, we get

$$J(c) = \inf\{I(x) \mid x \in X \text{ and } T(x) = c\} = \frac{9}{4}c^2 \tag{59}$$

Recall that in section 4.2 we got the following result (equation (53)):

$$\mathbb{P}(\text{overload}) \approx \frac{2}{a\sqrt{\pi}} \exp\big[-9a^2/4\big]$$

Recall that in that computation, the line limit was $a\sigma$ instead of c. So to compare this result we change a into c/σ, and also turn σ^2 into ϵ (because we did that in defining the probability measure in equation (55)). Then, it becomes

$$\mathbb{P}_\epsilon(\text{overload}) \approx \frac{2\sqrt{\epsilon}}{c\sqrt{\pi}} \exp\left[-\frac{9c^2}{4\epsilon}\right]. \tag{60}$$

And if we now take the limit

$$-\lim_{\epsilon\to 0} \epsilon \log(\mathbb{P}_\epsilon(E)) = -\lim_{\epsilon\to 0} \epsilon \left[\log(2) - \log(c\sqrt{\pi}) + \frac{1}{2}\log(\epsilon) - \frac{9c^2}{4\epsilon}\right] = \frac{9c^2}{4}$$

we see that $J(c)$ is indeed a rate function for the most likely overload. So the result found by the contraction principle is confirmed by our earlier, more algebraic computation.

Both of these approaches have advantages and disadvantages. A major advantage of the direct computation is the forefactor besides the exponential: we cannot recover this if we only have the rate function. A downside of the direct computation is that it is only feasible for specific models: for more complicated networks, it quickly becomes difficult or even impossible to compute the function $\mathbb{P}(\text{overload})$ exactly.

4.4 Results for the general model

We have now studied the three-cycle in detail, found the (approximate) power flows and the overflow probabilities. It is now time to move on and focus on a general power network, and see what we can find in that case. However, we will still be referring to this simple model in order to check our more general results.

4.4.1 Continuity of power flows

The first question is how the power flows changes as a function of the power injections. Does the power flow $\mathbf{f}$ vary continuously as a function of the power injection $\mathbf{p}$? In the three-node case, the near-solutions certainly do, but that does not imply that the actual solutions also vary continuously! So we need a different approach. Note that the exact solutions can be found by solving a polynomial equation of degree 6. The coefficients of this polynomial vary as a continuous function of the power injection $\mathbf{p}$. It is a known result that the roots of a polynomial are a continuous function of its coefficients[12]. So we find that the power flows must be continuous in terms of the power injections for the three-cycle network. For a larger network, this strategy does not immediately work, since we first have to reduce it to a single polynomial in order to use this theorem.

Now we look at a general network. We assume that we have full information about the present state: we know the power injection $\mathbf{p}$, the phase angles δ and the power flow $\mathbf{f}$. Our goal is to figure out how $\mathbf{f}$ varies in terms of $\mathbf{p}$. For this, we first turn to δ. Unfortunately we do not have an explicit function of δ in terms of $\mathbf{p}$. We do, however, have an explicit function J of $\mathbf{p}$ in terms of δ:

$$[J(\delta)]_i = \sum_{j \neq i} w_{ij} \sin(\delta_i - \delta_j). \tag{61}$$

We want to know what happens to δ if $\mathbf{p}_0 = \mu$ is replaced by $\mu + \epsilon\mathbf{q}$. Unfortunately the function J is not easy to invert, so we turn the process around. If we change $\delta \to \delta + \epsilon\theta$, we get the following change in $\mathbf{p}$:

$$\begin{aligned}
[J(\delta + \epsilon\theta)]_i &= \sum_{j \neq i} w_{ij} \sin\big((\delta_i + \epsilon\theta_i) - (\delta_j + \epsilon\theta_j)\big) \\
&= \sum_{j \neq i} w_{ij} \sin\big((\delta_i - \delta_j) + \epsilon(\theta_i - \theta_j)\big).
\end{aligned}$$

For notational shorthand, define $\delta_{ij} = \delta_i - \delta_j$ and θ_{ij} likewise. Assuming that ϵ is small, we can make the following expansion:

$$\begin{aligned}
&= \sum_{j \neq i} w_{ij} \sin\big(\delta_{ij} + \epsilon\theta_{ij}\big) \\
&= \sum_{j \neq i} w_{ij} \big(\sin(\delta_{ij}) + \epsilon\theta_{ij}\cos(\delta_{ij})\big) \\
&= \mu + \epsilon \sum_{j \neq i} \theta_{ij} w_{ij} \cos(\delta_{ij}).
\end{aligned}$$

This is a linear system in θ. Remember that δ is known, hence $\cos(\delta_{ij})$ is also known and can be left in the resulting expression. We now have

$$Q_i = \sum_{j \neq i} \theta_{ij} w_{ij} \cos(\delta_{ij}) \tag{62}$$

and hence we can write

$$\mathbf{q} = M\theta, \quad M_{ij} = \begin{cases} -w_{ij}\cos(\delta_{ij}) & \text{if } i \neq j \\ \sum\limits_{i \neq j} w_{ij}\cos(\delta_{ij}) & \text{if } i = j. \end{cases} \tag{63}$$

Note that the sum of every row in M is zero, and hence $\theta = (1, 1, \ldots, 1)$ yields $\mathbf{q} = 0$, as expected. In order to solve θ and in terms of $\mathbf{q}$, we can replicate the strategy of solving the power flow equations in [7]. To find θ as a function of $\mathbf{q}$, we would like to write $\theta = M^{-1}\mathbf{q}$, but that is not possible since M is not invertible. For this reason, we use the *Moore-Penrose inverse* of the matrix, which is a generalization of the inverse that exists for any matrix[13]. In this case, it is equal to

$$M^+ = (M + \frac{1}{n}J)^{-1} - \frac{1}{n}J \tag{64}$$

where $J \in \mathbb{R}^{n \times n}$ is the all-ones matrix. Now we have that $\theta = M^+\mathbf{q}$ solves[14] equation (63). If we have found θ, then we can compute the power flow changes as follows.

$$\begin{aligned} P_{ij} + \epsilon Q_{ij} &= w_{ij} \sin\left(\delta_i + \epsilon\theta_i - (\delta_j \ + \epsilon\theta_j)\right) \\ P_{ij} + \epsilon Q_{ij} &= w_{ij}\sin(\delta_{ij}) + \epsilon\theta_{ij} w_{ij}\cos(\delta_{ij}) \\ Q_{ij} &= \theta_{ij} w_{ij}\cos(\delta_{ij}). \end{aligned}$$

We can write this in matrix form too. Suppose that if $\mathbf{p} = \mu + \epsilon\mathbf{q}$, then $\mathbf{f} = \psi + \epsilon\mathbf{g}$. Then we can write

$$\mathbf{g} = N\theta, \quad N_{ei} = \begin{cases} w_{ij}\cos(\delta_{ij}) & \text{if } e = (i,j) \text{ for some } j \\ -w_{ij}\cos(\delta_{ij}) & \text{if } e = (j,i) \text{ for some } j \\ 0 & \text{otherwise.} \end{cases} \tag{65}$$

Now we can write

$$\mathbf{g} = NM^+\mathbf{q} \tag{66}$$

as the relation between the power injection and power flow. The matrix $V = NM^+$ is known as the Power Transfer Distribution Factor (PTDF). We have now found a total derivative of the transformation $\mathbf{p} \to \mathbf{f}$ in the point (μ, ψ). Does that mean that the transformation is always continuous and even differentiable? Well, in the vast majority of cases it is, but there are situations in which our argumentation fails. Specifically, it goes wrong when we try to compute the Moore-Penrose inverse of a matrix M of dimension lower than $n - 1$. Suppose that we have $\dim(M) \leq n - 2$ and try to apply this procedure. We can still compute the

[13] If you are interested in learning more about it, see [13].

[14] This implicitly picks $\theta_1 + \ldots + \theta_n = 0$, which does no harm, since only phase angle differences are important.

Moore-Penrose inverse M^+ and choose $\theta = M^+\mathbf{q}$: we now hope that $\mathbf{q} = MM^+\mathbf{q}$. But now $\text{rank}(MM^+) \leq \text{rank}(M) \leq n-2$, which means that $\mathbf{q}$ has at most $n-2$ degrees of freedom, contradicting the fact that it should have $n-1$: it means that we can no longer freely choose the power inputs! What happens here?

It is known that if $\delta_i = 0$ for all i then M always has rank $n-1$ as long as the graph is connected. So how can it happen that $\text{rank}(M) \leq n-2$? Well, we have the additional cosine factors, and it could be that a c_{ij} is equal to 0. Mathematically this would be equivalent to setting $w_{ij} = 0$, or removing the edge $\{i, j\}$, disconnecting the graph. However, in practice it is very unlikely that $c_{ij} = 0$, since it would require that $|\delta_i - \delta_j| = \pi/2$, which is very large and hence corresponds to a very strange configuration: it would also imply that $|s_{ij}| = 1$, while s_{ij} is usually close to 0. So except for these unlikely cases, the power flow vector $\mathbf{f}$ is continuous, and actually locally Lipschitz continuous in terms of $\mathbf{q}$.

4.4.2 Probability of overflow

Using the relation between power injections and power flows we can compute the probability of overloads. Assume that at first we have known power injections μ and line power flows $\mathbf{f}$. After a perturbation, we find new power injections $\mu + \epsilon\mathbf{q}$ and power flows $\mathbf{f} + \epsilon\mathbf{g}$. The line limits are summarized in the vector $\mathbf{f}^{max}$. So if the component-wise inequality $|\mathbf{f}| \leq \mathbf{f}^{max}$ is satisfied then the power flow is safe[15].

The perturbation $\mathbf{q}$ cannot simply be chosen as a normal distribution, since we need to have $\sum_{i=1}^{n} Q_i = 0$. So instead we choose (as generally as possible) a vector $\mathbf{r} \propto \mathcal{N}(0, \Sigma)$. Then, we can define the transformation matrix $A \in \mathbb{R}^{n \times n}$ by $A_{ij} = -\frac{1}{n} + \delta_{ij}$. We then have $\mathbf{q} = A\mathbf{r}$. We can now compute the probability that a specific line i fails. Recall that line i fails if $|F_i + \epsilon G_i| > F_i^{max}$, which we can also write as $|F_i + \epsilon(VA\mathbf{r})_i| > F_i^{max}$. So we have

$$\epsilon(VA\mathbf{r})_i > F_i^{max} - F_i \text{ or } \epsilon(VA\mathbf{r})_i < -F_i - F_i^{max}. \tag{67}$$

Note that both of these equations define a half-space for r, and hence the probability of overflow becomes an error function[16], which is easy to approximate.

What we are more interested in is the probability that *at least* one line fails. The set of all points $\mathbf{r}$ for which there are no line failures can be defined as an intersection of $2m$ half-spaces (two for every line i), and hence this is a polytope H. Hence we have to compute the probability that the point $\mathbf{r}$, which follows a multivariate normal distribution, is inside H. This corresponds to computing the integral

$$\int_H \exp[-\mathbf{r}^T\Sigma\mathbf{r}]\, d\mathbf{r}, \quad \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2 \ldots, \sigma_n^2), \quad H = \{\mathbf{r} \in \mathbb{R}^n : |\mathbf{f} + VA\mathbf{r}| \leq \mathbf{f}^{max}\}. \tag{68}$$

[15]Note that we assume that the limit of a line is the same in both directions. The approach if direction matters is similar, but does not add much insight and to keep things simple we make this assumption.

[16]The fact that the dimension of the half-space is not necessarily 3 (which it was in the three-node model) does not change this fact.

Note that in contrast to in the three-node case we integrate over $\mathbf{r}$ instead of P_1, P_2, so we integrate over n variables instead of[17] $n-1$. This integral can be computed numerically if we know the network structure and parameters. We can also try to make an approximation of the probability. For this, we start with the fact that

$$\mathbb{P}(\text{the line most likely to fail overloads}) \leq \mathbb{P}(\text{there is an overload}) \leq \sum_i \mathbb{P}(\text{line } i \text{ overloads}). \tag{69}$$

This is simply a result of

$$\mathbb{P}(H_1) \leq \mathbb{P}(H_1 \cup H_2 \cup \ldots \cup H_{2m}) \leq \mathbb{P}(H_1) + \mathbb{P}(H_2) + \ldots + \mathbb{P}(H_{2m}). \tag{70}$$

The reason that this might work is best seen with the following thought experiment. There are $2m$ possible overloads, two for every line (the upper and lower limit of line flow). The likeliness of one of these overflows is determined by the number of standard deviations an overflow is away. As ϵ gets very small (which is the case for rare events), the standard deviations get larger. At some point, the overflows with larger standard deviations become increasingly unlikely as compared to the smallest standard deviation, and hence the left and right side of the inequality will be very close to each other.

Let's make this exact. We are going to compute the right hand side of equation (69). The distances from the boundaries before the perturbation are given by:

$$\mathbf{d}^+ = \mathbf{f}^{max} - \mathbf{f}, \quad \mathbf{d}^- = \mathbf{f}^{max} + \mathbf{f}. \tag{71}$$

For the initial solution to be safe, we need both $\mathbf{d}^+ \geq 0$ and $\mathbf{d}^- \geq 0$. To compute the chance of $\epsilon g_i > d_i^+$ we need to know the standard deviation of g_i. Since $\mathbf{g} = VA\mathbf{r}$ we have that $\mathbf{g}$ follows a normal distribution with mean 0 and variance $\Sigma_{\mathbf{g}}$. Note that for individual lines we can ignore correlation between components of $\mathbf{g}$, so we only need the diagonal elements, let $\tau = \text{diag}(\Sigma_{\mathbf{g}})$. Then we can collect all information in the matrix

$$T = \begin{bmatrix} \mathbf{d}^-/\tau & \mathbf{d}^+/\tau \end{bmatrix}. \tag{72}$$

The division here is component-wise, so $\mathbf{d}^-/\tau = (d_1^-/\tau_1, \ldots, d_m^-/\tau_m)^T$. By defining a new random variable $N \propto \mathcal{N}(0,1)$, we can write

$$\begin{aligned} \mathbb{P}(\text{line } i \text{ overloads}) &= 1 - \mathbb{P}(-T_{i1} \leq \sqrt{\epsilon}N \leq T_{i2}) \\ &= P(\sqrt{\epsilon}N < -T_{i1}) + P(\sqrt{\epsilon}N > T_{i2}) \\ &= P(N > T_{i1}/\sqrt{\epsilon}) + P(N > T_{i2}/\sqrt{\epsilon}). \end{aligned}$$

Note that the smaller T_{ij} is, the more likely the overload is. Letting T_{min} be the smallest element of matrix T, equation (69) becomes

$$\mathbb{P}(N > T_{min}/\sqrt{\epsilon}) \leq \mathbb{P}(\text{there is an overload}) \leq \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq 2}} \mathbb{P}(N > T_{ij}/\sqrt{\epsilon}). \tag{73}$$

[17] Note that changing $\mathbf{r} \to \mathbf{r} + c\mathbb{1}$ with $\mathbb{1}$ the all-ones vector does not change $\mathbf{q}$. So if we change the coordinates to make one coordinate parallel to $\mathbb{1}$, we can integrate out this coordinate and are left with an integral over $n-1$ variables.

And since we have $\mathbb{P}(N > T_{ij}/\sqrt{\epsilon}) \leq \mathbb{P}(N > T_{min}/\sqrt{\epsilon})$ for any i, j, we get

$$\sum_{\substack{1\leq i\leq m\\ 1\leq j\leq 2}} \mathbb{P}(N > T_{ij}/\sqrt{\epsilon}) \leq 2m\mathbb{P}(N > T_{min}/\sqrt{\epsilon}) \tag{74}$$

and hence $\mathbb{P}(\text{there is an overload}) = \Theta\big(\mathbb{P}(N > T_{min}/\sqrt{\epsilon})\big)$. Note that we take m (the number of edges in the graph) to be a constant, since we consider the same graph for all ϵ. Using an error function, we can compute

$$\begin{aligned}
\mathbb{P}(N > T_{min}/\sqrt{\epsilon}) &= \frac{1}{2}\operatorname{erfc}\big(T_{min}/\sqrt{2\epsilon}\big) \\
&\approx \frac{1}{2}\frac{\exp\left[-\frac{T_{min}^2}{2\epsilon}\right]}{\frac{T_{min}}{\sqrt{2\epsilon}}\sqrt{\pi}} \\
&= \frac{\sqrt{\epsilon}}{T_{min}\sqrt{2\pi}}\exp\left[-\frac{T_{min}^2}{2\epsilon}\right].
\end{aligned}$$

We find the final result that

$$\mathbb{P}(\text{there is an overload}) = \Theta\left(\sqrt{\epsilon}\exp\left[-\frac{T_{min}^2}{2\epsilon}\right]\right). \tag{75}$$

Note that this is still the same dependency on ϵ as for the three-cycle!

5 Analysis of a system with variable voltages

In this model, we drop the assumption that all voltages are approximately the same, and we no longer ignore resistance in the network. We do make the assumption that phase angle differences are small (in fact, we set all phase angle differences to zero), and ignore reactive power inputs or outputs. We assume the network to be a line graph. This gives us the *Distflow* power flow model. It is not as artificial as it may seem, as it corresponds to an important practical problem: a street of houses that is supplied from a single point. The problem is now to determine the probability of an overload, given a probability distribution for the power consumption at every house (including possible correlations between houses). Despite the assumption of a specific and relatively simple graph, the problem has not been solved yet for a line of arbitrary length. Earlier research did manage to solve it by ignoring the active power loss[15]. Our goal is to solve it without that assumption.

First, we will describe the model, and derive a recursive relation for the voltages (assuming the power consumption to be known), allowing us to compute the voltages as a function of the power consumption vector $\mathbf{p}$. Then, we suggest a semidefinite program that can be used for exactly finding the *most likely overload configuration* (i.e. the most likely distribution of power consumption that still leads to an overload), as well as an explicit upper and lower bound for the probability of an overload.

5.1 Model description

We consider a line graph consisting of $N+1$ points. The root node, labeled as 0, is the only source that provides power for the entire street. The other nodes are labeled $1, 2, \ldots, N$, where node 1 is the closest to the source. These nodes represent houses that consume electricity[18]. For simplicity we assume all lines to have the same resistance r. We also assume that the voltage phase angles are very small, so that we can take all voltage angle differences to be zero and hence the complex voltage $\widetilde{V}_j$ is a positive real number for all j. See Figure 5. We want to avoid large voltage differences with the source: we require that $\widetilde{V}_j \geq (1-\Delta)\widetilde{V}_0$ for all $1 \leq j \leq n$. Our goal is to find out for which power consumption configurations this constraint is violated. Such a violation is the *rare event* that we are interested in. We assume that if every house consumes the expected amount of power (by the probability distribution), then there is no violation. Then, we assume the variance in the power consumption to become very small, so that an overload becomes increasingly rare, and we are interested in the evolution of this probability.

Define $\widetilde{S}_j = \widetilde{P}_j + i\widetilde{Q}_j$ as the power consumed at node j. Since nodes can only consume active power, $\widetilde{Q}_j = 0$. As we assumed that houses consume power, we have $\widetilde{P}_j > 0$. We let $I_{j-1,j}$ and $\widetilde{S}_{j-1,j}$ be the complex current resp. power flow from node $j-1$ to node j. Let $\overline{I}$ denote

[18]It is possible that houses generate electricity (e.g. with solar panels), but for now we assume that houses consume more power than they generate.

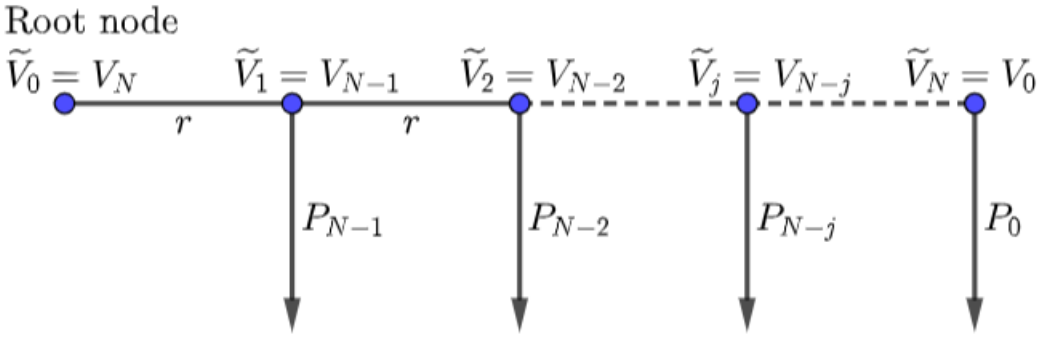

Figure 5: The line model, representing a street that receives its power from a single source. The root node that supplies the entire street is $\widetilde{V}_0$. We can prove a recursive relation that describes $\widetilde{V}_{j-1}$ in terms of $\widetilde{V}_j$ and $\widetilde{V}_{j+1}$. To turn this into a normal recursive relation that starts at $j = 0$ we reverse the order and define $V_j = \widetilde{V}_{N-j}$ and $P_j = \widetilde{P}_{N-j}$.

complex conjugation. Now, we have the following three equations:

$$\begin{aligned} \widetilde{S}_{j-1,j} - r|I_{r-1,r}|^2 &= \widetilde{S}_j + \widetilde{S}_{j,j+1} \quad \text{(power balance)} \\ \widetilde{V}_{j-1} - \widetilde{V}_j &= rI_{j-1,j} \quad \text{(Ohm's law)} \\ \widetilde{S}_{j-1,j} &= \widetilde{V}_{j-1}\overline{I}_{j-1,j} \quad \text{(definition of complex power).} \end{aligned} \tag{76}$$

Combining all of this yields the following recursive relation (see [15] for a proof):

$$\begin{aligned} \widetilde{V}_{N-1} &= \widetilde{V}_N + \frac{r\tilde{P}_N}{V_N} \\ \widetilde{V}_{j-1} &= 2\widetilde{V}_j - \widetilde{V}_{j+1} + \frac{r\widetilde{P}_j}{\widetilde{V}_j}. \end{aligned}$$

Now we can define $V_j = \widetilde{V}_{n-j}$, $P_j = \widetilde{P}_{n-j}$ and $X_j = rP_j$ to obtain the recursion

$$\begin{aligned} V_1 &= V_0 + \frac{X_0}{V_0} \\ V_{j+1} &= 2V_j - V_{j-1} + \frac{X_j}{V_j}. \end{aligned}$$

What this reversing of order physically means is that we now consider the voltage of the leaf node V_0 to be fixed, and the voltage of the source node V_N has to accommodate this. It is important to note that V_N is increasing in X_j for any $0 \leq j \leq N-1$, since extracting more power requires a higher voltage. Previously we had the condition that $\widetilde{V}_j \geq (1-\Delta)\widetilde{V}_0$ for all $1 \leq j \leq n$. This becomes $V_j \geq (1-\Delta)V_N$ for all $0 \leq j \leq n-1$. We can actually reduce this to a single condition:

Lemma 5.1 *V_j is an increasing sequence: we have $V_j > V_{j-1}$ for all $1 \leq j \leq n$.*

Proof. Note that since we assumed that $P_i > 0$, we find that $X_i > 0$ for all i. Hence $V_1 > V_0$. We can now proceed by induction. Assume that $V_j > V_{j-1}$. Then we find

$$\begin{aligned} V_{j+1} &= 2V_j - V_{j-1} + \frac{X_j}{V_j} \\ &= V_j + (V_j - V_{j-1}) + \frac{X_j}{V_j} \\ &> V_j + 0 + 0. \end{aligned}$$

□

This implies that if $V_0 \geq (1-\Delta)V_N$, then $V_j \geq (1-\Delta)V_N$ for any $j \geq 0$, so we only need to check the first. We may just as well work in a per-unit system with $V_0 = 1$. Then, the condition reduces to

$$V_N \leq \frac{1}{1-\Delta} \tag{77}$$

All together, we find the following recurrence relation

$$\begin{aligned} V_0 &= 1 \\ V_1 &= 1 + X_0 \\ V_{j+1} &= 2V_j - V_{j-1} + \frac{X_j}{V_j} \end{aligned} \tag{78}$$

and we are interested in the value V_N. The numbers $X_0, X_1, \ldots, X_{N-1}$ follow some stochastic distribution. We assume that if all X_i are equal to their expected value, then $V_N \leq 1/(1-\Delta)$, and want to approximate the probability of the rare event

$$\mathbb{P}\left(V_N > \frac{1}{1-\Delta}\right). \tag{79}$$

For notational simplicity, we define

$$\Lambda = \frac{1}{1-\Delta} \tag{80}$$

and, similar to in section 4, we define $\mathbf{V} = (V_1, \ldots, V_N)$ and $\mathbf{X} = (X_0, \ldots, X_{N-1})$.

5.2 A semi-definite program for the line

For the line model, we suggest a semi-definite program that can be used for numerical analysis. This is useful because we can program it to work until a desired accuracy is achieved, but the downside is that it does not yield a simple approximation. We are now analyzing the case that $\mathbf{X} = \mu + \sqrt{\epsilon}\mathbf{Y}$, where μ is the expected value of $\mathbf{X}$ and $\sqrt{\epsilon}\mathbf{Y}$ is a perturbation. If $\mathbf{X} = \mu$ then there is no overload, but the perturbation $\sqrt{\epsilon}\mathbf{Y}$ may cause this. As ϵ approaches zero, an overload will be increasingly unlikely: we would like to know how the probability evolves. An approximation in the limit of $\mathbf{X} \to 0$ is made later.

5.2.1 The rate function

To compute overload probabilities, we need the contraction principle, and for that we need a transformation function that transforms $\mathbf{X}$ (of which the probability distribution is known) into $\mathbf{V}$. A first guess may be to define

$$T_1(\mathbf{X}) = V_N(\mathbf{X}). \tag{81}$$

After applying the contraction principle, we would obtain the following rate function for V_N:

$$J(v) = \inf\{I(\mathbf{X}) \mid V_N(\mathbf{X}) = v\}. \tag{82}$$

If we could compute this, then since J is a rate function we know that (by equation (2))

$$\lim_{\epsilon \to 0} \epsilon \log(\mathbb{P}_\epsilon(V_N \geq \Lambda)) = -\inf_{v \geq \Lambda} J(v) \tag{83}$$

which means that we can approximate the probability of an overload directly. The problem is that there is not really a direct way to work J out beyond this implicit definition. So we have to take a different approach, and for this we need a different transformation. We define the transformation $T : \mathbb{R}^N \to \mathbb{R}^N$ as

$$T(\mathbf{X}) = \mathbf{V}(\mathbf{X}) \tag{84}$$

so the output of the transformation is the entire vector $\mathbf{V} = (V_1, V_2, \ldots, V_N)$. This function is continuous and well-defined, so we can apply the contraction principle. We also need a probability density function for $\mathbf{X}$: assume that all X_i are i.i.d. $\propto \mathcal{N}(\mu_j, 1)$ (we will later relax the assumption that all standard deviations are equal to 1, this is just to make the computation simpler). So the rate function for $\mathbf{X}$ is

$$I(\mathbf{X}) = \frac{1}{2} \sum_{i=0}^{j-1} (X_i - \mu_i)^2. \tag{85}$$

Recall that a lower I corresponds to a higher probability. So for the most likely overload configuration, we wish to find the $\mathbf{X}$ with the lowest $I(\mathbf{X})$ that still satisfies $V_N(\mathbf{X}) \geq \Lambda$. Although we are essentially solving an optimization problem now, we will continue to use capital letters for $\mathbf{X}$ and $\mathbf{V}$ to avoid confusion. By the contraction principle, we obtain the rate function

$$I_N(\mathbf{V}) = \inf\{I(\mathbf{X}) \mid T(\mathbf{X}) = \mathbf{V}\} \tag{86}$$

or, written out,

$$I_N(\mathbf{V}) = \inf_{\mathbf{X}: T(\mathbf{X}) = \mathbf{V}} \frac{1}{2} \sum_{i=0}^{j-1} (X_i - \mu_i)^2. \tag{87}$$

There is in fact only one $\mathbf{X}$ that satisfies $T(\mathbf{X}) = \mathbf{V}$. This is because we have the relation

$$V_{j+1} = 2V_j - V_{j-1} + \frac{X_j}{V_j} \tag{88}$$

which can also be written as

$$X_j = V_j V_{j+1} + V_j V_{j-1} - 2V_j^2 \tag{89}$$

which holds for all $j \geq 1$. For $j = 0$ we have $X_0 = V_1 - 1$.
This means that the rate function I_N is equal to

$$I_N(\mathbf{V}) = \frac{1}{2}(V_1 - 1 - \mu_0)^2 + \frac{1}{2}\sum_{j=1}^{N-1}(V_j V_{j+1} + V_j V_{j-1} - 2V_j^2 - \mu_j)^2. \tag{90}$$

It may seem that this rate function is rather pointless, as in equation (86) we take the infimum over just one value. The first reason it is of interest is because of the following theorem.

Theorem 5.2 *In order to find the infimum for the rate function over all possible* $\mathbf{V}$ *satisfying* $V_N \geq \Lambda$*, the constraint* $V_1 \leq V_2 \leq \ldots \leq V_N$ *is redundant assuming that* $\mathbf{V} > 0$*. Equivalently:*

$$\inf_{\mathbf{V}: V_N \geq \Lambda} I_N(\mathbf{V}) = \inf_{\substack{\mathbf{V}: V_N \geq \Lambda \\ V_1 \leq V_2 \leq \ldots \leq V_N}} I_N(\mathbf{v}). \tag{91}$$

Proof. Suppose that we have a configuration of voltages in which the list of inequalities is not satisfied. Then there is an j such that $V_j > V_{j+1}$: take the smallest such j for convenience. Since this j is the smallest, we have $V_j > V_{j-1}$ and hence

$$V_j > V_{j+1} = 2V_j - V_{j-1} + \frac{X_j}{V_j} \geq V_j + \frac{X_j}{V_j}.$$

Since $V_j > 0$ this implies that $X_j < 0$. Since V_N is increasing in $\mathbf{x}$ and there is an overload, there must also be at least one k such that $X_k > \mu_k$.

The point is now that we can find a different configuration with the same V_N, but lower I_N. For this, we choose two δ_i, δ_k such that if we replace X_i by $X_i + \delta_i$ and X_k by $X_k - \delta_k$ to obtain $\mathbf{x}'$, the voltage V_N remains unchanged (this is possible since V_N is increasing in both X_i and X_j). Since $X_i - \mu_i < 0$ we find

$$(X_i - \mu_i)^2 > (X_i + \delta_i - \mu_i)^2 = (X_i' - \mu_i)^2$$

and since $X_j - \mu_j > 0$ we find

$$(X_j - \mu_j)^2 > (X_j - \delta_j - \mu_j)^2 = (X_j' - \mu_j)^2$$

provided that δ_i, δ_j are sufficiently small. Hence, if we replace the initial $\mathbf{V} = T(\mathbf{X})$ by $\mathbf{V}' = T(\mathbf{X}')$ then we find that $I_N(\mathbf{V}') < I_N(\mathbf{V})$, while $V_N = V_N'$. Hence any configuration with $\mathbf{V}$ not satisfying $V_1 \leq V_2 \leq \ldots \leq V_N$ can be improved upon. We conclude that this constraint is redundant for finding the infimum. □

The reason why this result is useful is that (90) gives us a function of $I_N(\mathbf{V})$ in terms of $\mathbf{V}$ only. However, for a network to be realistic, we need to have $V_1 \leq V_2 \leq \ldots \leq V_N$. But now we know that if we want to compute $\inf_{V_N \geq \Lambda} I_N(\mathbf{V})$ we can actually remove this constraint, since any $\mathbf{V}$ not satisfying it can never yield the lowest possible value of $I_N(\mathbf{V})$. Hence, in order to compute the infimum in equation (91), we can solve the following optimization problem:

$$\begin{aligned} &\text{minimize } I_N(\mathbf{V}) \\ &\text{subject to } V_i > 0 \text{ for all } 1 \leq i \leq N-1 \\ &\qquad\qquad V_N \geq \Lambda. \end{aligned}$$

The solution of this optimization problem will provide us with the most likely overload configuration $\mathbf{X}$. Note that this means (by definition) that this optimization problem is equivalent to computing $J(V_N)$ in (82). In order to simplify this problem even further, we can try to get rid of the constraint $V_i > 0$ as well. Note that any configuration with $V_i < 0$ is not physically possible due to the definition of our model, but this does not prevent such 'solutions' from arising mathematically. Since I could not find any concrete proof that a configuration with $V_i < 0$ cannot give a smaller $I_N(\mathbf{V})$ but also did not find any counterexamples, for now, we will just assume that the constraint that $V_i > 0$ is redundant: further research would be needed to either confirm or disprove this conjecture. This leaves us with the problem to minimize $I_N(\mathbf{V})$ given by (90), with the constraint $V_N \geq \Lambda$. It is easy to see that the minimum will occur at a configuration with $V_N = \Lambda$, so we can replace the inequality with an equality if we wish.

5.2.2 The standard approach to make a semidefinite program

We would like to solve the polynomial optimization problem

$$\begin{aligned} &\text{minimize } I_N(\mathbf{V}) \\ &\text{subject to } V_i > 0 \text{ for all } 1 \leq i \leq N-1 \\ &\qquad\qquad V_N = \Lambda. \end{aligned}$$

As stated before, we will assume that the constraint $V_i > 0$ is redundant. The problem is that the objective function $I_N(\mathbf{V})$ has degree 4. For this reason, the problem cannot be solved exactly. So what we do instead is create an equivalent *semidefinite program*.

Definition 5.1 *Let $\mathbb{S}^n$ be the space of symmetric $n \times n$-matrices. Let $A_1, \ldots, A_k, C \in \mathbb{S}^n$, $b_1, \ldots, b_k \in \mathbb{R}$ and define the inner product*

$$\langle A, B \rangle_{\mathbb{S}^n} = \sum_{1 \leq i,j \leq n} A_{ij} B_{ij}.$$

Then the optimization problem

$$\begin{aligned} &\min_{X \in \mathbb{S}^n} \langle C, X \rangle_{\mathbb{S}^n} \\ &\text{subject to } \langle A_k, X \rangle_{\mathbb{S}^n} \leq b_k \\ &\qquad\qquad X \succeq 0 \end{aligned}$$

is a semidefinite program. Here $X \succeq 0$ means that X is positive semidefinite.

Any program satisfying this definition is a semidefinite program, or SDP in short. So we would like to write our optimization problem as an SPD. A method often applied in similar problems is defining

$$W_{i,j} = V_i V_j$$

for every edge $\{i, j\}$, so that we can now write

$$I_N(\mathbf{v}) = \frac{1}{2}(W_{0,1} - 1 - \mu_0)^2 + \frac{1}{2}\sum_{j=1}^{N-1}(W_{j,j+1} + W_{j-1,j} - 2W_{j,j} - \mu_j)^2. \tag{92}$$

It is important to note that the $W_{i,j}$ are not independent now. For every edge $\{i, j\}$, we have the relation

$$W_{i,i}W_{j,j} = W_{i,j}^2 \tag{93}$$

which can also be written as

$$\det \begin{pmatrix} W_{i,i} & W_{i,j} \\ W_{j,i} & W_{j,j} \end{pmatrix} = 0. \tag{94}$$

So now, we get the optimization problem

$$\begin{aligned} &\text{minimize } I_N(\mathbf{W}) \\ &\text{subject to } \det(W(e_{ij})) = 0 \text{ for every edge } e \\ &\qquad W_{0,0} = 0, W_{N,N} = \Lambda. \end{aligned}$$

We would like to relax the condition that the determinant has to be zero, in order to create a semidefinite program. Look at the formula (92) and recall that this formula is in fact $\sum_{i=0}^{N-1}(X_i - \mu_i)^2/2$ written out. In case of the most likely overload, all $X_i - \mu_i$ are positive (or at least non-negative), due to the same argument as in Lemma 5.1. This means that if $W_{j,j+1}$ is higher than the lowest value allowed by the constraints, then we can reduce $I_N(\mathbf{V})$ by replacing $W_{j,j+1}$ with a lower value. Hence we can relax the determinant constraint to

$$\det(W(e_{ij})) \leq 0 \tag{95}$$

without changing the optimal solution.

So we obtain the following optimization problem:

$$\begin{aligned} &\text{minimize } I_N(\mathbf{W}) \\ &\text{subject to } \det(W(e_{ij})) \leq 0 \text{ for every edge } e \\ &\qquad W_{0,0} = 1, W_{N,N} = \Lambda^2 \end{aligned}$$

where the objective function $I_N(\mathbf{W})$ is of degree 2 instead of 4, a considerable improvement. The problem, however, is that it cannot be written as an SDP since the determinant is smaller than zero, not larger. This means that we do not obtain the SDP that we were looking for.

5.2.3 Full linearization

A strategy that does work is the *full linearization*: we assign a new variable to every term of degree 4. To work this out we need some more definitions.

Notation. We take $\mathbb{N} = \{0, 1, 2, \ldots\}$ as the set of non-negative integers. Define $\mathbb{N}_t^n$ as all n-tuples $\alpha \in \mathbb{N}^n$ such that $|\alpha| = \sum_i |\alpha_i| \leq t$. For every $\alpha \in \mathbb{N}_4^{N-1}$, we define

$$y_\alpha := \mathbf{V}^\alpha = V_1^{\alpha_1} V_2^{\alpha_2} \cdots V_{N-1}^{\alpha_{N-1}}. \tag{96}$$

Note that every $\alpha \in \mathbb{N}_4^{N-1}$ corresponds to a monomial V^α of degree at most 4. Now we can write out all the squares in the sum $I_N(\mathbf{V})$ and fill in $V_N = \Lambda$, then every term is a monomial of the form $\mathbf{V}^\alpha$ for some $\alpha \in \mathbb{N}_4^{N-1}$. Let the coefficient of $\mathbf{V}^\alpha$ be f_α, then we can write

$$I_N(\mathbf{V}) = \sum_{\alpha \in \mathbb{N}_4^{N-1}} f_\alpha y_\alpha. \tag{97}$$

With these new variables, $I_N(\mathbf{V})$ becomes a linear function. As constraints, we first have that $\alpha = \emptyset = (0, 0, \ldots, 0)$ yields $y_\alpha = \mathbf{V}^\alpha = 1$. Second, the $\mathbb{N}_2^{N-1} \times \mathbb{N}_2^{N-1}$-matrix[19] M defined by

$$M_{\alpha,\beta} = y_{\alpha+\beta} \tag{98}$$

must be positive semidefinite. Essentially, this constraint guarantees that after solving the SDP and finding y_α, we can find a $\mathbf{V}$.

As an example, we take the case $N = 3$. In this case we start with

$$I_3(\mathbf{V}) = \frac{1}{2}(V_1 - 1 - \mu_0)^2 + \frac{1}{2}(V_1 V_2 + V_1 - 2V_1^2 - \mu_1)^2 + \frac{1}{2}(V_2 \Lambda + V_1 V_2 - 2V_2^2 - \mu_2)^2.$$

[19] This means: the matrix has a row and column for every element of $\mathbb{N}_2^{N-1}$.

This gives us the coefficients

$$
\begin{aligned}
f_{(0,0)} &= \frac{1}{2}(1+\mu_0)^2 + \frac{1}{2}(\mu_1^2+\mu_2^2)\\
f_{(1,0)} &= -1-\mu_0-\mu_1\\
f_{(0,1)} &= -\Lambda\mu_2\\
f_{(2,0)} &= 1+2\mu_1\\
f_{(1,1)} &= -\mu_1-\mu_2\\
f_{(0,2)} &= \frac{1}{2}\Lambda^2+2\mu_2\\
f_{(3,0)} &= -2\\
f_{(2,1)} &= 1\\
f_{(1,2)} &= \Lambda\\
f_{(0,3)} &= -2\Lambda\\
f_{(4,0)} &= 2\\
f_{(3,1)} &= -2\\
f_{(2,2)} &= 1\\
f_{(1,3)} &= -2\\
f_{(0,4)} &= 2
\end{aligned}
$$

and the matrix

$$
M(y) = \begin{pmatrix}
y_{(0,0)} & y_{(1,0)} & y_{(0,1)} & y_{(2,0)} & y_{(1,1)} & y_{(0,2)}\\
y_{(1,0)} & y_{(2,0)} & y_{(1,1)} & y_{(3,0)} & y_{(2,1)} & y_{(1,2)}\\
y_{(0,1)} & y_{(1,1)} & y_{(0,2)} & y_{(2,1)} & y_{(1,2)} & y_{(0,3)}\\
y_{(2,0)} & y_{(3,0)} & y_{(2,1)} & y_{(4,0)} & y_{(3,1)} & y_{(2,2)}\\
y_{(1,1)} & y_{(2,1)} & y_{(1,2)} & y_{(3,1)} & y_{(2,2)} & y_{(1,3)}\\
y_{(0,2)} & y_{(1,2)} & y_{(0,3)} & y_{(2,2)} & y_{(1,3)} & y_{(0,4)}
\end{pmatrix} \tag{99}
$$

that is required to be positive semidefinite. Remember that $y_{(0,0)} = 1$, which we can already fill in. Plugging this linear problem into a computer with some test values for $\mu_0, \mu_1, \mu_2, \Lambda$ yields the same answer as a 'normal' polynomial optimization program, so this strategy is indeed correct. See Appendix 1 for a numerical experiment.

What this means is that we found an SDP to compute the most likely overload configuration. This is useful if we have specified parameters and want to analyze the overload probability, because an SDP can usually be solved more quickly than a general polynomial optimization problem. This means that we can relatively quickly compute $J(\Lambda)$ and then apply equation (83) to compute the evolution of the overload probability.

5.3 Approximating the solutions

We now move on to a different way of looking at the problem. Instead of making an algorithm to numerically solve the problem, we approximate the solution in the limit of small deviations in $\mathbf{X}$. We start by assuming that $\mu = 0$ and that N is very large. Later we will relax these

two assumptions, but in this case we can find a nice and simple relation, which is why we do this first.

5.3.1 A conjecture if $\mu = 0$

If we compute all X_i for the most likely overload case, then all X_i appear to be on a straight line that intersects with the horizontal axis at $i = N$. This leads us to the wild guess

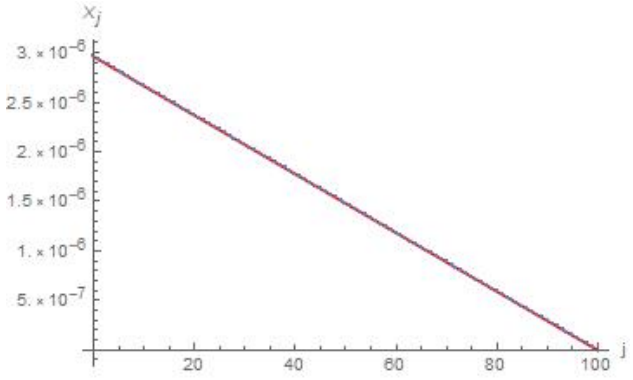

Figure 6: All values X_i in the most likely overload case, for $N = 100$ and $\Lambda = 1.01$. Note that all points are on a line, and that the line intersects with the horizontal axis almost precisely at $i = N = 100$. This gives us a good guess of the values of X_i and hence V_i for the optimal solution.

$$X_i = (N - i)x \tag{100}$$

for a fixed value x. We will prove this guess later: first we will see what this result means for $I_N(\mathbf{V})$. Using the recurrence relation we can now write V_i as a function of x only. After some computation we get the expression

$$V_i = 1 + \frac{1}{6}i(i+1)(3N+1-i)x + \mathcal{O}(x^2) \tag{101}$$

Note that since Λ is close to 1, x is very small, and hence we can ignore higher order terms. We can now compute Λ in terms of x as follows:

$$\begin{aligned}\Lambda &= V_N \\ &= 1 + \frac{1}{6}N(N+1)(2N+1)x.\end{aligned}$$

This implies that

$$x = \frac{6(\Lambda - 1)}{N(N+1)(2N+1)} \approx \frac{3(\Lambda - 1)}{N^3} \tag{102}$$

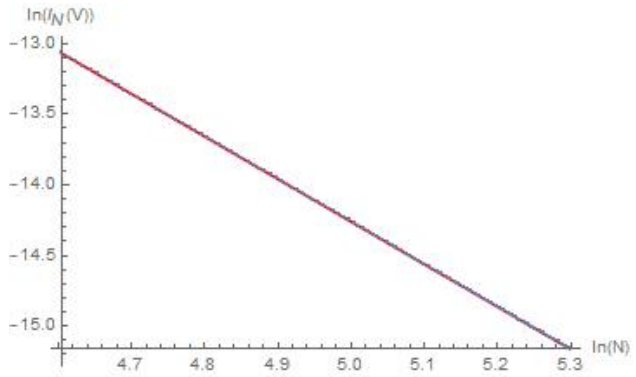

Figure 7: A plot of all $\inf_{\mathbf{V}} I_N(\mathbf{V})$ for $100 \leq N \leq 200$, with $\Lambda = 1.01$. Since we expected a relation of the form $I_N(V) \propto N^c$ for some constant c, we made a log-log plot, which shows all points on a straight line, confirming that expectation. The relation that we numerically find is $\inf_{\mathbf{V}} I_N(\mathbf{V}) \propto N^{-3.011}$, in agreement with our conjecture that it is proportional to N^{-3}.

where the latter approximation is valid if N is large. Assuming that the solution we guessed is indeed optimal we get

$$\begin{aligned}
\inf_{\mathbf{V}} I_N(\mathbf{V}) &= \sum_{i=0}^{N-1} \frac{1}{2} X_i^2 \\
&= \frac{1}{2} x^2 \sum_{i=0}^{N-1} (N-i)^2 \\
&= \frac{18(\Lambda-1)^2}{\big(N(N+1)(2N+1)\big)^2} \cdot \frac{1}{6} N(N+1)(2N+1) \\
&= \frac{3(\Lambda-1)^2}{N(N+1)(2N+1)} \\
&\approx \frac{3(\Lambda-1)^2}{2N^3}.
\end{aligned}$$

This gives us the conjecture that $I_N(\mathbf{V}) \propto N^{-3}$. We can again check this against computational results. In Figure 7, we can clearly see that in a log-log plot the points are indeed on a line, suggesting that we do indeed have a relation of the form $\inf_{\mathbf{V}} I_N(\mathbf{V}) \propto N^c$ for some constant c. And since the line is decreasing, we find $c < 0$, as we also found in our computation.

Recall that we typically use rate functions for probabilities that tend to zero, in which case the rate function becomes very large. So the rate function tending to zero means that we are not really in the terrain of rare events anymore. The reason that this happens is because we set all standard deviations to be 1. So when we want to make some actual estimates, we will choose a more realistic standard deviation instead.

5.3.2 A conjecture if $\mu \neq 0$

To make the conjecture we just found more universal, we now drop the assumptions that $\mu = 0$ and that N is very large. We do still assume μ to be relatively small, so that we can set $\mu_j = \mathcal{O}(\mu)$. The problem is that we no longer have the relation $X_j \approx (N-j)x$ in this case. To solve this, we define $Y_j = X_j - \mu_j$, so that the mean of Y_j is 0 for all $0 \leq j \leq N-1$. If we perform some computations, we now still seem to find that $Y_j = (N-j)y$. Taking this as our guess again, we find that

$$V_i = 1 + \frac{1}{6}i(i+1)(3N-i+1)y + \sum_{j=0}^{i-1}(i-j)\mu_j + \mathcal{O}(y^2, \mu y, \mu^2). \tag{103}$$

This gives us

$$y = \frac{6\left(\Lambda - 1 - \sum_{j=0}^{N-1}(N-j)\mu_j\right)}{N(N+1)(2N+1)}.$$

Hence, we get

$$\begin{aligned}
\inf_{\mathbf{V}} I_N(\mathbf{V}) &= \sum_{i=0}^{N-1} \frac{1}{2}\big((N-i)y\big)^2 \\
&= \frac{1}{2}y^2 \sum_{i=0}^{N-1}(N-i)^2 \\
&= \frac{1}{2}\frac{36\left(\Lambda - 1 - \sum_{j=0}^{N-1}(N-j)\mu_j\right)^2}{(N(N+1)(2N+1))^2} \cdot \frac{N(N+1)(2N+1)}{6} \\
&= \frac{3\left(\Lambda - 1 - \sum_{j=0}^{N-1}(N-j)\mu_j\right)^2}{N(N+1)(2N+1)}.
\end{aligned}$$

If we fill in $\mu = 0$ and assume that N is large, we get $N(N+1)(2N+1) \approx 2N^3$ and we recover our previous formula.

A special case to consider is if all μ_j are equal. In this case, we choose $\mu_j = a/N^2$, because it is known that $X_j = a/N^2$ leads to V_N becoming a constant in the limit $N \to \infty$[15]. In this case, we get the simpler formula

$$\inf_{\mathbf{V}} I_N(\mathbf{V}) = \frac{3\left(\Lambda - 1 - \frac{N+1}{2N}a\right)^2}{N(N+1)(2N+1)}. \tag{104}$$

5.4 Proof of the conjecture

In the case $\mu = 0$ and the limit of small X, we can actually prove that $X_i = (N-i)x$ for some constant x. If $\mu \neq 0$, then our guess turns out to be a little bit off.

5.4.1 The case $\mu = 0$

The proof is rather surprising, as it involves differentiation of $\mathbf{V}$ with respect to $\mathbf{X}$, and the Cauchy-Schwarz inequality. To get started we need this theorem.

Theorem 5.3 *The first order derivatives of* $\mathbf{V}$ *around* $\mathbf{X} = 0$ *are given by*

$$\left.\frac{\partial V_n}{\partial X_i}\right|_{\mathbf{X}=\mathbf{0}} = n - i \tag{105}$$

for all $1 \leq i < n$. *If* $i \geq n$ *then the derivative is 0. The second order derivatives are given by*

$$\left.\frac{\partial^2 V_n}{\partial X_i \partial X_j}\right|_{\mathbf{X}=\mathbf{0}} = -(n-i)(i-j) \tag{106}$$

if $j \leq i < n$. *Again, if* i *or* j *is* $\geq N$ *then this derivative is 0.*

If we make a Taylor expansion, this yields the following expressions:

$$\begin{aligned}
V_0 &= 1 \\
V_1 &= 1 + X_0 \\
V_2 &\approx 1 + 2X_0 + X_1 - X_0X_1 \\
V_3 &\approx 1 + 3X_0 + 2X_1 + X_2 - 2X_0X_1 - 2X_0X_2 - X_1X_2 \\
V_4 &\approx 1 + 4X_0 + 3X_1 + 2X_2 + X_3 - 3X_0X_1 - 4X_0X_2 - 3X_0X_3 - 2X_1X_2 - 2X_1X_3 - X_2X_3.
\end{aligned}$$

Proof. First of all, note that V_n can be written as a function of $X_0, \ldots, X_{n-1}$, and hence any derivative with respect to X_i with $i \geq n$ is obviously 0. Now for the non-trivial derivatives. We can prove the identities by induction on n: first we prove (105). As an induction basis, the identity is (rather trivially) true for $n = 0$ and $n = 1$. Now assume (as induction hypothesis) that (105) is true for $V_0, \ldots, V_n$: we will prove it for V_{n+1}. To save notation we omit the point of differentiation $\mathbf{X} = \mathbf{0}$ from now on. We have

$$V_{n+1} = 2V_n - V_{n-1} + \frac{X_n}{V_n}. \tag{107}$$

If $i \leq n-1$ we find

$$\begin{aligned}
\frac{\partial V_{n+1}}{\partial X_i} &= 2\frac{\partial V_n}{\partial X_i} - \frac{\partial V_{n-1}}{\partial X_i} - \frac{X_n}{V_n^2}\frac{\partial V_n}{\partial X_i} \\
&= 2(n-i) - (n-1-i) - 0 \\
&= n + 1 - i
\end{aligned}$$

which proves the induction step in this case. If $i = n$ we find

$$\begin{aligned}
\frac{\partial V_{n+1}}{\partial X_n} &= 2\frac{\partial V_n}{\partial X_n} - \frac{\partial V_{n-1}}{\partial X_n} - \frac{1}{V_n} \\
&= 2 \cdot 0 - 0 + \frac{1}{V_n}.
\end{aligned}$$

And since $V_n = 1$ if $\mathbf{X} = \mathbf{0}$, this means that the final outcome is 1. This proves the induction step, and hence the proof of (105) is completed. Now on to the second order derivatives. We have that

$$\frac{\partial^2 V_{n+1}}{\partial X_i \partial X_j} = 2\frac{\partial^2 V_n}{\partial X_i \partial X_j} - \frac{\partial^2 V_{n-1}}{\partial X_i \partial X_j} + \frac{\partial^2}{\partial X_i \partial X_j}\frac{X_n}{V_n}. \tag{108}$$

First, suppose that $i < n$. Then this becomes

$$\begin{aligned}
&= -2(n-i)(i-j) + (n-1-i)(i-j) + X_n \frac{\partial^2}{\partial X_i \partial X_j}\left(\frac{1}{V_n}\right) \\
&= -2(n-i)(i-j) + (n-1-i)(i-j) \\
&= -(2n-2i-n+1+i)(i-j) \\
&= -(n+1-i)(i-j).
\end{aligned}$$

The term X_n drops out since we are differentiating in the point $\mathbf{X} = 0$. This proves the induction step in this case. Now suppose that $i = N$ and $j < i$. Then (108) is equal to

$$\begin{aligned}
&= 2 \cdot 0 - 0 + \frac{\partial^2}{\partial X_n \partial X_j}\left(\frac{X_n}{V_n}\right) \\
&= \frac{\partial}{\partial X_j}\left(\frac{1}{V_n}\right) \\
&= -\frac{1}{V_n^2}\frac{\partial V_n}{\partial X_j} \\
&= -1 \cdot (n-j) \\
&= -(n+1-i)(i-j).
\end{aligned}$$

The last inequality holds since $n = i$. In this case the induction step is also proven. Now for the final case that $i = j = n$. In this case (108) is equal to

$$\begin{aligned}
&= 2 \cdot 0 - 0 + \frac{\partial^2}{\partial X_n \partial X_n}\left(\frac{X_n}{V_n}\right) \\
&= \frac{\partial}{\partial X_n}\left(\frac{1}{V_n}\right) \\
&= 0 \\
&= -(n+1-i)(i-j)
\end{aligned}$$

again proving the induction step. This means the induction step is true in all cases and the proof is complete. □

In the case $\mu = 0$, it is sufficient to go until the first order of $\mathbf{X}$ (since we are assuming the limit Λ to be just above 1 and N to be large, hence $\mathbf{X}$ will be very small). Then we find

$$V_N = 1 + \sum_{i=0}^{N-1}(N-i)X_i. \tag{109}$$

We want to minimize $I_N(\mathbf{V})$ over all configurations with $V_N = \Lambda$. This means that we have to solve the optimization problem[20]

$$\begin{aligned} &\text{minimize} \sum_{i=0}^{N-1} X_i^2 \\ &\text{subject to} \sum_{i=0}^{N-1} (N-i) X_i = \Lambda - 1 \end{aligned}$$

where $\Lambda - 1 > 0$ is a constant. There is an elegant way to solve this problem. Since we have the conjecture that that $X_i = (N-i)x$, it makes sense to define

$$Z_i = X_{N-i}/i \tag{110}$$

so that our conjecture becomes that all Z_i are equal. The minimization problem becomes

$$\begin{aligned} &\text{minimize} \sum_{i=1}^{N} i^2 Z_i^2 \\ &\text{subject to} \sum_{i=1}^{N} i^2 Z_i = \Lambda - 1. \end{aligned}$$

This can be solved by applying the Cauchy-Schwarz inequality to the vectors

$$\begin{aligned} \mathbf{A} &= (Z_1, 2Z_2, \ldots, NZ_N) \\ \mathbf{B} &= (1, 2, \ldots, N). \end{aligned}$$

Now Cauchy-Schwarz tells us that

$$\begin{aligned} (\mathbf{A} \cdot \mathbf{A})(\mathbf{B} \cdot \mathbf{B}) &\geq (\mathbf{A} \cdot \mathbf{B})^2 \\ \left(\sum_{i=1}^{N} i^2 Z_i^2 \right) \left(\sum_{i=1}^{N} i^2 \right) &\geq \left(\sum_{i=1}^{N} i^2 Z_i \right)^2 = (\Lambda - 1)^2 \\ \sum_{i=1}^{N} i^2 Z_i^2 &\geq \frac{(\Lambda - 1)^2}{\sum_{i=1}^{N} i^2} \end{aligned}$$

which gives us a lower bound for $\sum_{i=1}^{N} i^2 Z_i^2$ that is only attained if the vectors $\mathbf{A}$ and $\mathbf{B}$ are parallel. Now $\mathbf{A}$ and $\mathbf{B}$ are parallel exactly if $Z_1 = Z_2 = \ldots = Z_N$. So the solution to the optimization problem can be found by setting $Z_i = x$ for all i. Hence we have $X_i = (N-i)x$, as was claimed.

5.4.2 What if $\mu \neq 0$?

We assumed that $\mu = 0$ in the entire proof, so the question is what happens if $\mu \neq 0$. For this, we ignore any terms of order $\mathcal{O}(\mu^2)$ or higher, to keep things simple. The first order

[20] I left out the factor $\frac{1}{2}$ in $I_N(\mathbf{V})$ since it does not change the solution for $\mathbf{X}$.

relation in theorem 5.3 becomes

$$\left.\frac{\partial V_n}{\partial X_i}\right|_{\mathbf{X}=\mu} = (n-i) + \sum_{0\le j\le n-1} \mu_i \left.\frac{\partial^2 V_n}{\partial X_i \partial X_j}\right|_{\mathbf{X}=0} \tag{111}$$

which can also be written as

$$\left.\frac{\partial V_n}{\partial X_i}\right|_{\mathbf{X}=\mu} = (n-i) - \sum_{0\le i<j} (n-j)(j-i)\mu_i - \sum_{j<i\le n-1} (n-i)(i-j)\mu_i. \tag{112}$$

We can prove this by taking the second-order Taylor expansion around $\mathbf{X} = 0$, differentiating with respect to X_i, and then fill in $\mathbf{X} = \mu$. This is also where the second-order derivatives come in. Introducing the notation

$$c_{ni} = \sum_{0\le j\le n-1} \mu_i \left.\frac{\partial^2 V_n}{\partial X_i \partial X_j}\right|_{\mathbf{X}=0} \tag{113}$$

the optimization problem becomes

$$\begin{aligned} &\text{minimize} \sum_{i=0}^{N-1} Y_i^2 \\ &\text{subject to} \sum_{i=0}^{N-1} (N-i-c_{Ni}) Y_i = \Lambda - 1. \end{aligned}$$

And with the same trick as before with the Cauchy-Schwarz inequality, we find that $Y_i = (N-i-c_{Ni})y$. Note that this is not the $(N-i)y$ that we guessed! Does this mean we have to redo the computation? Well, c_{Ni} will usually be a lot smaller than $N-i$, due to the μ in c_{Ni}. However we cannot just eliminate this term as being of order $\mathcal{O}(\mu)$, since the factor $(N-i)(i-j)$ can compensate for this. This factor is $\mathcal{O}(N^2)$, and since there are N such terms (for every $0 \le i \le N-1$ it is $\mathcal{O}(N^3)$. For example, if $\mu_i = a/N^2$ for all i, then

$$c_{Ni} = \frac{a(n-i)}{6n^2}\big(3i(i+1) + (n-i-1)(n-i+1)\big) = \mathcal{O}(N)$$

unless we fix i to be small. We will see that c_{Ni} is proportional to N with a very small factor, so leaving it out is certainly not a bad approximation, but this does mean that our approximation does not become exact in the limit $N \to \infty$. Unfortunately the expressions become very long and complicated if we do take this into account, and for that reason I was not able to do it for this project. Further research would be needed to work out this case. In the rest of the thesis we will just approximate the most likely overflow configuration with $Y_i = (N-i)y$ in the case $\mu \neq 0$.

Another question that we haven't considered so far is what happens if not all nodes have the same standard deviation. The derivatives will not change, so in the optimization problem the constraint will be the same. What will change is the function $I_N(\mathbf{V})$: here every term

X_i^2 is replaced by X_i^2/σ_i^2. So we get a very similar optimization problem, that we can solve in a similar way. Let's do this: we want to solve

$$\begin{aligned} &\text{minimize} \sum_{i=0}^{N-1} X_i^2/\sigma_i^2 \\ &\text{subject to} \sum_{i=0}^{N-1} (N-i)X_i = \Lambda - 1. \end{aligned}$$

This time, we can make the substitution $X_i = \sigma_i^2(N-i)Z_i$, to obtain the new system

$$\begin{aligned} &\text{minimize} \sum_{i=0}^{N-1} (N-i)^2 Z_i^2 \sigma_i^2 \\ &\text{subject to} \sum_{i=0}^{N-1} (N-i)^2 Z_i \sigma_i^2 = \Lambda - 1. \end{aligned}$$

And by again applying the Cauchy-Schwarz inequality, we find that all Z_i need to be equal. Hence $X_i = \sigma_i^2(N-i)x$ with x some constant in this case. So the derivation will be proportional to the variance σ^2. Note that this strategy relies on there being no correlation between components of $\mathbf{X}$: if there was then this strategy does not work anymore. In that case, we can solve it by find a vector $\mathbf{Z}$ that makes the covariance matrix Σ diagonal, and then apply this strategy.

5.5 A lower and upper bound

For the case $\mu = 0$, we can actually make a lower and upper bound for $\inf_{\mathbf{V}} I_N(\mathbf{V})$, that converge to each other in the case $\Lambda \downarrow 1$.

The upper bound. There are two ways to create an upper bound. The first one is by finding the solution that satisfies $X_i = (N-i)x$, such that $V_N = \Lambda$. To do this, we define

$$f_N(x) = V_N(X_0 = Nx, X_1 = (N-1)x, \ldots, X_{N-1} = x). \tag{114}$$

Now $f_N(x)$ is strictly increasing in terms of x. This means that there is a unique x which satisfies $f_N(x) = \Lambda$. We can find this x using numerical methods (such as Newton-Rhapson's algorithm). Next we compute $I_N(\mathbf{V})$. This bound is valid since the infimum of $I_N(\mathbf{V}$ is at most the value attained for this particular $\mathbf{V}$. Although this method by solving for 1 unknown is computationally much more friendly than the exact method of optimizing for N unknowns, it is unfortunately still not an explicit formula.

Another approach is the following. Rather than choosing values for $\mathbf{X}$, we choose $V_i = 1 + \frac{1}{6}i(i+1)(3N+1-i)x$ and define our $\mathbf{X}$ to match that. Since $V_N = \Lambda$, we have that

$$x = \frac{6(\Lambda - 1)}{N(N+1)(2N+1)}. \tag{115}$$

We have $X_i = V_iV_{i+1} + V_iV_{i-1} - 2V_i^2$, which becomes

$$X_i = (N-i)x\big[1 + \frac{1}{6}i(i+1)(3N+1-i)x\big]. \tag{116}$$

Plugging all of this in, we find that

$$\inf_{\mathbf{V}} I_N(\mathbf{V}) \leq \sum_{i=0}^{N-1} \frac{1}{2} X_i^2.$$

This is an extremely long expression, here it is posted in two lines:

$$U := \frac{3(\Lambda-1)}{N(N+1)(2N+1)}\Bigg[1 + \frac{(N-1)(N+2)}{(2N+1)^2}(\Lambda-1) + \frac{(N-1)(N+2)(17N^4+34N^3+35N^2+18N+36)}{105N^2(N+1)^2(2N+1)^2}\Bigg]. \tag{117}$$

Note that in the limit $\Lambda = 1$ this reduces to the result we found in section 5.3.1. This expression is our upper bound.

Before we continue with the lower bound, we need the following lemma.

Theorem 5.4 *Suppose that for a certain $\mathbf{X} \geq 0$ leads to $V_N(\mathbf{X})$ as voltage of the last node. Then*

$$V_N(\mathbf{X}) \leq 1 + \sum_{i=0}^{N-1} (N-i)X_i. \tag{118}$$

Proof. This theorem is inspired by the Taylor expansions: since all second order derivatives are negative this statement seems very likely. Still we now require a more solid proof. First, we prove a small lemma:

$$V_n - V_{n-1} \leq \sum_{i=0}^{n-1} X_i. \tag{119}$$

This can be proved using induction. For $n = 1$, the statement is clearly true (it is in fact an equality). Suppose now that it is true for a certain n. Then we have

$$V_{n+1} - V_n = V_n - V_{n-1} + \frac{X_n}{V_n} \leq \sum_{i=0}^{n-1} X_i + \frac{X_n}{V_n} \leq \sum_{i=0}^{n-1} X_i + X_n = \sum_{i=0}^{n} X_i \tag{120}$$

which proves the induction step and thereby this small lemma. We can now prove the

theorem using a telescopic sum:

$$\begin{aligned} V_N &= V_0 + \sum_{j=0}^{N-1}(V_{N-j} - V_{N-j-1}) \\ &\leq 1 + \sum_{j=0}^{N-1}\sum_{i=0}^{N-j-1} X_i \\ &= 1 + \sum_{i=0}^{N-1}(N-i)X_i. \end{aligned}$$

□

The lower bound. We prove that if we take

$$x = \frac{6(\Lambda - 1)}{N(N+1)(2N+1)} \tag{121}$$

then

$$L := \frac{1}{2}\sum_{i=0}^{N-1}((N-i)x)^2 \leq \inf_{\mathbf{V}:V_N=\Lambda} I_N(\mathbf{V}). \tag{122}$$

We can do this as follows. Define $\mathbf{X}^0 = (Nx, (N-1)x, \ldots, x)$. Suppose that (122) does not hold, then there is a $\mathbf{V}^1$ with $I_N(\mathbf{V}) < L$. Let $\mathbf{X}^1$ be the corresponding vector of deviations. Using our theorem 5.4, we have

$$V_N(\mathbf{X}^1) \leq 1 + \sum_{i=0}^{N-1}(N-i)X_i^1. \tag{123}$$

Also by definition we have that $V_N(\mathbf{X}^1) = \Lambda$, and hence

$$\Lambda \leq 1 + \sum_{i=0}^{N-1}(N-i)X_i^1. \tag{124}$$

This implies that there exists a $\lambda \in [0,1]$ such that

$$\Lambda = 1 + \lambda\sum_{i=0}^{N-1}(N-i)X_i^1. \tag{125}$$

But then $\lambda\mathbf{X}^1$ satisfies the following two equations:

$$\sum_{i=0}^{N-1}(N-i)(\lambda X_i^1) = \Lambda - 1$$

$$\sum_{i=0}^{N-1}(\lambda X_i^1)^2 < \sum_{i=0}^{N-1}(X_i^0)^2$$

which contradicts the fact that $\mathbf{X}^0$ is known to be the the optimal solution to this optimization problem. Hence, our assumption that such an $\mathbf{X}^1$ exists is false and hence the inequality in (122) holds.

All together, we have the following definite bounds:

$$\frac{3(\Lambda-1)^2}{N(N+1)(2N+1)} \leq \inf_{\mathbf{V}} I_N(\mathbf{V}) \leq \frac{3(\Lambda-1)}{N(N+1)(2N+1)}\left[1+\frac{(N-1)(N+2)}{(2N+1)^2}(\Lambda-1)\right.$$
$$\left.+\frac{(N-1)(N+2)(17N^4+34N^3+35N^2+18N+36)}{105N^2(N+1)^2(2N+1)^2}\right]. \quad (126)$$

In the case $\mu \neq 0$, these same strategies can both be applied to compute bounds as well. However, in this case the formulas will become even longer than they are now, so this will not be done here.

5.6 Accuracy of the approximations

We have found an approximation in the case $\mu = 0$ that is exact if $\mathbf{X} \to 0$, and an approximation in the case $\mu \neq 0$. We will now analyze the accuracy of these approximations, and of the lower and upper bound that we found. This is also the point where we will get rid of the assumption that all standard deviations are 1, since that would lead to ridiculously low values of $I_N(\mathbf{V})$. Instead, note that if the standard deviation is replaced by σ, then (85) becomes

$$I(\mathbf{X}) = \frac{1}{2}\sum_{i=0}^{j-1}\frac{(X_i-\mu_i)^2}{\sigma^2}. \quad (127)$$

5.6.1 The bounds if $\mu = 0$

We have found two explicit bounds in (126), so let's see how close they are to the actual value of $\inf_{\mathbf{V}:V_N=\Lambda} I_N(\mathbf{V})$. As a sanity check, we can also see if the exact value is always between the bounds. For σ we choose $(\Lambda-1)/10^4$, to hopefully make an overload a rare event. Note that this is not relevant for how well the bounds work (it simply adds a factor to I) but it does make for results that are a lot more sensible.

$\inf_{\mathbf{V}:V_N=\Lambda} I_N(\mathbf{V})$	lower bound	exact value	upper bound
$N=10, \Lambda=1.1, \sigma=10^{-5}$	$1.2987 \cdot 10^5$	$1.33017 \cdot 10^5$	$1.33103 \cdot 10^5$
$N=100, \Lambda=1.1, \sigma=10^{-5}$	$1.47776 \cdot 10^2$	$1.51433 \cdot 10^2$	$1.51529 \cdot 10^2$
$N=10, \Lambda=1.01, \sigma=10^{-6}$	$1.2987 \cdot 10^5$	$1.30188 \cdot 10^5$	$1.30189 \cdot 10^5$
$N=100, \Lambda=1.01, \sigma=10^{-6}$	$1.47776 \cdot 10^2$	$1.48145 \cdot 10^2$	$1.48146 \cdot 10^2$

The sanity check is successful, since the exact value is between the bounds in all cases. What we can clearly see is that the upper bound is extremely tight, especially for the limiting case

$N \to \infty$ and $\Lambda \downarrow 1$. The much simpler lower bound is not bad at all, and also gets closer as we approach the limits.

5.6.2 In the case $\mu \neq 0$

The strategies that we have developed can also be used for the case $\mu \neq 0$, but the expressions become even longer making the analysis more difficult. So for now, we will be comparing the actual results to the results we got from the guessed relation of $Y_i = (N-i)y$. We will be taking $\mu_i = a/N^2$ for the average power output, since this leads to a constant V_N if[15] $\mathbf{X} = \mu$. Again, we take $\sigma = a/10^4$.

$\inf_{\mathbf{V}:V_N=\Lambda} I_N(\mathbf{V})$	exact value	estimate
$N = 10, \Lambda = 1.1, a = 0.1, \sigma = 10^{-5}$	$2.80178 \cdot 10^4$	$2.62987 \cdot 10^4$
$N = 100, \Lambda = 1.1, a = 0.1, \sigma = 10^{-5}$	$3.83711 \cdot 10^1$	$3.62088 \cdot 10^1$
$N = 10, \Lambda = 1.01, a = 0.01, \sigma = 10^{-6}$	$2.64707 \cdot 10^4$	$2.62987 \cdot 10^4$
$N = 100, \Lambda = 1.01, a = 0.01, \sigma = 10^{-6}$	$3.64254 \cdot 10^1$	$3.62088 \cdot 10^1$

The approximation is not as accurate as the upper bound we found in the case $\mu = 0$, but still appears to give descent results. Note that while decreasing the values of a and Λ does yield a considerably better approximation, increasing N has hardly any effect on improving the relative error. This makes sense, as we already saw that the error in the relation $Y_i = (N-i)y$ is proportional to μ, hence to a. Despite the fact that the approximation can certainly be improved upon, it is still not bad: it always has the right order of magnitude, and the relative error is less than 10% in all cases, even less than 1% in the bottom two cases. It is however likely that this approach is not very accurate if Λ became much larger, since the approximation $Y_i = (N-i)y$ we used will become much less accurate. Luckily all practical systems satisfy $\Delta < 0.5$, hence $\Lambda < 2$, so the order of magnitude is still likely to be correct.

5.7 Making $I_N(\mathbf{V})$ converge

In the line model as we studied it, we have the relation $I_N(\mathbf{V}) \propto N^{-3}$. However, from a practical standpoint we would like to see $I_N(\mathbf{V})$ converging to a fixed value as $N \to \infty$, so let's see if we can make that happen in a realistic system. One idea would be to change the resistance values. Since $X_i = rp_i$, changing the resistance will change the variance in X_i. So if r_i is changed into r_i/N, then the variance would change from σ^2 into σ^2/N^2. So in that case, the rate function would become

$$I_{N,r/N}(\mathbf{V}) = \sum_{i=0}^{N-1} \frac{1}{2}(NX_i)^2 \tag{128}$$

and hence we have $I_{N,r/N}(\mathbf{V}) \propto 1/N$ in this case.

The most likely overload configuration now involves that $X_0 > X_1 > \ldots > X_N$, which means that some nodes have a larger impact than others. To counter this, we can replace

the covariance matrix Σ of $\mathbf{X}$ by $\text{diag}(1/N, 1/(N-1), \ldots, 1)$ (i.e. not every node has the same variance). Then, the rate function would become $I_N(\mathbf{V}) = \sum_{i=0}^{N-1} \frac{1}{2}(N-i)X_i^2$ and the optimization problem would become

$$\begin{aligned} &\text{minimize} \sum_{i=0}^{N-1} (N-i)X_i^2 \\ &\text{subject to} \sum_{i=0}^{N-1} (N-i)X_i = \Lambda - 1 \end{aligned}$$

By the Cauchy-Schwarz inequality the optimal solution is attained when $X_0 = X_1 = \ldots = X_{N-1}$. This means that $X_i = \frac{2(\Lambda-1)}{N(N+1)}$, and hence

$$I_N(\mathbf{V}) = \frac{2(\Lambda-1)^2}{N(N+1)} \tag{129}$$

which is proportional to $1/N^2$. So if we combine these two ideas and take the covariance matrix $\Sigma = \frac{1}{N^2}\text{diag}(1/N, 1/(N-1), \ldots, 1)$ (corresponding to all resistances divided by N and nodes further away having lower variance), then $I_N(\mathbf{V})$ would converge to the constant $2(\Lambda-1)^2$ as $N \to \infty$. Although this does meet our goal, it is rather artificial. So although it is possible to make $I_N(\mathbf{V})$ constant as $N \to \infty$, it is difficult to find non-artificial models in which this happens.

6 Circular model

For this model, we make the same assumptions as in the line graph, except that we now take a cycle graph rather than a line. Intuitively, this corresponds to building a line from the last house on the street back to the supply point, with the same resistance. There are $N+1$ nodes labeled 0 through N. Node 0 is the source node of our network. We work in a per-unit system where node 0 has voltage $V_0 = 1$ and the power output X_0 at this node is exactly such that there is power balance in the network. We require that $V_i \geq 1 - \Delta$ for all $1 \leq i \leq N$. Since power is injected at node 0, we have $X_0 < 0$. Note that X_0 is not simply $-(X_1 + \ldots + X_N)$, since there is also power loss through resistance in the network. Since we don't know X_0 on beforehand, we define $\mathbf{X} = (X_1, \ldots, X_N)$. The power output X_i at every node other than 0 is known, and the problem is to now determine the values V_i. The equation

$$V_{i+1} = 2V_i - V_{i-1} + \frac{X_i}{V_i} \tag{130}$$

that we had for the line model still applies here, and applies to every node in the cycle, with the periodic boundary conditions $V_{-1} = V_N$ and $V_{N+1} = V_0$. For $i = 0$ it is rather useless since we don't know X_0. So we have a total of N equations, with N unknowns $V_1, V_2, \ldots, V_N$. This means that the system should be solvable. Written out, it is

$$\begin{aligned}
V_2 &= 2V_1 - 1 + \frac{X_1}{V_1} \\
V_3 &= 2V_2 - V_1 + \frac{X_2}{V_2} \\
V_4 &= 2V_3 - V_2 + \frac{X_3}{V_3} \\
&\ldots \\
1 &= 2V_N - V_{N-1} + \frac{X_N}{V_N}
\end{aligned}$$

Note that I filled in that $V_0 = V_{N+1} = 1$. Since the system is not linear we have very little hope of solving it exactly. If we were to solve it numerically, then we could apply the following plan:

1. The first equation states V_2 as a function of V_1
2. Filling this in, the second equation states V_3 as a function of V_1
3. Filling in V_2 and V_3, the third equation states V_4 as a function of V_1
4. Continue this until the semi-last equation, which now states V_N as a function of V_1
5. Now the final equation can be written in terms of V_1 only

It is easily seen by induction that the final equation becomes a rational function of V_1, which can then be written as a polynomial equation $q(V_1) = 0$, with q a polynomial of degree

2^N. This means that we get 2^N different solutions for V_1. However, as it turns out, only of these solutions is physically relevant, for the following reason. It is natural to require that if $\mathbf{X} = 0$, then we have $\mathbf{V} = \mathbb{1}$. However, if we compute all 2^N zeroes of q at $\mathbf{X} = 0$ with a computer, we find that the solution $V_1 = 1$ (corresponding to $\mathbf{V} = \mathbb{1}$) is a simple (i.e. non-multiple) root. All other solutions of $q(V_1) = 0$ in fact lead to at least one V_i being equal to zero. These 'solutions' exist due to turning the rational function into a polynomial when computing: they cannot exist in the original system of equations, since we divide by V_i there. However, when we perturbate $\mathbf{X}$ by a little, these 'solutions' also get perturbed, so that V_i is no longer zero, and then they may suddenly be physically possible solutions. Despite this, they are not the solutions we are looking for, since the solutions we look for require $V_i \geq \Delta$ for any i. Since we assume the system to start at $\mathbf{V} = \mathbb{1}$, it is highly unlikely that we accidentally end up in one of these unwanted solutions.

Now that we have a way of exactly computing $\mathbf{X}$, we would also like to have an approximation. Like in the linear mode, we can linearize around $\mathbf{X} = 0$. We find the derivatives

$$\left.\frac{\partial V_i}{\partial X_j}\right|_{\mathbf{X}=0} = \begin{cases} -\frac{(N+1-i)j}{N+1} & \text{if } i \geq j \\ -\frac{(N+1-j)i}{N+1} & \text{otherwise.} \end{cases} \tag{131}$$

This way, we can find the Jacobian matrix. Since all elements are negative, we define this matrix as $-M$. As an example, for $N = 5$ we find that

$$M = \frac{1}{6}\begin{pmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \tag{132}$$

and then we have

$$\mathbf{V} \approx \mathbb{1} - M\mathbf{X} \tag{133}$$

as the approximate relation. From now on, we will be using this approximation to evaluate rare event probabilities and find the most likely overload case.

6.1 Evaluation of rare event probabilities

We now want to find the most likely overload case $\mathbf{X}$ (i.e. the $\mathbf{X}$ with the lowest $I_N(\mathbf{X})$ such that $V_i(\mathbf{X}) \leq 1 - \Delta$ for some $1 \leq i \leq N$). Assuming a Gaussian probability distribution, the rate function is again equal to $I_N(\mathbf{X}) = \mathbf{X}^T\mathbf{X}/2$ (the fact that we use a different network does not change the probability distribution for $\mathbf{X}$). So to find the most likely overload case, we need to solve the optimization problem

$$\begin{aligned} &\text{minimize } \mathbf{X}^T\mathbf{X}/2 \\ &\text{subject to } \mathbb{1} - M\mathbf{X} \not> (1-\Delta)\mathbb{1}. \end{aligned}$$

The $\not\succ$ can be handled by splitting this into the N separate optimization problems

$$\begin{aligned} &\text{minimize } \mathbf{X}^T\mathbf{X}/2 \\ &\text{subject to } \mathbf{M}_i^T\mathbf{X} \geq \Delta \end{aligned}$$

where $\mathbf{M}_i^T$ is the i-th row of M. Now all of these problems yield a minimum for $\mathbf{X}^T\mathbf{X}/2$, and if we pick the lowest of these N numbers, we find the overall minimum.

Using the Cauchy-Schwarz inequality, we can prove that the optimal solution for the optimization problem i is attained when $\mathbf{X}$ is parallel to $\mathbf{M}_i$. So we write $\mathbf{X} = \lambda\mathbf{M}_i$. The optimum is attained when the constraint inequality becomes an equality, yielding

$$\lambda = \frac{\Delta}{||\mathbf{M}_i||^2} \tag{134}$$

and hence we find that the minimum to optimization problem i is equal to

$$\mathbf{X}^T\mathbf{X}/2 = \lambda^2\mathbf{M}_i^T\mathbf{M}_i = \frac{\Delta^2}{2||\mathbf{M}_i||^2}. \tag{135}$$

So to solve the optimization problem we need to find the i with the largest $||\mathbf{M}_i||$. Looking at the matrix in (132), we expect this i to be in the middle between 1 and N, which would yield $i = (N+1)/2$. Since we have the exact elements of the matrix, we can compute $||\mathbf{M}_i||^2$ for every i: it is equal to

$$||\mathbf{M}_i||^2 = \frac{i(N+1-i)(2i(N+1-i)+1)}{6(N+1)}. \tag{136}$$

In order to find the maximum we can differentiate this with respect to i, which yields

$$\frac{\partial}{\partial i}||\mathbf{M}_i||^2 = \frac{1}{6(N+1)}(N+1-2i)(4i(N+1-i)+1) \tag{137}$$

which is zero if $i = (N+1)/2$, or if $4i(N+1-i)+1 = 0$ but the solutions of the latter have either $i < 0$ or $i > N+1$, which eliminates them. We find that $i = (N+1)/2$ is a maximum, and the other two zeroes (outside $[0, N+1]$) are minima, and hence $i = (N+1)/2$ is the maximum that we are looking for. However, i needs to be an integer, and while $(N+1)/2$ is an integer for odd N, it is not for even N. So for even N the optima is attained at either $i = N/2$ or $i = N/2+1$. As it turns out, both of these i yield the same $\mathbf{X}^T\mathbf{X}/2$.

Assuming that N is odd and hence $(N+1)/2$ is an integer, we find that

$$||\mathbf{M}_{(N+1)/2}||^2 = \frac{1}{48}(N+1)(N^2+2N+3) \tag{138}$$

and hence

$$\mathbf{X}^T\mathbf{X}/2 = \frac{24\Delta^2}{(N+1)(N^2+2N+3)}. \tag{139}$$

Like in the line model, this is proportional to N^{-3}. In the line model, it was proportional to

$$\begin{aligned}(\Lambda-1)^2 &= \left(\frac{1}{1-\Delta}-1\right)^2 \\ &= \left(\frac{\Delta}{1-\Delta}\right)^2 \\ &\approx \Delta^2\end{aligned}$$

so assuming that Δ is small the proportionality to Δ is the same as well. Even though the proportionality relations are the same, the constant is different. Instead of a factor 3/2 we obtained in section 5.3.1, we now have a factor 24, which is 16 times higher. Effectively, this means that

$$\inf_{\mathbf{V}} I_{N,circ}(\mathbf{V}) \simeq 16 \inf_{\mathbf{V}} I_{N,line}(\mathbf{V}). \tag{140}$$

If we assume that the covariance matrix of $\mathbf{X}$ is equal to ϵI with I the identity matrix, the probability of an overload event E evolves as

$$\mathbb{P}_\epsilon(E) = \exp[-I_N(\mathbf{V})/\epsilon]. \tag{141}$$

Hence, we have that

$$\mathbb{P}_{\epsilon,circ}(E) \simeq (\mathbb{P}_{\epsilon,line}(E))^{16} \tag{142}$$

meaning that the chance of an overload for the circle model is far smaller than for the linear model. From a practical standpoint, it means that adding an additional power line to create a cycle makes overloads far less likely. This might be a cost-effective measure to prevent overloads.

6.2 The case $\mu \neq 0$

If $\mu \neq 0$, we can approach the problem as follows. The optimization problem for row i of the matrix becomes

$$\begin{aligned}&\text{minimize } (\mathbf{X}-\mu)^2 \\ &\text{subject to } \mathbf{M}_i^T\mathbf{X} \geq \Delta.\end{aligned}$$

Set $\mathbf{X} = \mu + \mathbf{Y}$, with $\mathbf{Y}$ averaging zero. Then this problem translates into

$$\begin{aligned}&\text{minimize } \mathbf{Y}^2 \\ &\text{subject to } \mathbf{M}_i^T(\mathbf{Y}-\mu) \geq \Delta\end{aligned}$$

or, equivalently,

$$\begin{aligned}&\text{minimize } \mathbf{Y}^2 \\ &\text{subject to } \mathbf{M}_i^T\mathbf{Y} \geq \Delta - \mathbf{M}_i^T\mu.\end{aligned}$$

Note that by applying the Cauchy-Schwarz inequality we now still find $\mathbf{Y}$ to be parallel to $\mathbf{M}_i$. We find that $\mathbf{Y} = \lambda \mathbf{M}_i$, with

$$\lambda = \frac{\Delta - \mathbf{M}_i^T \mu}{||\mathbf{M}_i||^2}. \tag{143}$$

We assume that $\lambda > 0$ (since otherwise an overload is not a rare event), hence we require that $\Delta > \mathbf{M}_i^T \mu$ for every $1 \leq i \leq N$. We now have

$$\begin{aligned} \mathbf{Y}^T \mathbf{Y}/2 &= \lambda^2 \mathbf{M}_i^T \mathbf{M}/2 \\ &= \frac{(\Delta - \mathbf{M}_i^T \mu)^2}{||\mathbf{M}_i||^4} \cdot \frac{||\mathbf{M}_i||^2}{2} \\ &= \frac{(\Delta - \mathbf{M}_i^T \mu)^2}{2||\mathbf{M}_i||^2}. \end{aligned}$$

We need to find the i for which this is minimal. Note that this is not necessarily the node furthest away from the origin anymore! For example, if we take $\mu = (a, 0, 0, \ldots, 0)$, it is possible that node 1 will overload for the smallest a, and hence $\mathbf{X}^T \mathbf{X}/2$ would be the lowest for $i = 1$. This means that we cannot know i without further information about μ. So if μ is not specified then this is as far as we can go.

6.2.1 Assuming that μ is constant on the circle

If we assume that μ is the same everywhere, the lowest $\mathbf{X}^T \mathbf{X}/2$ is attained when i is closest to $(N+1)/2$. This is easy to see, since $\Delta - \mathbf{M}_i^T \mu$ is decreasing in $\mathbf{M}_i$, and so is $1/||\mathbf{M}_i||^2$. So the highest $\mathbf{M}_i$ yields the lowest $\mathbf{X}^T \mathbf{X}/2$. If we take $\mu = c/(N+1)^2$ for every node, then we find

$$\mathbf{M}_{(N+1)/2}^T \mu = \frac{c}{(N+1)^2} \sum_{j=1}^{N} M_{ij} = \frac{c}{(N+1)^2} \cdot \frac{1}{2} \left(\frac{N+1}{2} \right)^2 = \frac{c}{8}. \tag{144}$$

Hence, we have that

$$V_{(N+1)/2} \approx 1 - \frac{c}{8}. \tag{145}$$

Since this approximation becomes increasingly accurate if $N \to \infty$, we find that $\lim_{N \to \infty} \inf_{1 \leq i \leq N} V_i$ is a constant, just like in the line model.

The rate function for the most likely overload case can be computed by plugging this in:

$$\mathbf{Y}^T \mathbf{Y}/2 = \frac{(\Delta - \mathbf{M}_i^T \mu)^2}{2||\mathbf{M}_i||^2} = \frac{24(\Delta - c/8)^2}{(N+1)(N^2 + 2N + 3)}. \tag{146}$$

7 Conclusion and outlook

7.1 Conclusion

In this thesis, we analyzed AC power flow networks without full linearization. We first studied a three-node cycle without the small angle assumption in depth. The solutions are the roots of a polynomial of degree 6. There are multiple realistic solutions, which turns out to be caused by phase angles being offset by π. The solutions cannot be computed exactly, but can be approximated very well by 'near-solutions', which become exact in the limit of small deviations. Next, we looked at the probability of an overload in this system, for which we could find an explicit function. After studying this model, we derived several results for a general power network, such as continuity of power flows in terms of power injections, as well as explicit bounds on the overflow probability in this case.

Next, we looked at a system with variable voltages and non-zero resistance. We first assumed a line graph, known as the *Distflow* model, corresponding to a street with one source node and N houses that are supplied by this source. The constraint here is that the voltage differences may not be too large. This model had previously only been solved by full linearization. First, we present an SPD that allows us to quickly compute a solution with a computer, if we wish to solve the problem exactly. Next we try to find a good approximation assuming small deviations. We show that in the most likely overload configuration, the power output at a node is proportional to how far this node is away from the source node. This means that the further a node is away from the source node, the larger its impact, and the more likely it is to cause an overload. We went on to compute explicit bounds on the likelihood of the most likely overload configuration, with very tight bounds that converge to each other in the limiting case of an extremely large line with almost no voltage difference allowed. The results were confirmed with numerical experiments. We also noted that the most likely overload situation becomes increasingly likely as the line becomes longer, even if we also decrease the resistance of lines.

Next, we looked at a slightly different configuration, where the expected power output at houses is not 0 but instead equal to a fixed constant. In this case, we can still apply the same strategies but the exact expressions become very complicated, which is why we only made an estimate in this thesis. Further research would likely be able to establish explicit bounds in this case as well.

The strategy that we used can be applied to other networks as well. As an example, we added a power line to turn the model into a cycle. In this case, we can prove that overloads become far less likely than for the line graph, using a similar technique. It is likely that many other interesting models can also be analyzed using the techniques used in this thesis.

7.2 Outlook

7.2.1 The rate function for a general network

We analyzed a two networks for a system with variable voltages: a line and a circle. Can we also get any results for a general model? Well, we need to redo the computation. Applying the contraction principle relies on:

1. A Large Deviations Principle for the power consumption, with a rate function $I(\mathbf{X})$
2. A continuous transformation T

The LDP for the power consumption relies on what distribution we choose for $\mathbf{X}$, which means that I does not change if we still assume a normal distribution. The transformation T will change. It is important to note that this model is fundamentally different from what we discussed in section 4, since we are now not looking at a transformation of power injections to line power flows, but from power injections to voltages. So we cannot really use the analysis we did there.

Since we are assuming small angles now, the power flow equations reduce to

$$\begin{aligned} P_{km} &= -V_k^2 G_{km} + V_k V_m G_{km} \\ Q_{km} &= V_k^2 B_{km} - V_k V_m B_{km}. \end{aligned}$$

We removed the absolute values since all voltages are assumed to be positive anyways. Even though all lines are equivalent, we cannot assume $G_{km} = G$, since the network is not a complete graph and hence G_{km} is zero for any line $\{k, m\}$ not in the network. So we just leave it like this. It is quite difficult to extract $\mathbf{V}$ from these equations, but we can work the other way around. If we change $\mathbf{V}$ by a bit, then we can compute the change in $\mathbf{P}$. Initially

$$P_k = \sum_{m=0}^{N} G_{km} V_k (V_m - V_k).$$

Suppose $\mathbf{V} \to \mathbf{V} + \epsilon \mathbf{W}$, then we get

$$\begin{aligned} P_k + \epsilon \Delta P_k &= \sum_{m=0}^{N} G_{km}(V_k + \epsilon W_k)(V_m + \epsilon W_m - V_k - \epsilon W_k) \\ &= \sum_{m=0}^{N} G_{km}\big(V_k V_m - V_k^2 + \epsilon(V_k W_m - V_k W_k + W_k V_m - W_k V_k) + \epsilon^2 (W_k W_m - W_k^2)\big) \end{aligned}$$

and ignoring higher order terms of ϵ we get

$$\Delta P_k = \sum_{m=0}^{N} G_{km}(V_k W_m - 2V_k W_k + W_k V_m).$$

We can collect all of this into a matrix M, such that $\delta\mathbf{P} = M\mathbf{W}$. To find this matrix, we collect the terms W_i for every i:

$$\Delta P_k = \sum_{m=0}^{N} (G_{km}V_k)W_m + \left(\sum_{m=0}^{N} V_m - 2V_k \sum_{m=0}^{N} G_{km}\right)W_k.$$

Hence we have

$$M_{km} = G_{km}V_k + \delta_{km}\left(\sum_{m=0}^{N} V_m - 2V_k \sum_{m=0}^{N} G_{km}\right). \tag{147}$$

Note that unlike in section 4, this matrix M is invertible. So by computing the inverse, we have a mapping that sends $\mathbf{\Delta P}$ to $\mathbf{W}$, which is continuous (as it is linear). We conclude that we may apply the contraction principle in this case.

We now assume that $\mathbf{X}$ follows a Gaussian distribution, so that $I(\mathbf{X}) = \mathbf{X}^T\Sigma\mathbf{X}/2$. Then we can either send $\delta\mathbf{P}$ to $\mathbf{W}$, or to $\min_i V_i(\mathbf{X})$. If we choose to do the first, the rate function becomes trivial, like in Section 5 and 6. In this case, we could try to apply a strategy similar to what we did for the line and circle model: differentiating. A big question is if it is still possible to compute the derivatives exactly for a more complicated network. Further research would be needed to see if this is possible: if it is, then this would certainly be a promising strategy. Another idea is to try and make a semidefinite program to compute the solutions, but this may be more difficult since our implementation for the line model relied on having a simple expression for $\mathbf{X}$ in terms of $\mathbf{V}$.

The second, more direct choice for the rate function yields

$$J(\Delta) = \inf\{I(\mathbf{X}) \mid \min_{0\leq i\leq N} V_i(\mathbf{X}) \leq 1 - \Delta\}. \tag{148}$$

If we can compute this function then we have an explicit large deviation principle. The problem is that J can usually not be computed: we only have this implicit definition. It is hard to see a way to continue from here for a general network: for a specific network one could try splitting this into $N + 1$ different problems, one for every i, and then finding an expression for every V_i. But that relies on having a simple expression of V_i in terms of $\mathbf{X}$, which we do not even have for the line model. So this choice seems to be not as promising as the first one.

Appendix 1: full linearization checked by computer

To check the linearization progran we made, we need a computer, since this cannot be solved exactly by hand. The easiest way to do this is by computing the objective function $I_3(\mathbf{V})$ and then replacing $\mathbf{V}^\alpha$ by y_α. This can be done more elegantly than just writing all terms y_α (as was done in the code), but the point of this example is to be simple and clear. Next, we minimize the objective function while the matrix has to stay positive definite. As a check, we can also compute the optimal values V_1, V_2 directly by minimizing $I_3(\mathbf{V})$. First of all, we expect both optimization problems to have the same optimal value. Second, the value of y_{10} found in the first problem should be the same as V_1 in the second problem, as we substituted y_{10} for V_1: similarly the value of y_{01} should be the same as V_2. All of these results are confirmed by the test.

```
I3 = 1/2 (V1 - 1 - μ0)^2 + 1/2 (V1 * V2 + V1 - 2 V1^2 - μ1)^2 + 1/2 (V2 * Λ + V1 * V2 - 2 V2^2 - μ2)^2;

objectivefunction =
  Expand[I3] /. {V1 → y10, V1^2 → y20, V1^3 → y30, V1^4 → y40, V2 → y01,
    V1 * V2 → y11, V1^2 * V2 → y21, V1^3 * V2 → y31, V2^2 → y02, V1 * V2^2 → y12,
    V1^2 * V2^2 → y22, V2^3 → y03, V1 * V2^3 → y13, V2^4 → y04};

entries = {Λ → 11 / 10, μ0 → 1 / 100, μ1 → 1 / 100, μ2 → 1 / 100};

list = SemidefiniteOptimization[objectivefunction /. entries,
  {{1, y10, y01, y20, y11, y02}, {y10, y20, y11, y30, y21, y12},
    {y01, y11, y02, y21, y12, y03}, {y20, y30, y21, y40, y31, y22},
    {y11, y21, y12, y31, y22, y13}, {y02, y12, y03, y22, y13, y04}} ≥ 0,
                                                                    S+^6
  {y10, y01, y20, y11, y02, y30, y21, y12, y03, y40, y31, y22, y13, y04}]

{y10 → 1.01887, y01 → 1.05333, y20 → 1.0381, y11 → 1.07321,
 y02 → 1.10951, y30 → 1.0577, y21 → 1.09347, y12 → 1.13045, y03 → 1.16868,
 y40 → 1.07766, y31 → 1.1141, y22 → 1.15178, y13 → 1.19074, y04 → 1.23101}

objectivefunction /. entries /. list

0.0000607428
```

We can check this against a direct computation:

```
NMinimize[I3 /. entries, {V1, V2}]

{0.0000607433, {V1 → 1.01887, V2 → 1.05333}}
```

Note that the optima are not exactly the same, which is most likely due to rounding errors. Another thing that you may have noticed is that this is not following the formal definition of a semidefinite optimization problem. In the formal definition there is a matrix X with all

entries free that needs to be positive semidefinite, here it is a matrix of a specific form (for example the top left entry is specified, and some other entries are equal). So is this even an SPD after all? Well, it is possible to rewrite this optimization problem so as to match the form of definition 5.1. For that, we would first need to write the objective function as $\langle C, X\rangle$, which is easy. Next, we need to force X to take the specific form that it takes in our optimization problem. This can be achieved by choosing matrices A_k and adding conditions $\langle A_k, X\rangle = 0$. For example, to enforce the fact that $X_{41} = X_{22}$, one could choose

$$A_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{149}$$

Recall that the matrix X is required to be symmetric (as all positive semidefinite matrices are), so $\langle A_1, X\rangle = 0$ implies that $X_{41} = X_{22}$. So with a number of additional equations, we can rewrite our optimization problem in the form of definition 5.1. In practice, however, it is not expected that rewriting the problem this way speeds up computations: although it removes constraints on the matrix, it induces new constraints in the form of $\langle A_k, X\rangle = 0$. So for the best computational results it is likely the best idea to use the matrix $M(y)$.

To see if this semidefinite program is actually faster than the normal polynomial optimization one would need to choose a much larger value N and compare the computation times then. As there was no time to do this in this research project, this would be a good idea for further research.

References

[1] D'Andrade, B. (2017). *The power grid: smart, secure, green and reliable.* Elsevier Academic Press

[2] Yamashita, K., Li, J.; Zhang, P.; Liu, C.C. (2009). *Analysis and control of major blackout events.* 2009 IEEE/PES Power Systems Conference and Exposition, pp. 1-4

[3] Pourbeik, P.; Kundur, P.S.; Taylor, C.W. (2006). *The anatomy of a power grid blackout - Root causes and dynamics of recent major blackouts.* IEEE Power and Energy Magazine vol. 4, no. 5, pp. 22-29, Sept.-Oct. 2006

[4] Veith, E. M. S. P. (2017). *Universal Smart Grid Agent for Distributed Power Generation Management.* Logos Verlag Berlin

[5] Lucarini, V., Faranda, D., Freitas, A. C. G. M. M., Freitas, J. M. M. d., Holland, M., Kuna, T., Nicol, M., Todd, M. and Vaienti, S. (2016). *Extremes and recurrence in dynamical systems. Pure and applied mathematics: a Wiley series of texts, monographs, and tracts.* Wiley, Chichester, pp. 312

[6] Hosein, S.; Hosein, P.; Kattick, W.; Ratan, V. (2016). *Web application for power grid fault management.* ICIAS 2016

[7] Nesti, T. (2020). *Stochastic analysis of energy networks.* Technische Universiteit Eindhoven.

[8] Alhelou, A.H.; Hamedani-Golshan, M.E.; Njenda, T.C.; and Siano, P. *A Survey on Power System Blackout and Cascading Events: Research Motivations and Challenges.* Energies 2019

[9] Den Hollander, F. (2000). *Large deviations.* American Mathematical Society, pp. 19

[10] Dembo, A.; Zeitouni, O. (1998). *Large deviations techniques and applications.* Applications of Mathematics (New York), chapter 4.2.1

[11] Veley, V. F. C. (1987). *The Benchtop Electronics Reference Manual (1st ed.).* New York

[12] Whitney, H. (1972). *Complex analytic variaties.* Institute for Advanced Study Princeton, New Jersey, USA

[13] Penrose, R. (1955). *A generalized inverse for matrices.* Mathematical Proceedings of the Cambridge Philosophical Society

[14] Touchette, H. (2012). *A basic introduction to large deviations: Theory, applications, simulations.* School of Mathematical Sciences, University of London

[15] Christianen, M.H.M.; Cruise, J.; Janssen, A.J.E.M.; Shneer, S.; Vlasiou, M., and Zwart, B. *Comparison of stability regions for a line distribution network with stochastic load demands.* Technische Universiteit Eindhoven, Universiteit Twente, Centrum Wiskunde & Informatica, Heriot-Watt University

[16] Benoist, O. *Writing Positive Polynomials as Sums of (Few) Squares.* Université de Strasbourg, France