Optimising Power Flow in a Volatile Electrical Grid using a Message Passing Algorithm

Elizabeth Mary Harrison

Doctor of Philosophy

ASTON UNIVERSITY

March 2017

Aston University

Optimising Power Flow in a Volatile Electrical Grid using a Message Passing Algorithm

Elizabeth Mary Harrison

Doctor of Philosophy

2017

Thesis Summary

Current methods of optimal power flow were not designed to handle the increasing level of volatility in the electricity networks, this thesis suggests that a message passing-based approach could be useful for managing power distribution in electricity networks. This thesis shows the adaptability of message passing algorithms and demonstrates and validates its capabilities in addressing scenarios with inherent fluctuations, in minimising load shedding and generation costs, and in limiting voltages. Results are promising but more work is needed for this to be practical to real networks.

Keywords: Power Grids, Smart Grids, Networks, Renewable Energy, Belief Propagation

Acknowledgements

I would like to thank David Saad for accepting me as his PhD student, for helping me constantly and for all his patience. Thanks are also due to my second supervisor Ian Nabney for his help with the programming.

I would like to thank ESPRC (Industrial CASE Studentship 12330048) for their funding and opportunities, as well as ALSTOM for the resources I received from them, their funding and support.

I am very grateful to Michael. K. Y. Wong for his wisdom during the problems I encountered and his readiness to help, Diar Nasiev for his help with Matlab and Bill Yeung for his knowledge and help throughout my first year. Also, to Sajjad Fekriasl for his interest and enthusiasm.

I would like to thank everyone who helped check my writing and calculations, my friends for their help on unfamiliar topics, and especially my fiancé Daniel Johnstone and my family for their relentless help, encouragement, advice and prayers.

Publications

The material presented in the thesis has appeared in the following publications:

- E. Harrison, D. Saad, K. Y. M. Wong, "Optimal distribution in smart grids with volatile renewable sources using a message passing algorithm," *International Journal of Smart Grid and Clean Energy*, vol. 5, no. 4, p. 221-228, 2016.
- E. Harrison, D. Saad, K. Y. M. Wong, "Optimal Load Shedding in Electricity Grids with Renewable Sources via Message Passing," *Energy Procedia*, vol. 107, p. 101-108, 2016.
- E. Harrison, D. Saad and K. Y. M. Wong, "Power Grids with Volatile Sources – Message Passing for Optimising Electricity Distribution and Load Shedding," in *2nd International Conference on Offshore Renewable Energy*, Glasgow, 2016.
- E. Harrison, D. Saad and K. Y. M. Wong, "Message Passing for Distributed Optimisation of Power Allocation with Renewable Resources," in *2nd International Conference on Intelligent Green Building and Smart Grid*, Prague, 2016.

Contents

List of Abbreviations

OPF	Optimal Power Flow
DCOPF	DC Optimal Power Flow (Not actually direct current)
ACOPF	AC Optimal Power Flow
VFE	Vertex Free Energy
BP	Belief Propagation
QAR	Quenched Averaging with Additional Resource
PG	Power Grid
RRG	Random Regular Graph
SW	Small-World
BA	Barabási-Albert
MD	Minimum Distance
CPL	Characteristic Path Length

List of Figures

List of Tables

List of Symbols

N	Number of nodes in a network
c, c_j	Number of connections per node (index j indicates the connectivity of a specific node, j)
P, Q	Active and reactive power
P^D, P^G	Active power of demand and generation nodes, respectively
Q^D, Q^G	Reactive power of demand and generation nodes, respectively
V_j	Voltage magnitude at node j
θ_j	Voltage phase angle at node j
I_{ij}	Current between node i to j
R_{ij}	Resistance over the end between node i and j
x_{ij}	Reactance on the edge connective nodes i and j
G_{ij}	The conductance, $\frac{R_{ij}}{R_{ij}^2+x_{ij}^2}$
B_{ij}	The susceptance, $\frac{-x_{ij}}{R_{ij}^2+x_{ij}^2}$
y_{ij}	Current from node j to i
$\mathcal{O}(N)$	Scales to the order of N
s_z	The state of a node z
$m_{z\to a}$	A message sent from a variable z to a factor a
$m_{a\to z}$	A message sent from a factor a to a variable z
E	Energy
ϕ_{ij}	Information on edges, made up of objective functions, ϕ^{OF}, and constraints, ϕ^{CO}
ψ_j	Information on nodes, made up of objectives, ψ^{OF}, and constraints, ψ^{CO}
P_u	Probability of a node being in state u (marginal probability of a node)
T	Temperature
β	$\frac{1}{T}$
Z	Partition function
F	Free energy
$\mathbf{T}_j$	A tree of node j
$N^{\mathbf{T}}$	Number of nodes in the tree **T**
F_{av}	Average free energy

F_V	Vertex free energy
$\mathcal{A}_{ij}$	Adjacency matrix, outputs 1 if j and i are connected by an edge, and 0 otherwise
Θ	The step, or Heaviside function. Outputs 0 for negative argument values and 1 for positive values
ε_{jk}	An adjustment to the edge y_{jk}
A_{jk}	The first derivative of $F_V + \phi$ with respect to y_{jk}
B_{jk}	The second derivative of $F_V + \phi$ with respect to y_{jk}
μ_{ij}, γ_{ij}	Lagrange multipliers
ϵ	A very small, positive value
$\bar{*}$	Expected value of $*$
σ	Standard deviation, e.g., of the capacity of a node
ρ	Percentile
ρ^*	Confidence percentile of a specific network according to the most likely failure
z	z-value according to tables (and inverse equations) of Gaussian distribution
$\mathcal{N}(a^m, a^{s^2})$	The normal distribution from which a will be sampled and the notation of its mean and variance
$\mathcal{U}[a^-, a^+]$	The uniform distribution from which a will be sampled with the minimum and maximum values, a^- and a^+
$\mathcal{F}(\{a^n\})$	A dirac delta distribution, at each value $\{a^n\}$, from which a will take equally portioned samples for all variables
$\mathcal{C}(b, f(\bullet))$	Variable b can take values -1 or 1, indicating the negative and positive correlation of a set of values according to the distribution $f(\bullet)$ from which a will be sampled
ζ_j	The deficit capacity at node j
α_j, I_j	The importance value given to node j
G_j, G_{max}	The generation cost value of node j
P_{min}	The minimum power a generator node can send
g_{min}	The generation cost value of node j when sent power is at P_{min}
v_i	Voltage at node i

Abstract

To reduce their environmental impact and make the most of freely available resource, future electricity grids will include a more significant penetration of renewable generators, such as wind and solar. This will introduce challenges to the grid's method of economic dispatch due to the uncertain and intermittent nature of such renewable generators. The traditional method of power distribution was not originally designed to consider such generation volatility and with the additional pressure of fluctuations in demand, alternative techniques may be necessary. Based on principled statistical physics methodologies, message passing is able to efficiently and inherently consider fluctuations and volatility within electricity grids. Message passing iteratively sends local conditional probabilities as messages and provides a good global approximate solution for a given objective function while maintaining a modest computational complexity which increases approximately linearly with the system size. This distributed optimisation method complements current techniques and can solve large-scale non-convex resource allocation problems. This thesis suggests and analyses message passing as an algorithm for power distribution and considers uncertainties drawn from Gaussian probability distributions. We see the effect of increasing uncertainty on the power grid and examine our results using a program designed for an academic-level understanding of current power grid distribution methods on synthetic benchmark IEEE networks. In addition to this we demonstrate how message passing can accommodate different objective functions such as distribution, generation and environmental costs, as well as load shedding in the case of insufficient resource. The method is adapted to adjust voltages and introduce resistance for DC Optimal Power Flow (DCOPF) and constraints that limit the voltage between fixed values are considered. We discuss our findings on the location and connectivity of volatile nodes, and examine the effects of weighting nodes according to importance or production costs. The quality of the obtained solutions and computational cost of message passing is compared to a leading optimisation technique. Results highlight the importance of correct weighting of objective functions; we see that when this is done correctly, priority consumers can be protected and generators can be turned on according to their cost and capacity. Reliability of a network with uncertainty can be managed through increasing reserve at volatile nodes and our method of quenched averaging with predetermined confidence levels ensures a higher reliability in the form of higher reserves at uncertain nodes as a form of protection. The algorithm can effectively consider weighing variables on nodes as well as on edges and can thus accommodate voltage variables and resistance, allowing for a closer imitation of DCOPF characteristics. The thesis shows how effective the message passing algorithm can be and demonstrates a fraction of its capabilities. However, the approach needs more work before it can be fully comparable with existing approaches.

1 Introduction

1.1 Background

The UK uses 311 TWh of electricity a year which is approximately 4,813 kWh per person [1]. This electricity is generated in a variety of power stations and sent to houses, businesses and industries via a network of power lines. The power generated throughout the UK in 2015 came from three main sources; approximately 52% from fossil fuels, 21% from nuclear power and 25% from renewable sources [2] (Figure 1.1).

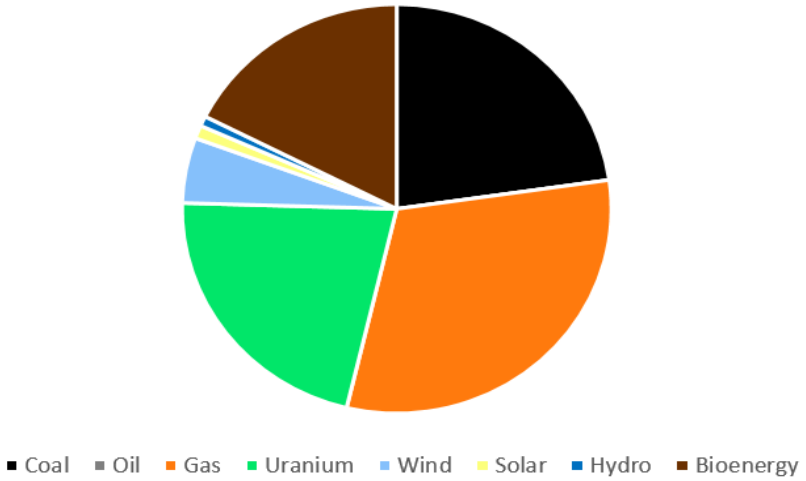


Figure 1.1 Composition of fuels used for electricity in the UK in 2015 [2, 3].

Fossil fuels dominate the electrical grid supply with a composition of coal (22%), oil (<1%) and gas (30%), these fuels are the fossilised remains of prehistoric plants and animals which have taken millions of years to form [4]. As they take so long to produce, the rate at which they are being harvested is unsustainable and they are expected to become depleted beyond an economic level of usage within a couple of hundred years. Estimates vary, however recent calculations suggest that this may happen as soon as 2040, 2042 and 2112[1], for oil, gas and coal, respectively, if usage continues at its current rate [5]. As well as their unsustainability, non-renewable fossil fuels produce undesirable by-products such as CO_2, nitrogen oxides, sulphur oxides, mercury and solid ash, generated by being burnt in a furnace for heat energy. CO_2 enters the Earth's atmosphere and disturbs the chemical balance which is essential for

[1] Estimates may be revised for gas as a relatively new method of retrieving gas from air pockets within the Earth, called fracking, makes estimating stores difficult. Gas is nonetheless still considered unsustainable.

reflecting most of the Sun's heat; this increases the Earth's temperature, leading to agricultural changes and rising sea levels and is just one of many concerns related to the by-products of burning fossil fuels.

One alternative to fossil fuels currently being utilised is nuclear power. Large amounts of energy are produced during nuclear fission through the process of splitting uranium atoms. Uranium supplies are expected to last for around 230 years with current technology and identified uranium sources [6]. The process produces unwanted by-products such as uranium, plutonium and barium which are highly radioactive, making them dangerous and expensive to dispose of. The expensive disposal process consists of being sealed and buried for thousands of years, in a place with minimal disruption to life, protected from natural disasters and human disruption. The past 40 years have seen several high-profile cases including Chernobyl (Ukraine) in 1986 [7] and Fukushima (Japan) in 2011 [8] which highlight the risks and lower public confidence (only 6% of Japan's population in 2011 thought that nuclear power was safe and new plants should be built, down from 21% in 2005 [9]).

Figure 1.2 shows some drawbacks of various sources of electricity generation, highlighting how costs, issues and dangers appear to be more prominent within non-renewable sources. Despite the high associated costs and risks, fossil fuels and nuclear sources have the advantage that the rate at which they produce energy is controllable. With good demand forecasts, generation of energy can be accurately forecasted and controlled to meet expected demand. This is beneficial to economies like the UK where there are changes in demand, dependent upon conditions such as time of day, season and cultural events.

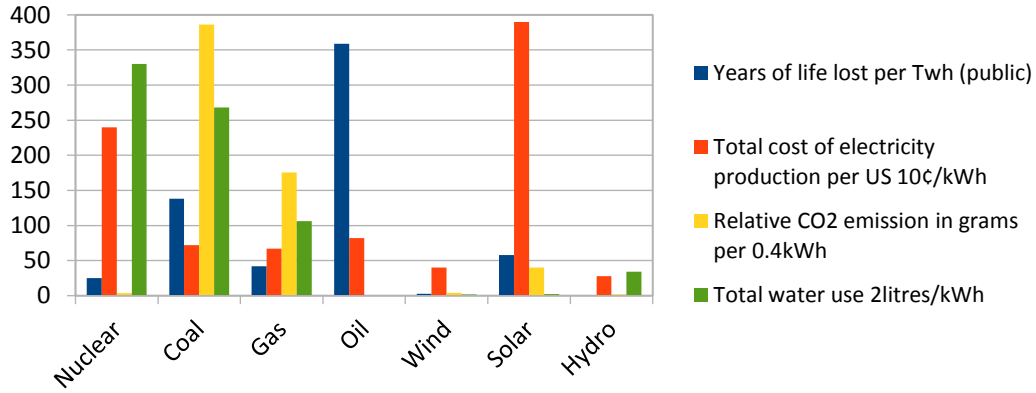


Figure 1.2 The associated costs of various power sources: Years of life lost per TWh in blue [10]. Total cost; production, construction and decommissioning in orange [11] (where wind represents onshore wind only and [12] suggests that solar costs more than halved within two years since these figures were taken), amount of CO_2 emissions per kWh in yellow [13] and water usage in green [11].

As an alternative to non-renewable energy sources, there have been recent ventures into the use of renewable energy sources. The UK government plan to invest £40 billion into renewable electricity generation by 2020 [14]. Increased investment is due to the perception that renewable energies are better for the environment, the technologies for harnessing them reaching a level where they can be used in power grids and the unsustainable nature of fossil fuels. Such sources produce significantly smaller amounts of harmful waste and as the name implies, are self-sustainable and temporally unlimited. Some sources of renewable energy are:

- Biofuel is from the anaerobic digestion of matter [15] which is mostly used as fuel for transportation. Studies have shown that burning certain crops for biofuel as an alternative to fossil fuels within electricity generation can actually contribute more to greenhouse emissions than standard fossil fuels due to high levels of N_20 emissions [16]. Rapeseed and maize are among these high N_20 emission fuels, but other fuels such as sugar cane, palm oil, poplar and willow trees produce lower levels and so remain a better alternative.
- Geothermal power is effective, reliable and sustainable, only emitting a small fraction of harmful emissions compared to fossil fuels. It works by heating water using the high temperatures inside the earth; the UK is not ideal for this as it does not lie on any actively volcanic land, but there are potential sites for geothermal technology to be harnessed. Discussions throughout the last 5 years with Iceland highlight the possibility of Iceland's volcanoes supplying geothermal electricity to the UK which could meet up to a third of the UK's demand [17, 18], however this could cost over £500 million and no decisions have been made [19].
- Hydroelectric power in the UK can currently generate around 5,885 GWh a year [20], almost 2% of the UK's consumer consumption. Power is generated from the gravitational force of running water, it is reliable and can store energy in a clean, environmentally friendly way[2], however estimates suggest its maximum potential could only meet up to 5% of the UK's consumption .
- Solar PV power converts sunlight into electricity using photovoltaics. Although the weather in the UK is not ideal for this due to regular, high levels of cloud coverage,

[2] Pumped storage uses low cost off-peak power to pump water to a higher elevation and stores it as gravitational potential energy. At peak times the water can be released through turbines.

solar power had a capacity of producing 1.4 TWh (0.4%) in 2013, and a technical maximum potential of 18 TWh (5.8%) by 2020 [12][3].

- Wind turbines have the highest renewable potential in the UK, they convert kinetic wind energy to electrical energy. The UK currently has 5,809 onshore turbines and 1,465 offshore, totalling an annual energy produced in 2015 of 39 TWh (12.5% of total UK consumption) [21]. Plans for more wind farms are ongoing throughout the country with a huge potential.
- Wave, tidal and ocean technologies are also being developed. They harness the movement, pressure and power of the sea using technologies such as pistons and turbines. Considering additional issues such as the hostility of salt water, marine life, additional forces from the waves and costs these technologies have until now had a much smaller audience and impact. However, with such huge potential in the ocean, many believe these sources will be one of our biggest electricity contributors in future years with [22] suggesting it could be used to supply up to 75% of the UK's electricity demand.

A major drawback of renewable energies with significant potentials like wind and solar[4] is that power sources cannot be controlled or accounted for at any one time (for example the wind cannot be turned on or up to a higher level to suit customer demand at any given time) and the predictions are not always completely accurate. Therefore, even with demand and renewable source forecasts, control of energy production is limited in comparison to those of a non-renewable nature[5]. Coupled with limited technology for storing large amounts of produced energy this means that renewable energy's capacity for supplying a large economy is currently restricted.

Despite this, the benefits of having clean power sources as a replacement for ageing non-renewable sources and technologies have been recognised by many organisations and governments, and corresponding legislations have been set. According to the Department of Energy and Climate Change the UK has a target of 15% of power in the electrical grid being

[3] Future technology could see a 24-hour solar panel available [188] which would reduce volatility and temporal fluctuations. But as these technologies are currently unavailable, for the purpose of this thesis, they will be ignored.

[4] Throughout this thesis "renewable energies" will imply only the fluctuating wind and solar sources only, according to their readily available technologies (not controllable hydro, geothermal, nor 24 hour solar panels).

[5] It is expected that tidal and wave sources will not be fully controllable, but may be more predictable and therefore more reliable.

renewable by 2020 and is already reaching 8.3% (using the methodology of the directive, provisional calculations) of energy coming from renewable sources in 2015 (Figure 1.3) [3]. Targets also hope to reduce greenhouse emissions by 80% by 2050 [23].

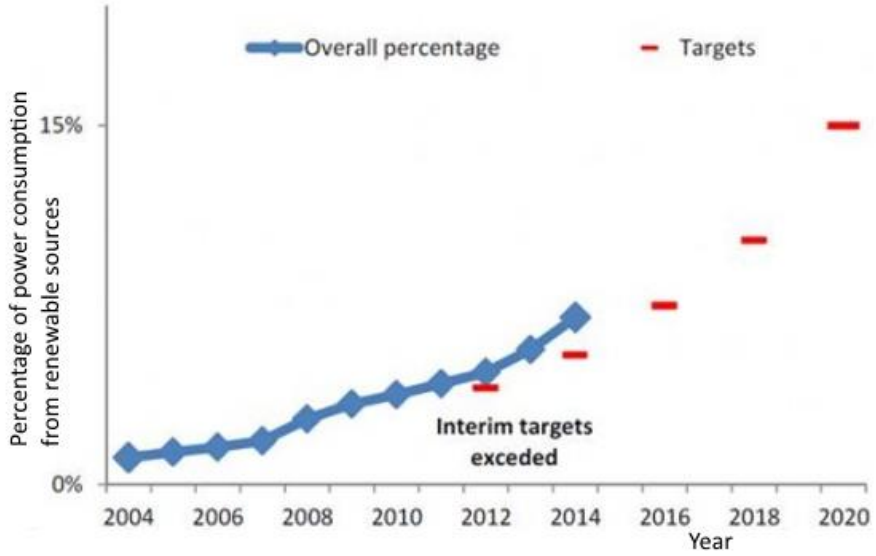


Figure 1.3 The UK's progress towards its European Directive target of 15% of power in the electrical grid being renewable by 2020 [24]. © Copyright under the Open Government Licence.

The growing reliance on renewable sources is aided by the increased use of microgenerators, which provide small scale generation of electricity to individuals, small businesses and communities; for example, a wind turbine on the top of a block of flats or a solar panel on the roof of a house. The UK government supports microgeneration for instance through an £86 million grant for microgeneration technologies [25], and although the intention that all homes built from 2016 would be zero carbon [26] was since scrapped, regulations and incentives ensure that homes will continue to meet high standards and will be "nearly zero energy buildings" by 2021 [27]. One example of this is the "Low Carbon Buildings Programme" which allows the public sector and non-for-profit organisations to request 50% of their microgeneration installation costs and offers grants to householders for domestic microgenerators [28]. Microgenerators at consumer nodes may increase the correlation between connecting nodes and this needs considering when power distribution is being calculated. For simplicity within this research, it is assumed that volatility from microgeneration is already absorbed within the expected capacity, production or surplus, of consumers (consumer average capacities and variance will include both their expected consumption and microgeneration) and correlations are ignored.

For consumers to receive the power generated from power sources, current must flow from generation plants through a network of power lines, substations and transmission stations; this is the electricity grid. Within an electricity grid there are two main types of networks; transmission and distribution. Transmission networks are more complex structures with loops, but are mostly sparsely connected ($N \gg c$, where N is the number of vertices in a network and c is the connectivity of each vertex in the network); they generally run between substations to generation stations and to landing stations (a substation which leads to areas of consumption), and are characterised by high voltage lines to reduce power loss during transmission. Distribution networks are the sections of the network between the landing stations and the consumers; at this point the networks are typically tree-like structures and are characterised by lower-voltage lines (Figure 1.4). This thesis will focus on the transmission stage and any talk of consumer nodes will imply the landing station leading to the distribution stage. This research has the capability to work successfully within the distribution stage, but it is not the focus of this study. One should point out that due to the complexity of the electricity grid, the distinction between the different stages and their characteristics is not clear-cut.

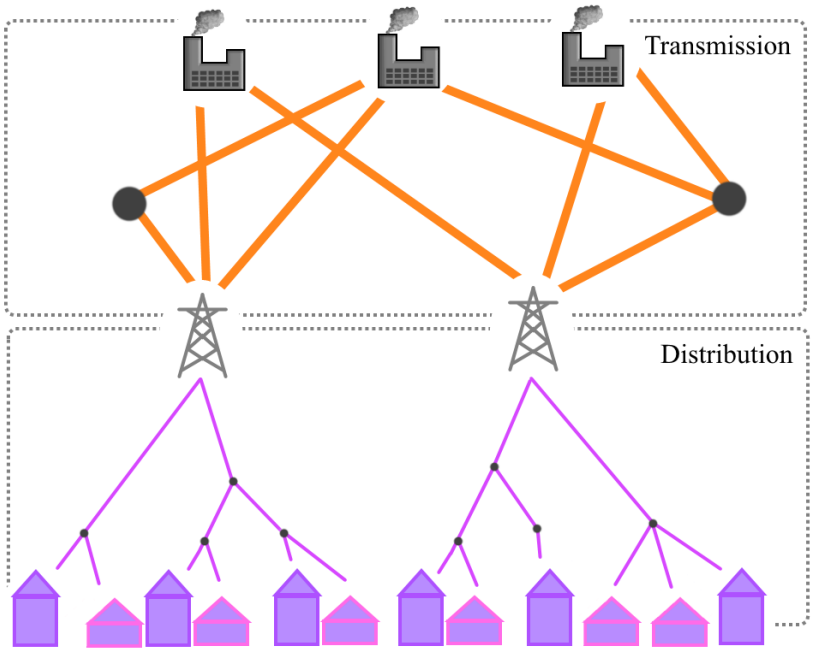


Figure 1.4 Sketch of the structural difference between transmission and distribution stages in the electricity grid.

1.2 DC and AC Circuit Theory

To understand complex AC power flow used within an electrical power grid, we must first understand the fundamentals of electrical circuits. Starting with atoms, which are made up of protons which have a positive electrical charge, electrons which have a negative electrical charge, and neurons which have no electrical charge. An atom is arranged with the neurons and protons at the centre and the electrons in an outer shell, attracted to the positive electrical charge of the protons in the nucleus.

1.2.1 DC Power Flow

When the electrons are separated from the nucleus, there is a pull for them to be reunited and this is called the potential difference. Within a circuit, loose electrons are attracted towards the protons and this flow is the electrical current.

Electric current (I) is the flow of electrical charge. The continuous flow of electrons, being pushed by a voltage source. Voltage (V) is the potential energy of an electrical supply. The difference between two voltages is the potential difference. Potential difference, essentially, represents the work required to move electrons away from the attractive force of the nucleus. Voltage can exist without current, but current cannot exist without voltage. The material that the electrical current moves through can be highly conductive (such as copper) or highly insulated (such as plastic), each has some level of resistance which restricts the current passing through it, this is Resistance (R). Any power lost through resistance appears as heat or light. Alternately, the resistance can be expressed as conductance $G = \frac{1}{R}$, which is the ability of the material to conduct electricity. Voltage, resistance and current are closely related through Ohm's Law:

$$V = IR. \tag{1}$$

In DC circuits, the voltage source is constant and the voltage source (e.g. battery) has a positive and negative side, which indicates the singular direction of current flow.

There are additional laws within electrical power flow that restrict a network. Two of which are Kirchhoff Laws; the first required the sum of all currents entering and leaving a node must be equal to zero:

$$\sum_{i=1}^{c} I_i = 0, \tag{2}$$

where c is the number of currents flowing through a node, and is commonly known as the Conservation of Charge, and the second is that the total voltage around a circuit is equal to the sum of all potential differences (or voltage drops) in the circuit, this is called the Conservation of Energy:

$$\sum_{i=1}^{N} V_i = 0, \tag{3}$$

where N is the number of nodes in the circuit.

Electric power (P) is the rate that energy is consumed or generated within a circuit. An energy source produces power, and a load absorbs it. It can be expressed through the electrical power formula:

$$P = VI. \tag{4}$$

Power over time gives electrical energy ($\mathbb{E}$), which is the capacity to do work and is measured in watts per hour, or in terms of power grids kWh (kilowatts per hour), MWh (megawatts per hour) and TWh (terawatts per hour). Turning on a 100-watt lightbulb for 1 hour will cost 0.1kWh total consumption:

$$\mathbb{E} = tP. \tag{5}$$

1.2.2 AC Power Flow

In AC power flow, voltage is generated by a coil rotating around a magnetic field, this results in the voltage source varying periodically with time and as they are proportional, current also varies periodically with time.

The AC voltage, v, over time can be expressed as a sine wave, as seen in Figure 1.5, or any periodic function; for derivation purposes, we will mention the combination of sin and cos (rewritten as an exponential) expressed as:

$$v_i = |V_i| \sin(\omega t + \theta_i), \quad or \quad v_i = |V_i| e^{\boldsymbol{i}(\omega t + \theta_i)}, \tag{6}$$

where v_i is the AC voltage, $|V_i|$ is the voltage magnitude, ω is the angular frequency ($\omega = 2\pi f$, where f is the frequency, in a power grid this is around 50Hz), t is time, θ_i is the phase difference (or phase angle) between the voltage and the AC current, and bold $\boldsymbol{i}$ is the imaginary number $\sqrt{-1}$. In Figure 1.5 the phase difference (or phase angle) is 0, and the only element preventing current flow is resistance, therefore AC current follows a similar waveform $i_i = |I_i|\sin(\omega t + \theta_i)$, or $i_i = \frac{v_i}{z_{ij}} = \frac{|V_i|}{z_{ij}}\sin(\omega t + \theta_i)$, where the ratio of voltage to current is called the impedance (a complex version of resistance), $Z_{ij} = R_{ij} + \boldsymbol{i}X_{ij}$. In AC circuits, complex or apparent power is $S_j = P_j + iQ_j = v_j i_j^*$, where i^* is the complex conjugate of the AC current i, so the two sin curves are multiplied as shown in Figure 1.6, and all power is real.

Figure 1.5 The waveforms of AC voltage and current with a potential difference of $\theta = 0$ [29].

Figure 1.6 The apparent power consumed according to the in-phase AC voltage and current [29], x-axis shows voltage, current or power (according to each curve).

In other cases, the current may be lagging or ahead of voltage as shown in Figure 1.7 and Figure 1.8. In this case, there is a phase difference, θ, which can be found by plotting the path of v and i as a phasor diagram (shown at the left of Figure 1.7), and finding the angle between values. The apparent power can be calculated as shown in Figure 1.9. Resistance is not affected by

frequency and so AC impedance is equal to DC resistance and the average power loss from AC is equivalent to that of DC.

Figure 1.7 The AC voltage with AC current lagging behind by phase angle θ [29].

Figure 1.8 The AC voltage with AC current ahead by phase angle θ [29].

Figure 1.9 The apparent power calculated for a voltage and current at phase angle $90°$ [29], x-axis shows voltage, current or power (according to each curve).

Some of the reasons why the UK uses AC power flow instead of DC are; the cheaper manufacturing cost of AC generators, and the ability step up or down voltages using transformers (two or more coils of wire where an electromagnetic field through one induces a higher/lower voltage in the other) to reduce power loss when electrical power is transmitted over long distances (because voltages are oppositely proportional to resistance).

1.3 Control of Electrical Power Flow

Electrical power flow is the distribution of electricity, within an electrical power grid, from generators to consumers. The electricity industry in the UK works by a combination of four main elements; **Generators, Suppliers,** the **National Transmission Network** and **Distributors** (These are the industries involved, and will be denoted in bold, not to be confused with either generators or distribution mentioned throughout the thesis). The **generators**, such as Alstom Grid, are the industries which own and operate the coal, oil, gas, nuclear, hydro and wind power stations. The **suppliers**, such as British Gas, are those who supply and sell electricity to consumers. The **distributors**, such as Northern Power grid and Western Power Distribution are those which own and operate the distribution network towers and power lines. And the **Transmission Network**, such as National Grid, is the industry which maintains the flow of generated electricity on a minute to minute basis (National Grid controls the transmission network for England and Wales, it also has interconnects with France, Northern Ireland and the Netherlands so that spare power can be traded between them). Each of the other industries pays for the right to connect to the transmission network [30].

The work in this thesis focuses mostly on issues relevant to the **Transmission Network** stage and how companies such as National Grid can maintain the correct flow of electricity "as safely, efficiently and smoothly as possible at all times", with 99.99999% reliability [31].

At the transmission stage of distribution, the power needs to be transferred through a network of power lines to the landing stations (and ultimately to the consumers). This is done by the continuous use of four methods of control:

- *Primary (Frequency Response)* - This is the automatic millisecond to millisecond balance between system demand and generation, controlled by change in frequency [32]. If generation is greater than demand the frequency rises, or if demand exceeds generation, the frequency falls (Figure 1.10). The change in frequency prompts an automated secondary response.

Figure 1.10 Real frequency response data at 15-second intervals over one hour [32].

- *Secondary (Automatic Generation Control)* - Restores frequency to its scheduled value of 50Hz by providing additional active power or reducing it to ensure the frequency stays within ±1% of nominal 50 Hz system frequency limits. This is done by adjusting voltages and happens within 30 seconds of the frequency change. Sufficient power or demand needs to be readily available for all credible adjustments.
- *Tertiary (Economic Dispatch)* - This is the manual stage of optimally distributing power in order to satisfy the demand and it uses predictions of the consumers expected demand and producers expected generation. This happens every 15-60 minutes and involves the turning up and down of electricity to be produced by the power stations to maintain appropriate voltage and frequency levels over the following time-period [33].
- *Tertiary (Unit Commitment)*- The decision of turning on and off power stations according to expected power consumption, with a time-period between one hour and a number of days.
- *Tertiary (Planning)*- Major decisions of construction and decommissioning of power stations according to demand forecasts for the upcoming years and decades.

This research will consider the economic dispatch stage. Computers are used for the non-trivial task of calculating the most efficient distribution route through the electrical network which minimises given objective functions according to the limits of the grid and considering changes over the given time-period. The terms ‘economic dispatch’, ‘power flow’ and ‘optimal power flow’ all slightly vary in meaning throughout literature. In this thesis, ‘economic dispatch’ will refer to the distribution stage where decisions are made about how much power each generator should be producing to satisfy the primary and secondary control stages for the following time-period. ‘Optimal power flow’ will describe the current method of power distribution at the economic dispatch stage and ‘power flow’ will mean the flow of power within a network.

1.4 Optimal Power Flow

The current method of optimal power flow (OPF) was first suggested in 1962 by Carpentier [34] and the method has changed little since then [35]. OPF includes the process of power flow to satisfy Kirchhoff's law and it also optimises the objective functions to find a more optimal solution while considering the constraints. The method suggested by Carpentier works by alternately repeating two steps until the overall optimum is found (Figure 1.11). The first step uses the Newton Raphson method (a root-finding algorithm which uses the first few terms of a functions Taylor series to find an accurate solution for the root) or an iterative method like the fast-decoupled technique to estimate initial values for electricity distribution to achieve zero excess, enforcing Kirchhoff's law, according to fixed power inputs. The second finds a better solution of power inputs to optimise the objective function while ensuring that all the constraints are met, using techniques such as gradient-descent or the interior point method. For the last 55 years, this has been the basis on which economic dispatch of electricity has been calculated. The advantages are that an optimal solution can be found quickly, but as it is a deterministic method that does not inherently consider probabilities, although it can consider small scale fluctuations from consumers, it was not designed to handle more substantial fluctuations such as those from renewable sources.

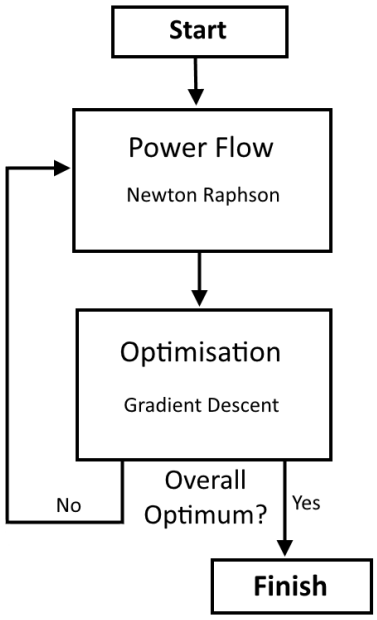


Figure 1.11 The stages of an OPF algorithm.

The objective functions are to minimise:

- Power loss – When current flows through a power line, energy is lost through heat. This follows the equation, $P_{\text{loss}} = I^2R$, where P_{loss} is active power loss, I is current and R is resistance.
- Load shedding – In the event where demand is higher than generation, the electricity provided to some consumers must be dropped or cut to balance demand against generation. Load shedding aims to minimise the number of consumers effected in this situation.
- Generation costs – The cost of power generation for different power plants varies largely for many reasons; cost of fuel, process of conversion to electricity, transportation of fuel, maintenance, efficiency etc. In addition to this the life cycle costs of the plant must be considered; building costs and expected decommissioning costs. This results in each power station having a different levelized cost of energy; one of the main objectives is to minimise these generation costs while meeting consumer demand.
- Bus voltage deviation – It is desirable to keep all voltages as close to 1 per unit as possible while distributing load; this can be added as an objective or soft constraint.
- Emissions of generation un its – Each power station emits different levels of a variety of pollutants and OPF aims to minimise the most harmful ones, such as CO_2, NO_2 and toxic waste.
- Number of control actions – reducing the number of manual controls will reduce the need for human intervention and consequently the risk of human error. It also may increase computational convergence speed of the control process.

The constraints include:

- Conservation of charge (Kirchhoff's law) – This states that the sum of all currents entering and leaving each node must be equal to zero.
- Bandwidth – A power line can only withstand a certain amount of current along it; if exceeded the line can overheat and warp, which could result in the line tripping.
- Minimum and maximum active and reactive power generation – Power generators cannot generate infinite amounts of power; each is limited to its own maximum. Similarly, when a power station is on, there is a minimum amount of power it can economically produce.

- Voltage, frequency and phase constraints – Electricity networks operate within strict voltage, frequency and phase angle constraints that cannot be violated due to their impact on the remainder of the network and its stability.

And the control variables are:

- Reactor status (discrete)
- Switch capacitors (discrete)
- Phase shifters
- Transformer tap settings
- Real power outputs
- Voltage magnitudes and phase angles
- Frequencies

There are also parameters to consider, such as resistance, reactance and impedance values, and the topology of the network.

The goal is to find the optimal setting to optimise the objective functions, satisfy the power flow equations, maintain a secure system, adhere to system operating limits, be self-healing and to use ACOPF (optimal power flow for an AC network) for unit commitments within a short period. A method which can consider all these objectives and constraints for such a complex problem, within the time constraints, is yet to be found because the problem is nonconvex, nonlinear, has discrete and continuous variable and may have multiple optimal solutions, and so compromises are made. The mentioned existing methods are used in combination to minimise all of the given objective functions under constraints. Usually the set of equations will be decoupled into the active power and phase angle, $P\theta$, and the reactive power and voltage magnitude, QV, due to tight coupling between the two pairs of variables. Generally, ACOPF is calculated by first minimising costs by varying real power using DCOPF (optimal power flow for a simplified AC network, with similar properties to DC power flow equations), the generator outputs are then fixed and an ACOPF is done to minimise losses by varying reactive power, capacitors, etc. These are done at 15-60 minute time intervals and are measuring and predicting the next time-window, time is not considered at this stage but conducted during the primary and secondary stages.

The method can be written mathematically as minimising the operating costs, f, according to the control variables, u:

$$\min_{u} f(u), \tag{7}$$

where the operation cost is a weighted combination of minimising the cost of generation, $\sum_i^{\boldsymbol{G}} g_i(P_i)$, the cost of changing controls, $\sum_i^N h_i(u_i - u_i^0)$, and minimising system losses, $\phi(u)$, where g_i is the specific generation cost function of generator i according to the active power it produces, P_i, for all generators, $\boldsymbol{G}$. h_i is the cost function of changing any control variable, u_i, from its previous state, u_i^0, for all nodes in the network, N, and ϕ is the cost of transmission in terms of power lost through heat over power lines (other objective functions can also be included here). The objective function is subject to some equality and inequality equations:

$$\begin{aligned} R(u) &= 0, \\ K(u) &\geq 0, \end{aligned} \tag{8}$$

where $R(u)$ is the equality constraints from the standard polar representation of the power flow equations, derived from previous definitions of AC power and current:

$$S_j = P_j + \boldsymbol{i}Q_j = v_j i_j^* = v_j \sum_{i=1}^{c} \frac{v_j}{Z_{ij}}, \tag{9}$$

impedance can be expressed as the complex function of resistance and reactance, and the AC voltages are expressed in terms of their voltage magnitude and phase angles:

$$P_j + \boldsymbol{i}Q_j = \sum_{i=1}^{c} \frac{|V_i||V_j|e^{\boldsymbol{i}(\omega t - \theta_i)}e^{\boldsymbol{i}(\omega t + \theta_j)}}{R_{ij} + \boldsymbol{i}X_{ij}}, \tag{10}$$

The time can be ignored as we are concentrating on one instance and the exponential functions can be rewritten using trigonometric identities, $e^{ia} = \cos(a) + i\sin(a)$:

$$P_j + \boldsymbol{i}Q_j = \sum_{i=1}^{c} \frac{\left(|V_i||V_j|\left(\cos(\theta_j - \theta_i) + i\sin(\theta_j - \theta_i)\right)\right)\left(R_{ij} - \boldsymbol{i}X_{ij}\right)}{R_{ij}^2 + X_{ij}^2}. \tag{11}$$

Finally, the real and imaginary parts can be split into the active and reactive equations:

$$P_j^{\boldsymbol{G}} - P_j^{\boldsymbol{D}} = \sum_{i=1}^{c_j} V_j V_i \left[G_{ij} \cos\left(\theta_j - \theta_i\right) + B_{ij} \sin\left(\theta_j - \theta_i\right) \right], \quad (12)$$

$$Q_j^{\boldsymbol{G}} - Q_j^{\boldsymbol{D}} = \sum_{i=1}^{c_j} V_j V_i \left[G_{ij} \sin\left(\theta_j - \theta_i\right) - B_{ij} \cos\left(\theta_j - \theta_i\right) \right], \quad (13)$$

where $P_j^{\boldsymbol{G}}$ and $Q_j^{\boldsymbol{G}}$ are the active and reactive power generated at node j, and $P_j^{\boldsymbol{D}}$ and $Q_j^{\boldsymbol{D}}$ are the active and reactive power consumed at node j. V_j and θ_j are the voltage magnitude and phase angle of node j and the sum function is over all nodes that node j is connected to, c_j, via a power line. G_{ij} and B_{ij} are the conductance and susceptance of the power line between nodes j and i, they represent the real and imaginary parts of the admittance, $Y_{ij} = \frac{1}{Z_{ij}} = \frac{R_{ij} - \boldsymbol{i} X_{ij}}{R_{ij}^2 + X_{ij}^2}$, where $G_{ij} = \frac{R_{ij}}{R_{ij}^2 + X_{ij}^2}$ and $B_{ij} = \frac{-X_{ij}}{R_{ij}^2 + X_{ij}^2}$, where R_{ij} is the resistance, and X_{ij} is the reactance. These power equations are true for each node j [36] and represent that the overall power demanded or available from a node needs to be equivalent to the incoming or outgoing power from all connected edges.

$K(u)$ represents the inequality constraints such as the upper and lower limits of the control variables, $u^{lower} \leq u \leq u^{upper}$, and the limits of the flow over edges, $|y_{ij}| \leq y_{ij}^{upper}$, where y_{ij} is the directional flow over the power line from node j to i.

This is a very complicated resource allocation optimisation problem because ACOPF is nonlinear due to the objective functions and constraints, and non-convex due to the non-linear AC voltage terms; to simplify this a set of approximations are made:

- Start by neglecting Equation (13), the reactive power.
- Then the conductance can be neglected because the resistance within the transmission stage is significantly less than reactance; so $G_{ij} \leftarrow 0$ and $B_{ij} \leftarrow -\frac{1}{X_{ij}}$.

$$P_j^{\boldsymbol{G}} - P_j^{\boldsymbol{D}} = \sum_{i=1}^{c_j} V_j V_i \left[B_{ij} \sin\left(\theta_j - \theta_i\right) \right]. \quad (14)$$

- Assume voltage magnitudes are 1 per unit.

$$P_j^{\boldsymbol{G}} - P_j^{\boldsymbol{D}} = \sum_{i=1}^{c_j} B_{ij} \sin\left(\theta_j - \theta_i\right). \quad (15)$$

Also, if voltage magnitudes are approximately equal to 1, then power P can be approximated to equal I.

- Also assume that the phase angles are small, allowing us to write $(\theta_j - \theta_i)$ instead of $\sin(\theta_j - \theta_i)$.
- Also, remembering that $B_{ij} = \frac{-1}{X_{ij}}$, we get:

$$P_j^{\boldsymbol{G}} - P_j^{\boldsymbol{D}} = \sum_{i=1}^{c_j} \frac{(\theta_i - \theta_j)}{X_{ij}}. \tag{16}$$

Also, it can be written that the overall power at node j ($P_j = P_j^{\boldsymbol{G}} - P_j^{\boldsymbol{D}}$) minus any power leaving the node j through power lines to neighbouring nodes i, P_{ij}, must be equal to zero (using Kirchhoff's Current Law where $P \approx I$).

$$P_j - \sum_{i=1}^{c_j} P_{ij} = 0. \tag{17}$$

Leading to the conclusion that:

$$P_{ij} = \frac{(\theta_j - \theta_i)}{X_{ij}}, \tag{18}$$

which can be compared with the DC equation for flow of current:

$$I_{ij} = \frac{V_j - V_i}{R_{ij}}, \tag{19}$$

where I_{ij} is the current over the edge connecting nodes j and i. Therefore, DCOPF does not actually consider a DC network flow, it merely considers AC under strict assumptions, resulting in equations similar to those of a DC flow.

Some software available to simulate AC and DC power flow and OPF includes Powerop [37], or Matpower [38].

1.5 Problems Associated with OPF

Renewable energies such as wind and solar are uncontrollable, which means they cannot be turned up on demand, they produce electricity independently of consumption (Figure 1.12 and Figure 1.13), and ultimately do not remain constant over the 15-60 minute intervals of economic dispatch. Predictions are available but are not entirely accurate because of the complicated and stochastic nature of the weather [39] and of the generators themselves. For these reasons, they are labelled as *volatile*. The increase in volatility in electrical grids puts extra pressure on the existing distribution methods because they were not designed to operate under these conditions, but are still required to run smoothly and reliably, providing correct amounts of power to each consumer while balancing generation and demand. If more power is produced than consumed, the power is lost and wasted through heat with clear costs implications; for example, [40] suggests (assuming that wind probability density function is a Beta distribution and using 1 hour-ahead predictions) that costs due to intermittency could be around 12,875 €/MW annually for Spanish wind production, which is approximately 9.7% of its annual gross income. Study [35] suggests that current solvers are around 10% off optimal solutions on average even before considering volatility of renewable generators. Alternatively, if insufficient power is generated, consumers are left without power which, depending on the scale and time-periods, can be harmful to homes, businesses and even lives [41, 42]. Table 1 shows that all over the world the number and scale of blackouts have increased steadily and significantly. This is not entirely due to the introduction of renewables, but it is related to the increased correlations between probabilistic elements in the system; these are present in renewable energies as most are linked to the weather. Blackouts are usually mitigated by using reserve; spare power produced to prevent insufficiencies during lower than expected power generation, but this is costly to produce. The current system is becoming outdated and a new method of distributing power may be necessary, one which is robust to accidental damage or deliberate attack and can effectively consider system limits, all within a volatile network and within a short time-period. A recent report [43] shows that the UK is not expected to meet its renewable energy target of 15% by 2020, whereas other European countries are expected to surpass theirs. However, a power grid with a large proportion of intermittent renewable energy penetration requires an algorithm that can consider fluctuations and correlations, is probabilistic and works well on a large scale; this requires finding a new method of distribution.

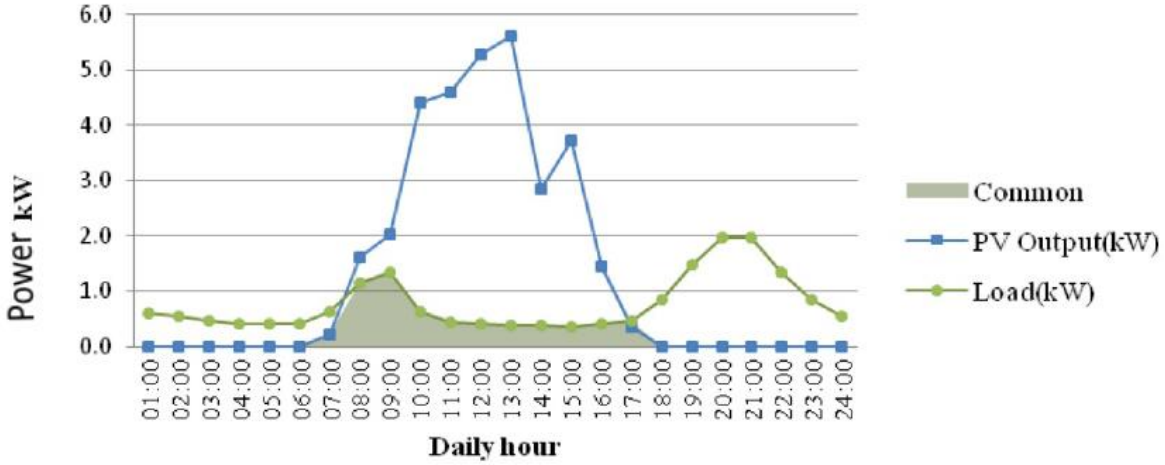


Figure 1.12 Solar PV output and consumer load over 24 hours [44].

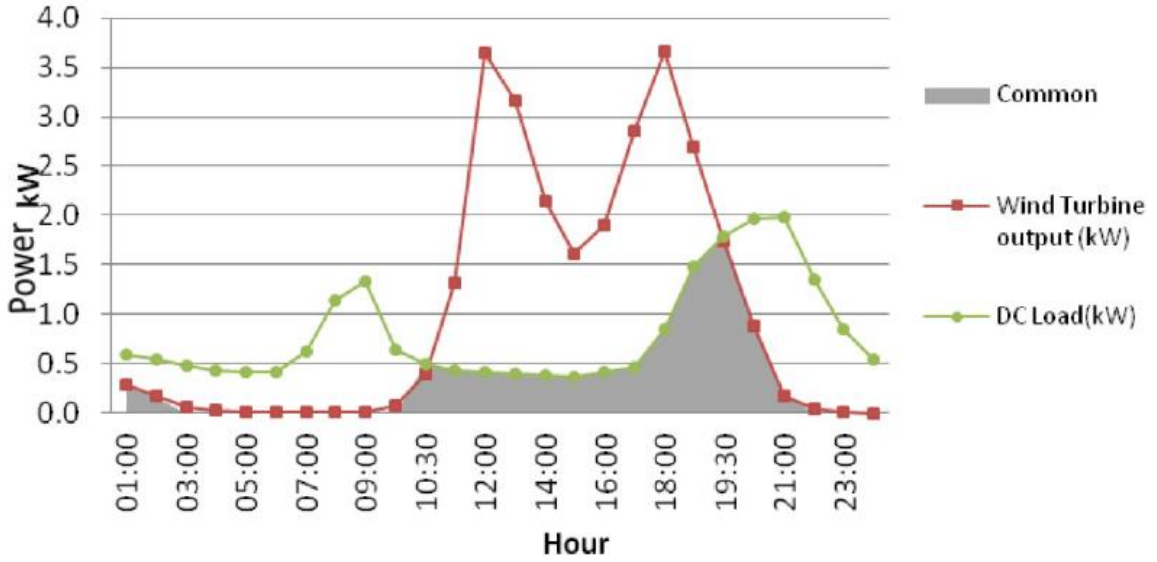


Figure 1.13 Wind turbine output and consumer load over 24 hours [44].

Table 1. The number of major blackouts that effected over 30 million people over six decades [45, 46, 47, 48, 49, 50] and the global cumulative wind capacity (in MW by the end of each decade) [51].

Years	# Major Blackouts	# People Affected	Wind Capacity
1960-1969	1	30m	-
1970-1979	1	40m	-
1980-1989	0	0m	-
1990-1999	1	97m	13,600
2000-2009	5	527m	159,089
2010-2015	4	980m	369,553

1.6 Existing Methods

There are a number of methods used, and suggested for the economic dispatch of power in an electrical grid. Methods 1-5 are known as conventional methods, and 6-7 are intelligent methods:

A1. Newton Raphson is a commonly used method for power flow, it works well in transmission systems and has a fast, quadratic exact convergence to the overall minimum. The method uses sequential linearisation and depends on an initial guess to find a power flow solution.

The algorithm is adapted to power flow by writing the power flow equations from Equations (12) and (13) as $P_*(\boldsymbol{x})$ and $Q_*(\boldsymbol{x})$:

$$\boldsymbol{x} = \begin{bmatrix} \theta_2 \\ \vdots \\ \theta_n \\ V_2 \\ \vdots \\ V_n \end{bmatrix}, \quad f(\boldsymbol{x}) = \begin{bmatrix} P_2(\boldsymbol{x}) - P_2^G + P_2^D \\ \vdots \\ P_n(\boldsymbol{x}) - P_n^G + P_n^D \\ Q_2(\boldsymbol{x}) - Q_2^G + Q_2^D \\ \vdots \\ Q_n(\boldsymbol{x}) - Q_n^G + Q_n^D \end{bmatrix}, \quad \boldsymbol{x}^{a+1} = \boldsymbol{x}^a - J(\boldsymbol{x}^a)^{-1} f(\boldsymbol{x}^a). \qquad (20)$$

Here θ_1, V_1, P_1, and Q_1 refer to the variables of the slack bus (node), this is the node that is used to absorb any excess power in the network to maintain Kirchhoff's Law. J is the Jacobian matrix. The equation shown uses an initial guess, $\boldsymbol{x}^0$, to calculate the subsequent iterations; $\boldsymbol{x}^1$ is the first iteration and this is continued until convergence, $|f(\boldsymbol{x}^a)| < \epsilon$, where ϵ is a small predetermined accuracy level [52].

A phenomenon has been found that in the distribution stage of power flow, due to the high Resistance:Reactance ratio, initial conditions outside the convergence area (basin of attraction) will result either in a local optimum or in non-convergence. Research has been done to overcome these problems, such as; using Newton Raphson based on current injection [53] which bypasses the difficult process of computing the Jacobian at every iteration, and instead computes the whole Jacobian once and recalculates the diagonals at each iteration, the Broyden method [54] which computes the whole Jacobian once and uses a rank-one update for the following iterations and a forwards-backwards sweep [55] which updates the upstream and downstream of each current from according to Kirchhoff's Law. The downside of these approaches includes slower convergence, local convergence, inferior precision, and higher complexity when considering PV nodes.

A2. Gradient descent is a commonly used dynamic method which uses the objective function in conjunction with Lagrange multipliers to enforce operational constraints

e.g.: $\mathcal{L}(u,\mu,\lambda) = f(u) + \lambda R(u) + \mu S(u)$. The function is differentiated to indicate the direction of steepest ascent and an adjustment is made in the opposite direction. This is repeated until $\nabla\mathcal{L}(u,\mu,\lambda) \approx 0$. The method usually has a linear convergence rate, however when the objective function is irregular, convergence can be slow in areas where $\nabla\mathcal{L} = 0$, and only local minima will typically be found. The adjustment parameter with respect to the computed gradient $\nabla\mathcal{L}$ directly influences convergence properties; if it is too small the algorithm may only converge to local minima or take too long to converge, however if it is too large, the iterations may result in oscillations around a global minimum. Another drawback is that the functions must be differentiable. The method has adjustments and variations which include the Differential Algorithm (DSD) [56] and the Least Mean Square (LMS) [57] however these do not guarantee an optimal solution and still have slow convergence rates, respectively.

A3. The Interior Point method (otherwise known as the Barrier method) has been widely considered [58, 59], it reaches an optimal solution by following the interior of the feasible region, it uses linear methods to satisfy constraints and a heuristic strategy of predictor-corrector to minimise a cost. The method is advantageous as it can solve linear and non-linear convex optimisation problems, it does not appear to be largely effected by network size and it can handle inequality constraints well by using barrier functions. Reference [35] describes it as the best technique for ACOPF. These methods comprise of two main interior point based algorithms, Infeasible Pure Primal-Dual algorithm (PDIPA) and Infeasible Primal-Dual Predicator-Corrector algorithm (PCIPA). PCIPA is advantageous in terms of computation time if the higher-order corrector variables are quadratic, in simple problems PDIPA is typically faster [60]. The number of iterations is only slightly dependent on the system size, with an order of approximately $\mathcal{O}(N)$. Drawbacks of interior point method include: 1) If the step size is not chosen properly, the sub-linear problem may have a solution that is inaccessible in the original nonlinear domain. 2) They also tend to have sensitive initial and termination criteria resulting in non-linear functions not being solvable for many choices.

A4. Linear programming is a widely-used optimization technique that linearises the objective function and constraints and finds a solution to the linear optimum. Linearising the complex problem introduces inaccuracy as the linearised functions may not be a good representation of the original ones, although it offers a guaranteed solution. One solution to this is using piecewise linear approximations, the more segments used the more accurate the method becomes; however, the number of segments does affect the speed and precision of the obtained solutions [61]. Another

solution is to minimise the deviation of control variables and although this can be effective, it slows the rate of convergence. Due to its simplistic approach, linear programming usually only considers DCOPF. Recent literature suggests that with additional approximations, such as the piecewise linear representation of the cosine curve, the algorithm is able to include reactive power and voltage magnitudes for ACOPF, this is called LPAC [62], they suggest that it works well with "hot-starts" (where a solved AC solution is already available and only adjustments are made), but described it as a building block to be used in conjunction with other optimisation techniques for "warm" and "cold-starts", where initial conditions and solution are further away.

A5. Quadratic programming is another commonly-used optimization technique. It is based on using a quadratic objective function and aims to provide an exact solution; for non-convex quadratic programming problems defined by a specific type of (Metzler) matrices an exact solution can be found through semi-definite programming relaxation [63]. A noise control is added to the Metzler matrices to maintain non-zero off-diagonal elements at all times. Quadratic programming also guarantees convergence, it is fast, reliable and efficiently-coded optimisation engines are already available for use in electricity-grid problems. However, constraints need to be linearised and if they are not fully obeyed, only suboptimal solutions are found, which can also happen if the approximation to a quadratic objective function is unrepresentative of the true objective for some parameter values. Nonlinear programming is a similar concept but without being constrained to quadratic objective functions it does not guarantee convergence to a global optimum; these methods are not robust and can be slow.

A6. Genetic Algorithms are based on the concept of natural selection and genetics, they search a rugged cost landscape by evaluating many configurations in a space in parallel, this reduces the chances of converging to a local optimum [64]. The objective function is encoded to binary codes, then a fitness (cost) function is calculated; the operations: reproduction, mutation and crossover are applied to produce offspring and repeated until the best offspring is found. It does not require the objective function to be continuous or differentiable and does not depend on a nearby starting point to find a global solution. The algorithm uses probabilistic evolution rules and assesses the objective function at parallel points. It is heuristic and not well-understood; difficulty in specifying the convergence criteria due to the stochastic nature of the algorithm, makes the choice of termination criteria problematic and so after a certain number of iterations, the best of

the set of offspring is chosen [65]; early termination could result in a suboptimal solution and performance has been found to be deficient in [66].

A7. Particle Swarm Optimisation (PSO) [67] combines social behaviour and evolutionary computations inspired by the behaviours of schools of fish and flocks of birds. The notion is that each particle can move around in multi-dimensional space inside a set of constraints. It starts as a uniform, global search and results in a more specific, local search at later stages. The method uses two main criteria; social-only which considers the belief of other particles and cognition-only which considers the individual particle separately. It is a simple heuristic concept which is claimed to be computationally efficient, flexible to adaptation and easy to implement. It is less likely than gradient methods to converge to local minima, it does not require approximations, nor does it require the function to be differentiable or convex, or require specific initial conditions. The method is heuristic and performance analysis is anecdotal in nature; one criticism is that local optimums remain likely and convergence is slow in high dimensional spaces, [68] suggests a method which allows particles to jump out of local optimums but increases computational time and is largely unexplored.

A8. Other heuristic methods include Ant Colony optimisation [69], Shuffling Jumping Frog algorithm [70], Cuckoo Search algorithm [71], Smart Evolutionary algorithm [72] and the Artificial Neural Networks model [73].

Other approximation methods which could be considered are Variational methods such as the Expectation Maximisation algorithm but they have an inherent bias towards simpler models [74]. They rely on having a limited number of parameters to be optimised which does not necessarily suit this type of global optimisation problem and computational complexity increases quickly with the dimensionality.

A9. Message passing or belief propagation (BP) is going to be discussed at length in the remainder of the thesis. The approach is principled which means it is based on solid probabilistic foundations which lead, through a simple derivation and clear approximations, to the algorithm. It passes conditional probability messages locally around the system using received conditional messages and priors; this gives a computational complexity of order $\mathcal{O}(N \log N)$ according to the system size, instead of global optimisation techniques which usually increase quadratically or cubically with the system size, and this quality allows the method to quickly find an optimal solution, even on large networks. Message passing gives good approximate solutions on sparse networks, which suits the topology of power grids, and its probabilistic properties will be shown to be an advantage when considering fluctuations and uncertainty of consumer

load and renewable generators. The method has been used for many optimisation problems as it is easily adaptable and can consider a range of objective functions, constraints and parameters.

These are some of the methods suggested and used for deterministic optimal power flow. The most efficient and commonly used is the interior point method. Section 3.1.4 will expand on some of these methods and how they have been adapted to consider the stochastic nature of renewable sources.

1.7 Aim

Existing methods for OPF and power flow with volatility are good, but have shortcomings. The aim of this thesis is to suggest a more suitable method for power distribution in electricity networks with volatility at both supply and demand nodes.

We suggest that the algorithm should be probabilistic to accommodate the volatility and account for correlations between components. It should be able to consider emergent behaviours and as the UK power grid is large, its computational complexity should increase approximately linearly with the system size; this will be vital for the program to converge quickly within the appropriate time-scale.

According to these specifications, we suggest a statistical mechanics inspired algorithm called Message Passing (MP) [75, 76]. This approach can inherently consider the uncertainties and fluctuations of generators and consumers within a grid at the economic dispatch stage. As a probabilistic algorithm, message passing offers a principled, distributed optimisation method which passes messages locally to obtain a global optimum; this allows for the computational complexity of the method to be of $\mathcal{O}(N\ln N)$ where N is the number of vertices in the grid and $\log N$ is related to the length of loops in a network. A global minima can be guaranteed through the consideration of a general T within the VFE calculations, a finite temperature quantifies increased information flow, as opposed to a steady state description often identified with a zero-temperature description, ensuring that the entire system is considered as to guarantee a global minima [77]. It can optimise a large number of parameters without becoming too costly computationally and while remaining effective; its distributive nature is particularly suitable for large or evolving networks. Additionally, this fast and reliable method has huge potential for adaptation. This thesis highlights the capabilities of message passing as a form of power

distribution in electricity grids, with the hope that further work would produce appropriate practical algorithms.

1.8 Statistical Mechanics for Hard Problems

As this problem requires an approach which can inherently consider probabilities, be suitable for large systems and use a principled method to find an optimal solution for electricity networks, a statistical mechanics-inspired approach is suggested. Statistical mechanics is used to understand the global properties of a system from its microscopic properties. It proposes that if you minimise the *free energy* of system, its most probable state would be found, which is also the optimal one when the objective function considered is the Hamiltonian. Statistical mechanics has previously been used for tasks such as decoding in the area of error correcting codes [78] and hard computational problems [79] and has been very positively received, but until recently focus has mostly been on discrete variables (the advantage to an optimisation problem with discrete states is that the possible combinations is finite), as the free energy is a complicated function in many cases, and the consideration of continuous variables make it even more difficult to compute (with continuous variables the combinations of sets of variables are infinite and therefore may not be found through exhaustive methods).

Resource allocation has been considered in computer science [80] and operations management [81] for many years. It has also been used to reduce internet traffic congestion, and streaming network flow [82]. Previously, a form of probabilistic modelling was suggested for OPF methods but dismissed as it was based on too simplistic assumptions. For instance, reference [83] suggested evaluating the propagation of data inaccuracies by using load flow calculations, the latter provides a nonlinear system which describes the energy flow through each edge. It was not feasible to use load flow calculations for AC power flow but it can handle a linear DC power flow. This thesis aims to make probabilistic modelling more relevant by finding ways to deal with the fluctuating properties of capacity using message passing techniques.

A message passing method described by Wong and Saad [76] suggests that instead of sending messages for continuous variables (e.g., functions), which were difficult to handle, one can send the first and second derivatives of the function, allowing the function to be represented by the first two terms of its Taylor expansion. They found that the method was successful and effective

as long as the message passing function could be described sufficiently accurately by the first few terms in its expansion. The message passing method is designed so that messages are passed locally to obtain, after convergence, a global optimum; this uses the assumption that since the network is sparse, the network can be considered locally as a tree and ignore recurrent messages because the probability of loops is negligible and can be ignored. Some studies investigating the role of loops [76] have found that even in the presence of some loops the algorithm converged to near optimal solutions over a broad range of parameters. As the global optimisation problem is solved by looking at the network locally, the computational complexity of the algorithm only increases as $\mathcal{O}(N \ln N)$ with the system size N, while still finding the global optimum. The message passing method is very similar to the probabilistic modelling algorithm of BP and the two are usually used as synonyms. The advantages of using this message passing approach are that it is inherently probabilistic and is able to consider the fluctuations within the network. As the computational complexity only increases $\mathcal{O}(N \ln N)$ with the system size, the algorithm can cope with large networks such as the UK electrical grid. Finally, message passing has the capacity to include other extensions, such as identifying possible instabilities in the grid for design, prevention and protection; it can also be used together with information visualisation methods to aid electricity-grid controllers, and theoretically it could work successfully with ACOPF.

The aim of this project is to suggest and develop a message passing approach to demonstrate its capabilities as a future OPF algorithm. This work was done in collaboration with ALSTOM UK.

1.9 Layout of Thesis

The proposed method of optimal distribution, message passing, will be explained in more detail in Section 2 and its relevance to the electricity grid setup will be explained in Section 2.4.4. Section 3 will discuss how the message passing algorithm can accommodate fluctuations and volatility while Section 4 will focus on how to distribute electricity when failure is inevitable by minimising load shedding. Section 5 explains the use of message passing for minimising the cost of generation as an objective function followed by Section 6 which explains how these calculations can be adapted to consider adjusting voltages for current to flow instead of power over lines. Each of these will contain their own results generated from simplified IEEE buses and synthetically generated systems. Section 7 will discuss the results of this research and conclude the thesis and in Section 8 we will suggest related work that could be addressed in future work.

2 Message Passing

2.1 Belief Propagation

Belief propagation (BP) (or sum-product message passing) was first proposed [84] in 1982. The inference method could find exact solutions to hard computational problems on tree-like graphs (graphs without loops) and it was later extended to find the approximate solution for general graphs (graphs containing loops). It suggests the use of local message passing to find an optimal global solution, an approach that is equivalent to dynamic programming.

Assume a set of variables (triangles) which can each take a value drawn from a set of discrete values, or *states*. Variables interact with one another and this interaction between variables can be described using a *factor* (squares) as shown in Figure 2.1. Each variable can assume a state, but its state is dependent on the states of other variables around it. The aim is to find the best solution of what state each variable should take and this is done through messages sent between the variables and the factors. The BP technique is usually visually described by rearranging the network into two columns of nodes on one side, and factors on the other, this is called a bipartite graph (Figure 2.2).

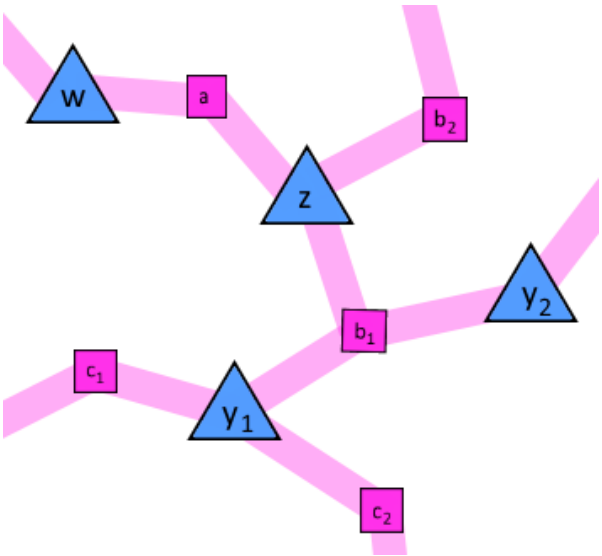


Figure 2.1 Some variables w, z, y_1 and y_2 (blue) and their interactions a, b_1, b_2, c_1 and c_2 (pink).

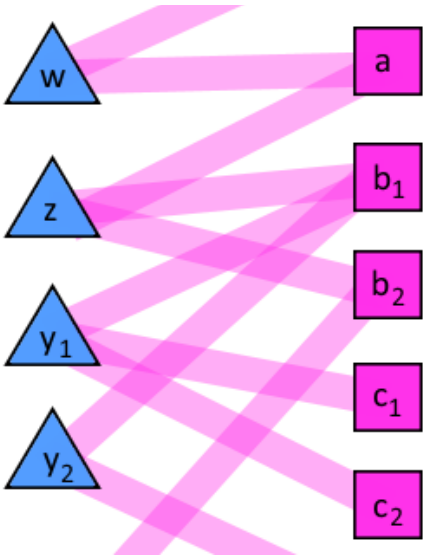


Figure 2.2 Example of the set of variables and interactions from Figure 2.1 rearranged, demonstrating a bipartite graph.

BP can be used to estimate the most likely state of a network, or the *probability marginals*, by sending messages locally, through the factors, between local variables (any variables it shares

an interaction with, or *neighbours*). Messages in BP are conditional probabilities, they can be from a variable to a factor, or from a factor to a variable (Both variables and factors can be attached to multiple others). Over a number of iterations, the passed messages converge to a consensus and a belief of the optimal state of each node can be calculated.

To start with we will explain the conditional probability message from a variable z to a factor a, about the probability of variable z being in state s_z, according to all the information it has recieved from its other interactions, but not a.

$$
\begin{aligned}
m_{z\to a} &= P\left(s_z \middle| \{t_b\}_{b\in\partial z_a}\right), && (21)\\
&\propto P\left(\{t_b\}_{b\in\partial z_a} \middle| s_z\right) P(s_z), && (22)\\
&\propto P(s_z) \prod_{b\in\partial z_a} P(t_b|s_z), && (23)
\end{aligned}
$$

where $m_{z\to a}$ is the message from variable z to factor a about the state s_z, t_b is the information at factor b, and this is for all factors connected to variable z, but not factor a (denoted $b \in \partial z_a$). Equation (22) is derived from Bayes theory (without normalization). In networks where the interactions are sparse, an assumption can be made that conditional probabilities are independent of one another; allowing them to be considered separately as shown in Equation (23). The assumption is true for all tree-like networks and gives an exact solution, however for graphs with loops, the assumption of independence may not be correct, giving only and approximate solution. Sparse graphs have been found to have a good approximate solution due to the sparsity and size of any loops.

The conditional probability messages from the factors to the variables can also be written as:

$$
\begin{aligned}
m_{b\to z} &= P(t_b|s_z), && (24)\\
&= \sum_{s_{y\in\partial b_z}} P\left(t_b \middle| s_z, \{s_y\}_{y\in\partial b_z}\right) P\left(\{s_y\} \middle| \{t_c\}_{c\in\partial y_b}\right), && (25)\\
&= \sum_{s_{y\in\partial b_z}} P\left(t_b \middle| s_z, \{s_y\}_{y\in\partial b_z}\right) \prod_{y\in\partial b_z} P\left(s_y \middle| \{t_c\}_{c\in\partial y_b}\right), && (26)\\
&= \sum_{s_{y\in\partial b_z}} P\left(t_b \middle| s_z, \{s_y\}_{y\in\partial b_z}\right) \prod_{y\in\partial b_z} m_{y\to b}, && (27)
\end{aligned}
$$

where $m_{b\to z}$ is the message from b to variable z expressing "what factor b thinks about z being in state s_z" and it is calculated according to all of z's neighbours. It is calculated by the probability of t_b given that z is in state s_z and that the neighbouring variables y, are in state s_y

multiplied by the condition probability of y being in state s_y (according to all of the other factors connecting to y). Equation (26) uses the assumption that the factors are independent, and the product of conditional probabilities can be recognised as the messages from variables y to factor b.

This results in the two messages:

$$m_{z\to a} \propto P(s_z) \prod_{b\in\partial z_a} m_{b\to z}\,, \tag{28}$$

$$m_{b\to z} = \sum_{s_{y\in\partial b_z}} P\left(t_b \middle| s_z, \{s_y\}_{y\in\partial b_z}\right) \prod_{y\in\partial b_z} m_{y\to b}\,, \tag{29}$$

where each new message uses the latest information it has received from its connecting variables/factors. For continuous variables, the messages become functions and the sum is replaced by an integral.

There are a few unique cases where this can be simplified; in the event of a network where nodes are the variables and interactions are modelled by edges (a connection between two nodes), then every factor is connected to only two variables (Figure 2.3 and Figure 2.4), resulting in a simplified message from factor b to variable z:

$$m_{b\to z} = \sum_{s_y} P(t_b|s_z, s_y) m_{y\to b}. \tag{30}$$

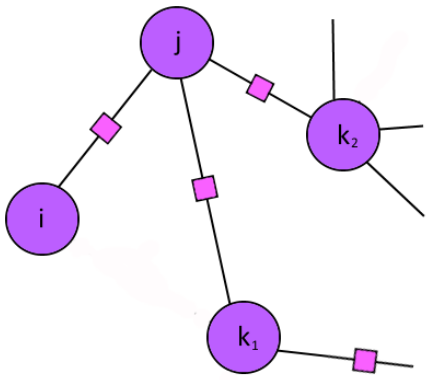


Figure 2.3 Example of part of a network, where variables are on nodes, and interactions are only between nodes connected by an edge [158].

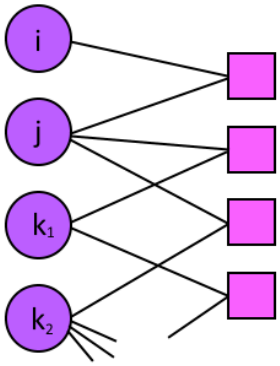


Figure 2.4 Example of the network from Figure 2.3 as a bipartite graph [158].

This allows the factor to be considered only as probability of some function of the interaction $g(s_z, s_y)$ and resulting in a message straight from node z to node w, essentially bypassing the factor. This is displayed through Figure 2.7-Figure 2.7.

$$m_{z\to w} \propto P(s_z) \prod_{y\in \partial z_w} \sum_{s_y} P(s_w|s_z, s_y) \prod_{y\in \partial z_w} m_{y\to z}. \tag{31}$$

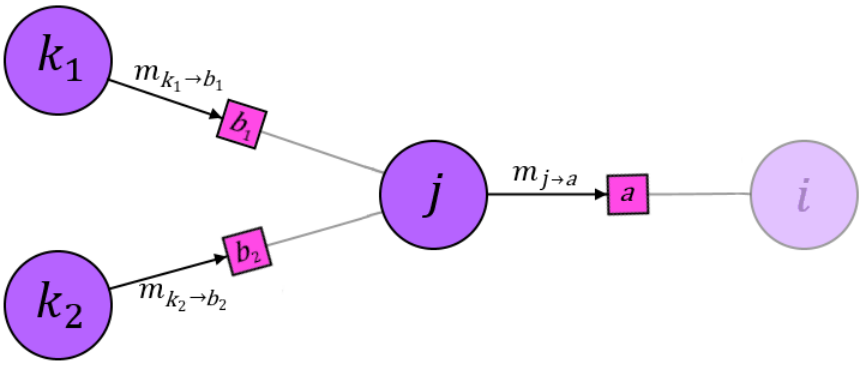


Figure 2.5 An example of messages from child variables to factors.

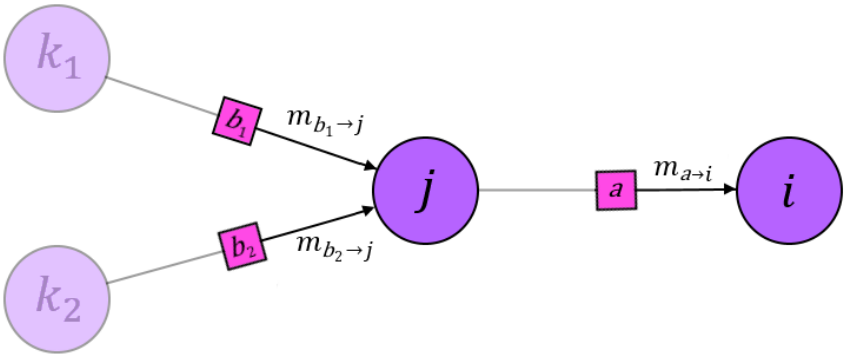


Figure 2.6 An example of messages from factors to parent variables.

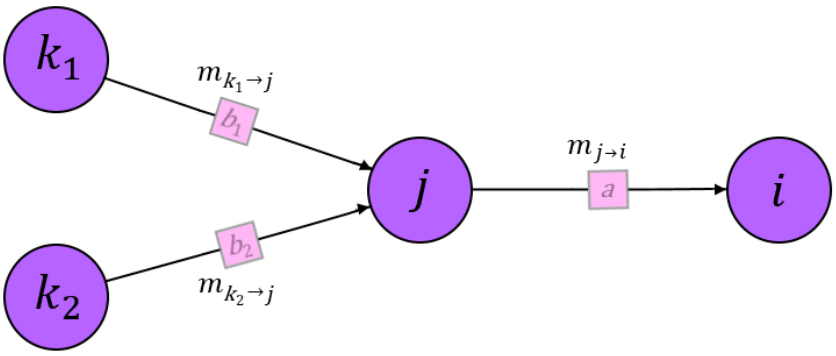


Figure 2.7 An example of messages from descendant nodes k to j, and one message from node j to i, bypassing the factors due to the specific network with interactions only connecting two variables.

Similarly, as will be discussed throughout the thesis, the variables could be the values of the edges instead (the connections between nodes in a network). In this case, every variable is only connected to two factors, but factors can still be connected to multiple variables (such that the nodes were the factors) and an identical equation is found.

The messages calculated from Equations (28) and (29) are iteratively passed on for the next message; either synchronously (each time step includes every connection being updated according to the previous time step's set of messages) or asynchronously (each time step represents one message sent and uses the most recent messages from other connections). The recursive message is calculated for every connection in both directions until all messages have converged to an agreed set of conditional probabilities. Once converged, a belief (or marginal distribution) of each variable can be calculated by considering the message information passed from all connecting factors and the prior knowledge of the variable; the belief determines the optimal state of the variable. The probabilistic marginal for variable z can be shown as:

$$b(s_z) \propto P(s_z) \prod_{b \in \partial z} P(t_b|s_z), \tag{32}$$

$$\propto P(s_z) \prod_{b \in \partial z} m_{b \to z}, \tag{33}$$

where the first term signifies any priors, or evidence about the probability of the variable being in state s_z, and this is multiplied by the likelihoods of z being in state s_z as stated by *all* neighbouring factors.

2.2 Statistical Physics

The same framework of messages and marginal approximations can be found through statistical physics techniques similar to the probabilistic modelling approach and the links between the two have been clearly identified [85].

Statistical physics allows us to describe the typical behaviour of the macroscopic state of a system and to analyse the relations between variables from the microscopic properties [86] in some systems. Existing derivations of the following equations are explained in detail in Appendix A.

From this point and onward, I will restrict the discussion to systems that are sparsely connected, i.e., where each variable interacts only with a small number of other variables. The general energy function, E, for a system with two variable interactions can be expressed in terms of the nodes and edges of the network where the i and j indices run over all system variables:

$$E = \sum_{ij} \phi_{ij} + \sum_{i} \psi_i. \tag{34}$$

This is termed in statistical physics as the Hamiltonian of a physical system, where ϕ and ψ are functions containing objective functions, prior beliefs (external fields) and/or constraints on the edges and nodes, ij indicates interactions between nodes i and j.

The Boltzmann distribution tells us the probability of the system being in a certain state u:

$$P_u = \frac{1}{Z} e^{-\beta E_u}, \tag{35}$$

where the normalization constant is the partition function:

$$Z = \sum_{u} e^{-\beta E_u}, \tag{36}$$

$\beta = \frac{1}{T}$ and T is the temperature, a parameter that determines the level of separation between probabilities relating to different free energy values. E_u is the energy of the system in state u. The free energy can be linked to the probability of a node being in a certain state.

$$F = -T \ln Z\,, \tag{37}$$

$$F = -T \ln \left\{ \sum_{u} \exp \left[-\beta \left(\sum_{ij} \phi_{ij} + \sum_{i} \psi_i \right)_u \right] \right\}. \tag{38}$$

Minimising the free energy of a function allows us to find the most likely (and optimal, if the Hamiltonian corresponds to the objective function) state. The free energy in real systems also includes the multiplication of the Boltzmann constant, k_B; however, since our systems are

simulated, the temperature can be rescaled using energy units so that $k_B = 1$, as explained in [77].

It is argued in the physics literature that if a large network is sparse ($N \gg c$) then the probability of small loops is negligible. Implying, that if a node j was removed, all of its neighbours would be probabilistically independent of each other. Using this *cavity method* (hence the term), one assumes the network to be locally tree-like and hence allows for the probabilistic independence (or weak dependence) between neighbours. This theory allows us to focus on small parts of a network at a time, iteratively, instead of the whole network at once.

The focus of each iteration is one node j and all nodes it is connected to via an edge. Using the cavity method, we can assume each neighbour of j would be independent of one another if j was removed. This allows us to ignore any interactions beyond the neighbouring nodes, consequently ignoring the rest of the network and the network is assumed to be locally tree-like. By separately considering one of the neighbours of j as the *ancestor* i, and labelling the rest *descendants* k (as seen in Figure 2.8), we can pass information to node i about the probability of i being in a certain state, according to current information at the node j; the interaction between nodes i and j, and the information that j has received from the descendant nodes about the state of j. This is where the link between BP and Statistical Physics can be observed.

To begin with, we will demonstrate the method of obtaining an optimal solution where the control variables are continuous directional "currents" flowing over edges, denoted by y_{ij} (see Figure 2.9), where i and j are the nodes connected by the given edge, with constraints on edges and nodes, ϕ_{jk}^{CO} and ψ_j^{CO}, and objective functions on edges and nodes, ϕ_{jk}^{OF} and ψ_j^{OF}.

The information passed to the ancestor node is the free energy according to the sub-tree $\mathbf{T}_j$ (the information from node j and its descendants as seen in Figure 2.10), it is proportional to the probability of the edge ij being in a certain state, and when the global free-energy is minimised one finds the optimal or *ground* state of the system.

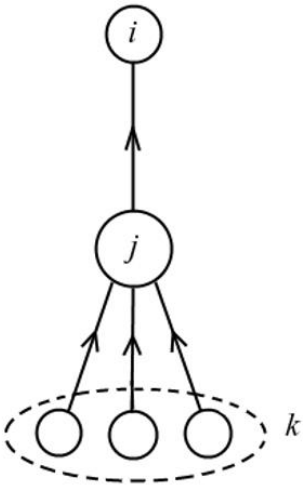


Figure 2.8 Considering a small section of the network and rearranging neighbours so one node appears as the *ancestor*.

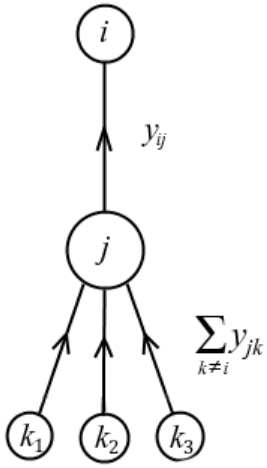


Figure 2.9 The current passed over edges from descendants to the parent node, j, and from node j to the ancestor.

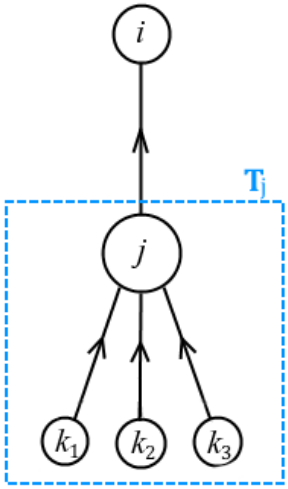


Figure 2.10 The tree of node j, $\mathbf{T}_j$.

The free energy can be written as:

$$F(y_{ij}|\mathbf{T}_j) = -T \ln \left\{ \sum_{jk} \left(\int dy_{jk} \right) \exp \left[-\frac{1}{T} \sum_{jk} F(y_{jk}|\mathbf{T}_k) + \phi(y_{jk}) + \psi_j \right] \right\}, \quad (39)$$

where $\phi(y_{jk}) = \phi_{jk}^{CO} + \phi_{jk}^{OF}$, $\psi_j = \psi_j^{CO} + \psi_j^{OF}$ and the integral of the continuous variable y_{jk} is equivalent to the sum of all states u. The free energy $F(y_{ij}|\mathbf{T}_j)$ can now be expressed as the product of the free energies of all the descendants, $F(y_{jk}|\mathbf{T}_k)$, in addition to constraints and objective functions, due to the tree-like assumption; we can assume all necessary information about the rest of the network is conveyed through the descendants' free energies, $F(y_{jk}|\mathbf{T}_k)$, allowing us to ignore the rest of the network. These equations can be translated directly to the BP formulation [76].

The free energy can be split into two parts, the average free energy per variable in the network, F_{av}, and the Vertex Free Energy (VFE), F_V, which gives the individual free energy contribution of the variable considered:

$$\left(y_{ij}\middle|\mathbf{T}_j\right) = N^{\boldsymbol{T}}F_{av} + F_V\left(y_{ij}\middle|\mathbf{T}_j\right), \tag{40}$$

where $N^{\boldsymbol{T}}$ is the number of variables in the tree $\mathbf{T}$.

$$F_V\left(y_{ij}\middle|\mathbf{T}_j\right) = -T\ln\left\{\prod_{jk}\int dy_{jk}\exp\left[-\beta\sum_{jk}F_V\left(y_{jk}\middle|\mathbf{T}_k\right) + \phi_{jk} - \beta\psi_j\right]\right\} - F_{av}, \tag{41}$$

where $F_V\left(y_{ij}\middle|\mathbf{T}_j\right)$ is the VFE function. The message passing algorithm is employing the VFE as the message passed, e.g., to node i from j given $\mathbf{T}_j$, the tree emanating from node j. The variable y_{ij} indicates the current (state) sent from node j to i; it is directional and so $y_{ij} = -y_{ji}$. The temperature determines how strictly the low free energy states are selected, calculations of the VFE assume a general temperature T in order to consider the global network and are then considered under strict optimisation ($T = 0$) to find the global optimum, giving a recursion relation:

$$\begin{gathered} F_V\left(y_{ij}\middle|\boldsymbol{T}_j\right) = \min\left[\sum_{jk}\left(F_V\left(y_{jk}\middle|\mathbf{T}_k\right) + \phi_{jk}^{OF}\right) + \psi_j^{OF}\right] - F_{av} \\ \text{such that } \sum_{jk}\phi_{jk}^{CO} \text{ and } \psi_j^{CO} \text{ are satisfied.} \end{gathered} \tag{42}$$

This is the continuous function "message" to be iteratively sent between nodes until convergence.

Using equation (31) we can map the belief propagation messages to the vertex free energy. If we first take the prior at variable z, which gives the probability of the variable being in a certain state s_z, as:

$$P(s_z) \propto e^{-\frac{\beta}{2}\phi(y_{ij})}, \tag{43}$$

And the conditional probability $P\left(s_w\middle|s_z, s_y\right)$ as the step function, which calculates, according to the descendants and information at the current variable, if the node can be satisfied for values of y_{ij} (s_w):

$$P(s_w|s_z, s_y) \propto \Theta\left(\sum_{jk} y_{jk} - y_{ij} + \Lambda_j\right), \tag{44}$$

Then the comparison can be found that $-T \log m_{z\to w} = F_V(y_{ij}|\boldsymbol{T}_j)$, where the function is integrated over the possible instances of the variable y_{jk} instead of summed, and the messages from descendants give $-T \log m_{y\to z} = F_V(y_{jk}|\boldsymbol{T}_k)$. A zero temperature limit is equivalent to a belief propagation scheme without memory whereas a finite temperature will correspond to a "hysteresis" effect, something that is being ignored in our present analysis.

2.3 The History of Message Passing

Message Passing is a class of algorithms which aim to find the probability distribution of the state of variables by passing messages locally over the edges of a factor graph. By sending a series of messages iteratively until a consensus is achieved, the algorithms obtain the most likely, or optimal, global solution. These algorithms are closely related with statistical physics concepts and are typically used for probabilistic graphical models.

BP is a type of discrete message passing which computes the probability marginals for tree-like graphs. The algorithm displays huge potential and can be used to draw random configurations from a distribution and compute the free entropy [77] but limitations include no guaranteed convergence, the possibility of converging to local maxima and that the algorithm only guarantees an exact solution if the network is tree-like (no loops). Ref. [84] suggests that in the presence of loops, BP could become stuck in a cycle and not converge or that an exact solution may not be found, but further research [87] shows how loopy graphs would still converge to "surprisingly good" approximate solutions even in the presence of many small loops. Loopy BP was able to optimise error-correcting codes such as low-density parity-check codes [88] and turbo decoding [89], although there remains a small risk of oscillation instead of convergence. Other extensions of BP include; generalised BP, which allows for larger regions to be considered in iterations for more accurate results and faster convergence, this is justified using a statistical physics concept called the "Kikuchi Approximation" which approximates the energy as the sum of all the energies of basic clusters, minus the energies of the over-counted intersections [90, 91], fractional BP which aims to acknowledge the cycles in the network, instead of computing as if there were none, by using a family of approximate free energies [92],

convex-concave procedure [93] which considers the objective function as a combination of convex and concave functions and plots the gradient of each one separately and calculates the next update accordingly, survey propagation [94, 95] which calculates approximate marginals, then iteratively assigns extreme probability values to variables to find satisfiable cases, affinity propagation [96] which establishes clusters by finding a suitable node to represent each cluster and has been successfully used to cluster images of faces, detect efficiently accessed airports and identify sentences in a manuscript. Message passing algorithms have been successful in finding good solutions in systems comprising of discrete variables such as; hard computational problems [97], error correcting codes [78, 98], probabilistic inference [99], Boltzmann tree structures [100] and real variable systems with a reduced number of parameters [101].

The aforementioned algorithms have typically concentrated on discrete variables. Work done on continuous messages includes using eigenfunction decomposition [102], expectation propagation which uses estimates of beliefs in terms of the mean and standard deviation until convergence [103] and Gaussian BP [104] which substitutes Gaussian distributions into continuous BP updates which has been used for machine learning [105] and connecting genotypes and phenotypes for greater understanding of underlying diseases [106]. Wong and Saad [75, 107] suggest a way of passing continuous messages which used edge variables. They suggest that instead of passing the full VFE function as a message, it can be accurately represented by its first and second derivative in the Taylor expansion if approximated by the Gaussian distribution. This is discussed in more detail in [76] and the results show that an exact solution can be found if the costs involved are of a quadratic form and the VFE assumes a Gaussian distribution; in other cases, where the VFE is not Gaussian, it still gives a good approximation.

2.4 Message Passing with Continuous Variables

2.4.1 Method

Since the derivation of Wong and Saad [76] will feature highly in the remainder of the thesis, we will review it in greater detail. They assumed a network of N nodes, each connected to c other nodes (though they suggested that c does not have to be homogeneous) in a sparse network. The given information on the network is a capacity at each node, Λ_j, indicating the available resource (which can be positive or negative), the constraint is that each capacity needs

to be made non-negative by the directional flow of current, y_{ij}, over edges to or from the nodes and the objective function is to minimise the cost of the flow of current, where the flow over the edges are the variables. It is a generic resource allocation problem.

The energy function used takes the form:

$$E = \sum_{(ij)} \mathcal{A}_{ij}\phi(y_{ij}) + \sum_{i} \psi(\Lambda_i, y_{ij}|\mathcal{A}_{ij} = 1), \tag{45}$$

where (ij) are all ordered pairs of nodes. The connectivity matrix $\mathcal{A}_{ij} = 1$ if an edge exists between nodes (ij) and $\mathcal{A}_{ij} = 0$ otherwise; ϕ is the transportation cost, which is a function of the current variable. In [76] they specifically looked at the case where $\phi_{ij} = \frac{y_{ij}^2}{2}$, but stressed that other objective functions could be considered; and indeed, some other types of functions were studied as well. The function ψ represents the performance cost; and this was considered to be a constraint, e.g., $\psi = \ln \Theta(\sum_i \mathcal{A}_{ij} y_{ij} + \Lambda_j)$ where $\Theta(\bullet)$ is the step function that returns 0 if $\bullet$ is negative, and 1 otherwise. The capacity of node j is Λ_j, this is a quenched variable, i.e., a given disorder variable, which indicates the demand or generation level (positive for generators, negative for consumers). Giving the condition:

$$\sum_{i} \mathcal{A}_{ij}\, y_{ij} + \Lambda_j \geq\ 0, \tag{46}$$

which demands that all nodes in the network will be *satisfied*; the constraint sums only the power sent along edges from neighbouring nodes to node j, and this total must be larger than or equal to the demand or smaller than or equal to the production at each node to satisfy the constraint.

Using the "locally" tree-like assumption, the free energy can be calculated recursively by considering at each point the free energy associated with node j, depending on the outgoing current y_{ij} and the tree below that node.

$$F(y_{ij}|\mathbf{T}_j) = -T \ln\left\{\prod_{k=1}^{c-1}\left(\int \mathrm{d}\, y_{jk}\right)\Theta\left(\sum_{k=1}^{c-1} y_{jk} - y_{ij} + \Lambda_j\right) \times \exp\left[-\beta \sum_{k=1}^{c-1}\left[F(y_{jk}|\mathbf{T}_k) + \phi(y_{jk})\right]\right]\right\}. \tag{47}$$

The VFE of y_{ij} describes the contribution to the free energy at a vertex j. Analysis using the Bethe approximation [108] and replica symmetry [107] result in similar expressions for the VFE:

$$\begin{aligned} F_V(y_{ij}|\mathbf{T}_j) = -T \ln\Bigg\{&\prod_{k=1}^{c-1}\left(\int \mathrm{d}\, y_{jk}\right)\Theta\left(\sum_{k=1}^{c-1} y_{jk} - y_{ij} + \Lambda_j\right) \\ &\times \exp\left[-\beta \sum_{k=1}^{c-1}\left[F_V(y_{jk}|\mathbf{T}_k) + \phi(y_{jk})\right]\right]\Bigg\} \\ &+ \left\langle T \ln\Bigg\{\prod_{k=1}^{c}\left(\int \mathrm{d}\, y_{jk}\right)\Theta\left(\sum_{k=1}^{c} y_{jk} + \Lambda_j\right) \times \exp\left[-\beta \sum_{k=1}^{c}\left[F_V(y_{jk}|\mathbf{T}_k) + \phi(y_{jk})\right]\right]\Bigg\}\right\rangle_{\Lambda_j,} \end{aligned} \tag{48}$$

where $\langle \bullet \rangle_{\Lambda_j}$ is the ensemble average of $\bullet$ with respect to Λ_j. At zero temperature Equation (48) reduces to:

$$\begin{aligned} F_V(y_{ij}|\mathbf{T}_j) = &\min_{\{y_k|\sum_{k=1}^{c-1} y_{jk} - y_{ij} + \Lambda_j \geq 0\}}\left[\sum_{k=1}^{c-1}\left[F_V(y_{jk}|\mathbf{T}_k) + \phi(y_{jk})\right]\right] \\ &- \left\langle \min_{\{y_k|\sum_{k=1}^{c} y_{jk} + \Lambda_j \geq 0\}}\left[\sum_{k=1}^{c}\left[F_V(y_{jk}|\mathbf{T}_k) + \phi(y_{jk})\right]\right]\right\rangle_{\Lambda_j.} \end{aligned} \tag{49}$$

The VFE is a continuous function and message passing is typically employed for systems with discrete variables. It was suggested [76] that if the VFE is Gaussian, then the first and second derivatives of the function would be sufficient to pass as messages instead of the full function, effectively expressing the first two terms of the Taylor expansion. If the VFE is not Gaussian, an approximate solution can be found. Therefore, to optimise the VFE we can utilise the Taylor expansion of the VFE, around a small adjustment, ε_{jk}, about y_{jk}:

$$_V = \sum_{k \neq i} \mathcal{A}_{jk} \frac{\partial F_V}{\partial y_{jk}} \varepsilon_{jk} + \frac{1}{2} \sum_{\substack{k \neq i \\ l \neq i}} \mathcal{A}_{jk} \mathcal{A}_{jl} \frac{\partial^2 F_V}{\partial y_{jk} \partial y_{jl}} \varepsilon_{jk} \varepsilon_{jl} + \cdots, \quad (50)$$

with a quadratic objective function, $\phi = \sum_{jk} \left(\frac{y_{jk}^2}{2} \right)$, any derivative of ϕ with respect to a variable y_{ja}, gives $\frac{\partial \phi}{\partial y_{ja}} = y_{ja}$; which is variable specific to the derivative. A second derivative will give: $\frac{\partial^2 \phi}{\partial y_{ja}^2} = 1$, or $\frac{\partial^2 \phi}{\partial y_{ja} \partial y_{jb}} = 0$ if $a \neq b$. Therefore, any summation where $k \neq l$ returns zero. The authors demonstrate that a quadratic function results in an exact solution as the VFE can be written precisely in terms of its first and second derivative, but that other objective functions can be optimised and give an approximate solution [76].

The component of the VFE that corresponds to the conditional probability message from *j* to i:

$$F_{ij} = \sum_{k \neq i} \mathcal{A}_{jk} \left[\left(A_{jk} + \phi'_{jk} \right) \varepsilon_{jk} + \frac{1}{2} \left(B_{jk} + \phi''_{jk} \right) \varepsilon_{jk}^2 \right], \quad (51)$$

where $A_{jk} = \frac{\partial F_{jk}}{\partial y_{jk}}$, $B_{jk} = \frac{\partial^2 F_{jk}}{\partial y_{jk}^2}$, $\phi'_{jk} = \frac{\partial \phi}{\partial y_{jk}}$ and $\phi''_{jk} = \frac{\partial^2 \phi}{\partial y_{jk}^2}$. Under the constraint that all nodes need to be satisfied (non-negative) according to all the resource coming in, any going out and any resource that the node starts with:

$$\sum_{k \neq i} \mathcal{A}_{jk} (y_{jk} + \varepsilon_{jk}) - y_{ij} + \Lambda_j \geq 0, \quad (52)$$

which will be considered using the Lagrange multiplier μ_{ij}. This function can be optimised to find A_{ij}, B_{ij} which are the messages passed to the ancestor node, i. Additional messages for the adjustments $y_{jk} \leftarrow y_{jk} + \varepsilon_{jk}$ are passed to the descendants.

Differentiating the localised component of the VFE, and a Lagrange multiplier μ_{ij} which incorporates the constraint, by ε_{jk} and making this equal to zero gives:

$$\varepsilon_{jk} = -\frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}. \quad (53)$$

Introducing the constraint as a Lagrange multiplier μ_{ij} the optimal solution is given by:

$$F_{ij}^{*} = \frac{1}{2}\sum_{k\neq i} \mathcal{A}_{jk} \frac{\mu_{ij}^{2} - \left(A_{jk} + \phi'_{jk}\right)^{2}}{B_{jk} + \phi''_{jk}}, \tag{54}$$

where,

$$\mu_{ij} = \min\left[\frac{\sum_{k\neq i} \mathcal{A}_{jk}\left(y_{jk} - \frac{A_{jk} + \phi'_{jk}}{B_{jk} + \phi''_{jk}}\right) - y_{ij} + \Lambda_{j}}{\sum_{k\neq i} \mathcal{A}_{jk} \frac{1}{B_{jk} + \phi''_{jk}}}, 0\right]. \tag{55}$$

The derivatives of F_{ij}^{*} with respect to y_{ij} lead to the forward messages A_{ij} and B_{ij} from node j to i:

$$A_{ij} \leftarrow -\mu_{ij}, \tag{56}$$

$$B_{ij} \leftarrow \frac{\Theta(-\mu_{ij} - \epsilon)}{\sum_{k\neq i} \mathcal{A}_{jk} \frac{1}{B_{jk} + \phi''_{jk}}}, \tag{57}$$

where ϵ is a small positive number. Note that in passing these when the descendant currents change continuously, both (54) and (55) also change as functions of $\phi(y_{ij})$ and its first and second derivatives.

As the messages are simplified to A and B, each node estimates the optimal current by approximating the VFE as a quadratic function. The remaining step is the determination of the drawn current y_{ij} at which the derivatives should be computed. Determining this *working point* is achieved by passing additional information-provision messages among the nodes, a step not present in conventional message passing algorithms.

One method proposed is, that when the messages are sent from j to i, backward messages y_{jk} computed from the same optimisation steps are sent from j to the descendants, informing them of the particular arguments to be used for calculating subsequent messages. The backward messages are given by $y_{jk} \leftarrow y_{jk} + \varepsilon_{jk}$, or:

$$y_{jk} \leftarrow y_{jk} - \frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}. \tag{58}$$

Alternatively, a forwards message could be sent to the ancestor and this is calculated by the energy difference from changing the edge y_{ij}:

$$E_{ij} = A_{ij}\varepsilon_{ij} + \frac{1}{2}B_{ij}\varepsilon_{ij}^2 + A_{ji}\left(-y_{ij} - \varepsilon_{ij} - y_{ji}\right) + \frac{1}{2}B_{ji}\left(-y_{ij} - \varepsilon_{ij} - y_{ji}\right)^2 + \phi'_{ij}\varepsilon_{ij} + \frac{1}{2}\phi''_{ij}\varepsilon_{ij}^2, \tag{59}$$

which, when optimised with respect to ε_{ij} gives:

$$\varepsilon_{ij} = \frac{-A_{ij} + A_{ji} - B_{ji}y_{ij} - B_{ji}y_{ji} - \phi'_{ij}}{B_{ij} + B_{ji} + \phi''_{ij}}, \tag{60}$$

resulting in a forward current of $y_{ij} \leftarrow y_{ij} + \varepsilon_{ij}$, or:

$$y_{ij} \leftarrow \frac{y_{ij}B_{ij} + y_{ij}\phi''_{ij} - A_{ij} + A_{ji} - B_{ji}y_{ji} - \phi'_{ij}}{B_{ij} + B_{ji} + \phi''_{ij}}. \tag{61}$$

Both work well for the quadratic cost function.

The updates of currents y_{ij} and y_{ji} in opposite directions of the same link allows one to use the criteria $y_{ij} = -y_{ji}$ to check for convergence. To test for the algorithms convergence, one requires $y_{ij} + y_{ji}$ to be less than a given threshold. The information-provision messages are also useful for monitoring the optimal cost function during simulations.

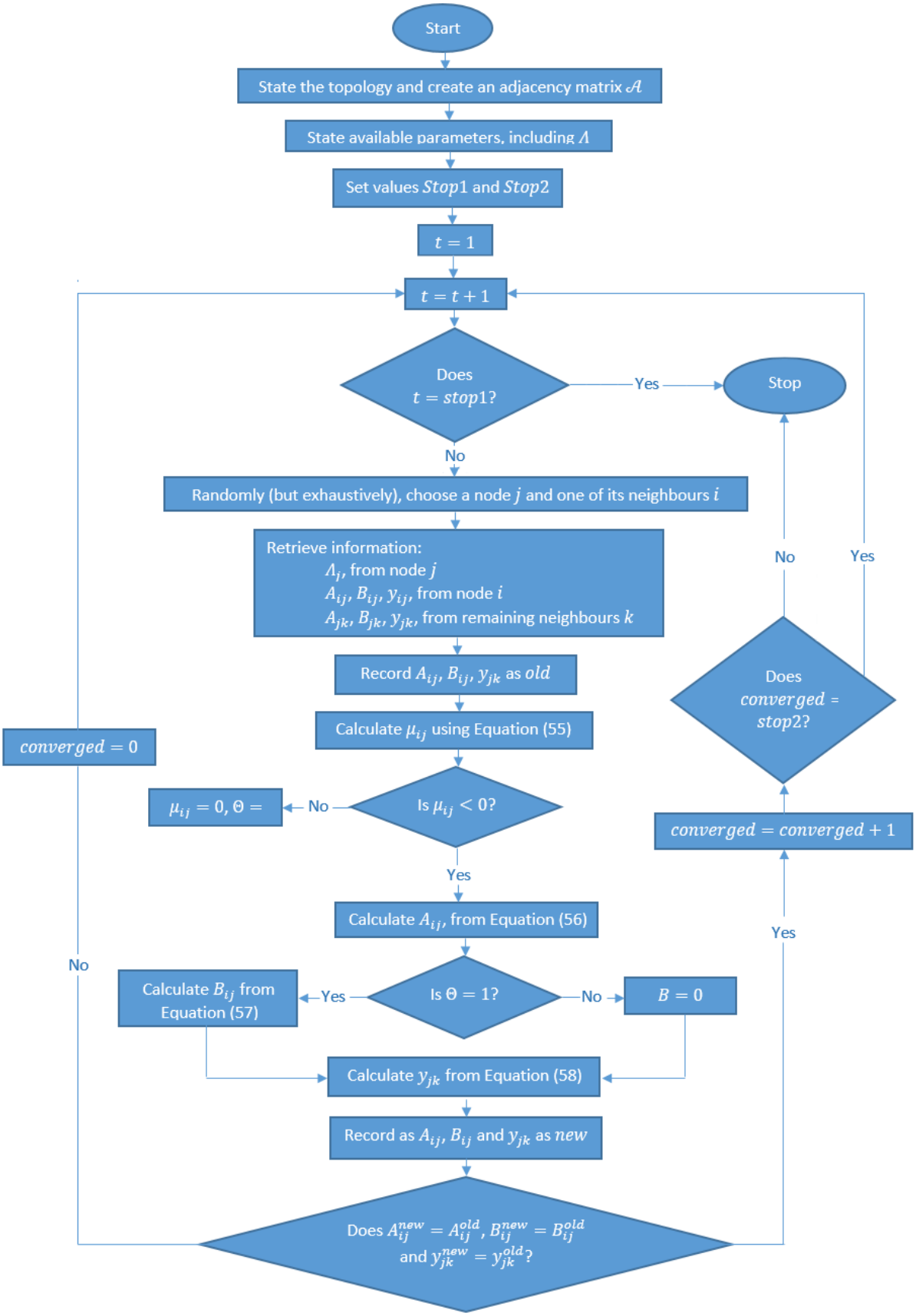
Using this message passing framework as a distributed optimisation algorithm was effective even on loopy networks. It has been found that the computational complexity increases with order $\mathcal{O}(N \ln N)$ in relation to the system size N, allowing the technique to find a global approximate solution within a reasonable time-scale, even for large networks. It has also been established that at the zero-temperature limit and with a quadratic objective function, the general VFE would be a piecewise quadratic with continuous slopes which allows for precise approximations and that the distribution in a network depends largely on the availability within the network and only loosely on the connectivity. In addition to message passing algorithm, Wong and Saad also considered the Kuhn-Tucker optimisation method which resulted in a similar algorithm to the message passing approach when considering a quadratic objective

function. The method gives solutions identical to the price iteration algorithm, based on recurrent corrections with respect to the objective function, which typically converges to a minimum, as long as the network is sparse. We can verify that at the steady state, $\mu_{ij} = -\phi'_{jk} - A_{jk}$ for all $k \neq i$, using Equation (58).

The second order does not play a part in the final solution but does help for a faster convergence. Wong and Saad find that the algorithm works well even on small networks and large well connected graphs.

Note: The method used throughout the thesis will aim to satisfy consumer requirements of sufficient resource in and electrical grid. From Section 2.4.4 flow over edges and resource at capacities will be considered in terms of power. In real electricity networks, Kirchhoff's Law states that the sum of all power must be equal to zero; in the analysis outlined below we will assume that any excess power either identifies power that the generator does not need to produce, or that can be lost through affine control (the primary and secondary control steps).

2.4.2 Flowchart of message passing process



2.4.3 Further Work on Continuous Message Passing

Other practical components of resource allocation include limiting flow over edges, considering resistance, softening constraints and being comparable to existing leading techniques. [109] and [110] show how this continuous message passing framework can accommodate such components.

Bandwidth

Bandwidth on an edge is the upper bound of current allowed to flow over the edge at any given time, in either direction. It can be applied to many real-life network problems, including; liquid through a pipe, vehicles on a road and power through a power line. Considering bandwidth in a network optimisation problem introduces discontinuous curves and this impacts on the differentiability of terms like $\frac{\partial F_{ij}}{\partial \mu_{ij}}$ at the cusp points. We will now describe the method suggested for considering bandwidth within a message passing algorithm [109].

Expanding upon the algorithm explained in Section 2.4, an explicit bandwidth constraint must be added:

$$-W \leq y_{jk} + \varepsilon_{jk} \leq W, \tag{62}$$

where W represents the maximum amount of current allowed to pass over any edge at any given time. The left-hand side of the inequality is due to the directional nature of y_{jk}, and the bandwidth is required to limit the adjusted current, $y_{jk} + \varepsilon_{jk}$; not just y_{jk}.

Given the previous derivation (35) this can be rearranged as:

$$y_{ij} - \Lambda_j = \sum_{k \neq i} \mathcal{A}_{jk} \max\left[-W, \min\left[W, \left(y_{jk} - \frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}\right)\right]\right]. \tag{63}$$

Using this, we can calculate the value of μ_{ij} by the intersection of the left-hand side with the piecewise linear graph representing the right-hand side of (63), as seen in Figure 2.11.

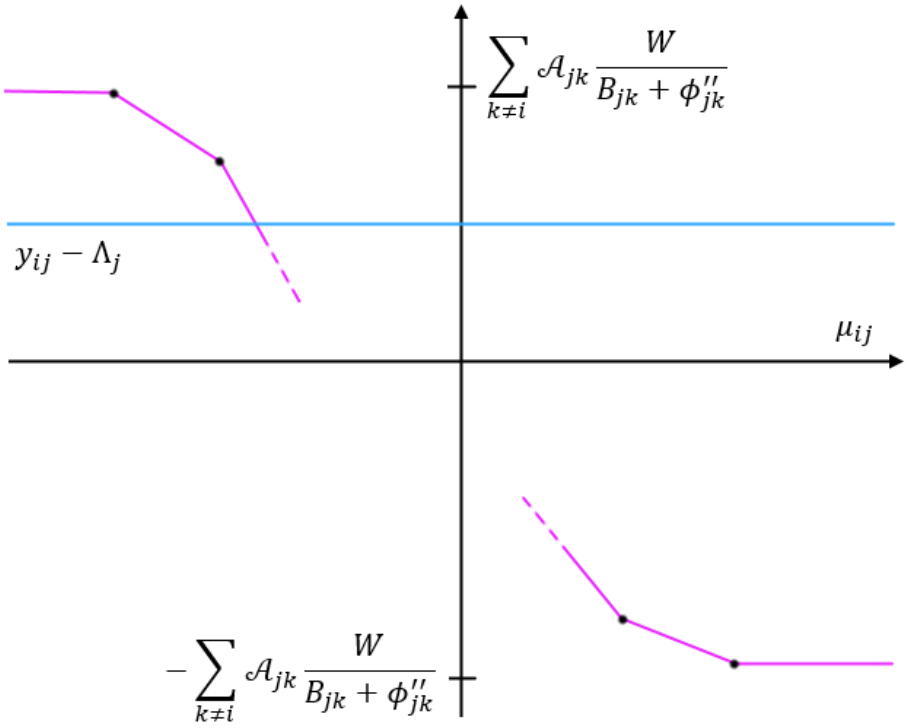


Figure 2.11 The piecewise linear representation of Equation (63).

Where the downward kinks happen when:

$$W = y_{jk} - \frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}, \tag{64}$$

and the x-coordinates at the point of downward kinks are:

$$\left(y_{jk} - W\right)\left(B_{jk} + \phi''_{jk}\right) - A_{jk} + \phi'_{jk}, \tag{65}$$

and for similar reasons the upwards kink x-coordinates are at:

$$\left(y_{jk} + W\right)\left(B_{jk} + \phi''_{jk}\right) - A_{jk} + \phi'_{jk}. \tag{66}$$

These are then rearranged into ascending order, and the y-coordinates can be calculated by:

$$y_1 = \sum_{k\neq i} \mathcal{A}_{jk} \frac{W}{B_{jk} + \phi''_{jk}}, \tag{67}$$

$$y_2 = y_1 + \eta_1(x_2 - x_1), \tag{68}$$

$$_n = y_1 - \left(\sum_{r=1}^{n-1} \eta_r x_r\right) + x_n \left(\sum_{r=1}^{n-1} \eta_r\right), \tag{69}$$

$$y_{2N} = -\sum_{k \neq i} \mathcal{A}_{jk} \frac{W}{B_{jk} + \phi''_{jk}}, \tag{70}$$

where $\eta_n = -\frac{1}{B_{jk}+\phi''_{jk}}$ for downward kinks, or $\eta_n = +\frac{1}{B_{jk}+\phi''_{jk}}$ for upward kinks.

Once a value for μ_{ij} is found, the algorithm follows the same steps as Section 2.4, but with the adjustments of:

$$B_{ij} \leftarrow \frac{\Theta(-\mu_{ij} - \epsilon)}{\sum_{k \neq i} \mathcal{A}_{jk} \frac{1}{B_{jk} + \phi''_{jk}} \Theta(W - |y_{jk} + \varepsilon_{jk}|)}, \tag{71}$$

and,

$$y_{jk} \leftarrow \max\left[-W, \min\left[W, \left(y_{jk} - \frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}\right)\right]\right]. \tag{72}$$

The algorithm can now limit any $|y_{jk}|$ value so that it does not exceed W.

Soft Constraints

An increase in hard constraints within a network may result in the optimisation problem being unsatisfiable. For example, [109] found that in certain networks, the introduction of bandwidth resulted in bottleneck scenarios; this is where all possible routes to a section of the network are limited as all connected edges have maximised their bandwidth limits, but some nodes remain unsatisfied because the two hard constraints are conflicting. This can be mitigated by softening constraints so that the network can still reach a minimum. To accommodate unsatisfied demands, an additional variable ζ_j was introduced, making the inequality (23), in the presence of a correction ε_{jk}, an equality. The variable ζ_j represents any deficit at node j and is determined at each iteration by the state of each node once adjusted current values are injected and sent from the node.

$$\sum_{\neq i} \mathcal{A}_{jk}\left(y_{jk} + \varepsilon_{jk}\right) + \Lambda_j - y_{ij} + \zeta_j = 0. \tag{73}$$

The algorithm then aims to minimise the overall deficit in the network by an additional objective function $\sum_j \zeta_j^2$.

Resistance

The work also included adding a homogeneous resistance value R to the distribution cost ϕ, which remains as a constant on the function throughout the algorithm.

Comparing message passing with quadratic programming

Work done in [110] compares this message passing method with a quadratic programming algorithm for a quadratic cost function; results show that they are both able to minimise the objective function efficiently, and that although quadratic programming requires less iterations for an optimised system for a selection of network sizes, connectivity's, network wealth (the networks average capacity) and converges to a steady state faster, the difference is minimal and one can highlight that message passing is more adaptable and relies less on assumptions, while finding solutions equivalent to leading optimization techniques.

2.4.4 Adapting Continuous Message Passing to the Electrical Grid

The reasons message passing is well suited as a distribution method for power flow are its distributive nature, its probabilistic framework that can accommodate volatile production and demand and its modest increase in computational complexity with the system size; this enables one to find global optimal solutions within a reasonable amount of time, even on large-scale non-convex networks. The method is adaptable and allows for different objectives to be minimised, such as generation costs and power loss. It can consider constraints such as the power flow equations and bandwidths on power lines, and message passing is suited well to sparse networks such as those of power networks. With the increase in renewable generators adding uncertainty into the network, this probabilistic method may be able to inherently consider these uncertainties to provide a more robust, reliable network. However, this message passing method was developed as a generic resource allocation technique and in order to use it within power networks, it is necessary to discover and understand its capabilities in terms of power flow, e.g. power flow equations, adjusting voltages, considering the costs of generation and adding generation limits, as well as making the program appropriate for power grid topologies and considering failures.

A power grid is a combination of power stations, substations and consumers. In our model these can be considered as network nodes and the power lines linking them are considered to be edges. We are assuming the power grid to be sparsely connected, and that each node, j, has its own capacity, taking either positive (generators) or negative (consumers) values of Λ_j. The aim of the optimisation is to make all the capacities positive or zero by passing power through edges to each node (this provides each consumer with sufficient power and prevents generators providing more power than they are able to generate), as well as minimising the quadratic cost of distribution along each edge. We will consider a simple network with N nodes, each connected to c others via edges, and power passed over an edge is our directional variable, y. A general formulation can accommodate heterogeneous connectivity. Appendix B gives a review on which synthetic network model best suits the topology of an electrical grid.

My goal in this thesis is to demonstrate some of the capabilities of message passing for use within electricity grids. The optimisation problem is considered to be static, where each optimisation problem is solved and then applied; it is not a dynamic approach that adjusts power flow in real-time, and therefore the time delay of passing messages is not problematic, nor is the time-delay of turning up generators.

2.4.5 Continuous Message Passing and the Interior Point Method

We will discuss the ability of message passing in comparison to other existing optimisation techniques. As mentioned in Section 0 work has been done on comparing a simple message passing algorithm with quadratic programming [110] on a simple network, which found that message passing was able to converge to equivalent optimal solutions, and the number of necessary iterations for convergence according to system size, connectivity and wealth of network were found to be higher than quadratic programming, but scaled similarly.

We compare the proposed continuous message passing with the interior point method via a professionally coded program called Matpower [38] to find if it can optimise the network to a similar standard, and at what computational cost. For these comparisons, each method considers only transportation costs and aims to satisfy each nodes power demand. Figure 2.12 shows the change in transportation cost as N increases, calculated using message passing (MP - black) and the interior point method (IPM - red) for $c = 5, 7$ and 10 (for $c = 10$ we only tested larger networks with $N \geq 40$ to reduce the likelihood of small loops) on a simple network with no fluctuations or leaf nodes. Comparing the performance of continuous message passing against

the interior point method under simple conditions, we find that both methods return exactly the same costs. Figure 2.13 indicates the average time taken (in seconds) to optimise the networks from Figure 2.12 as N increases. The figure shows that message passing takes longer and increases linearly with the system size, while the computing time interior point method remains small. The approximate linear increase in computational cost with the system size is consistent with the analysis and since we advocate message passing as a method for economic dispatch at the 15-60 minute time-scale it remains a competitive method that converges in seconds. We also expect that the message passing algorithms code could be optimised to reduce the associated computational time. The professionally coded Matpower program has a reduced computation time but is less flexible in the type of objective functions and uncertainties it can accommodate, computational cost is likely to grow for more complex scenarios and initial and termination criteria are sensitive. The main advantage of the message passing method is in its flexibility, its ability to accommodate different objective functions, uncertainties and constraints, and the potential of the method to consider the probabilistic nature of elements such as renewable sources.

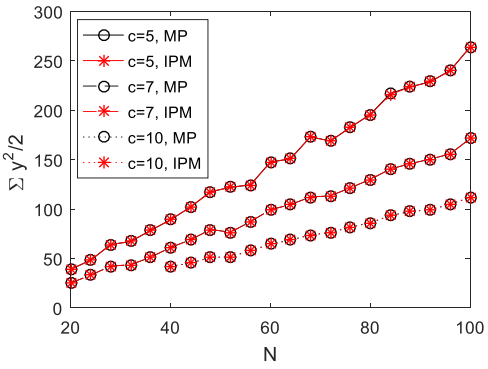


Figure 2.12 Line graph of the cost of distribution for increasing N when using a message passing algorithm (black) and the interior point method (red) algorithm from the Matpower program [38], for different connectivity's on a simple network without fluctuations [111]. One fifth of nodes are generators with $\Lambda = 10$, two fifths are consumers of $\Lambda = -3$ and the rest are substations of $\Lambda = 0$. The two trends follow the same path and appear as one line. © Creative Commons Licence, DOI: 10.1016/j.egypro.2016.12.139.

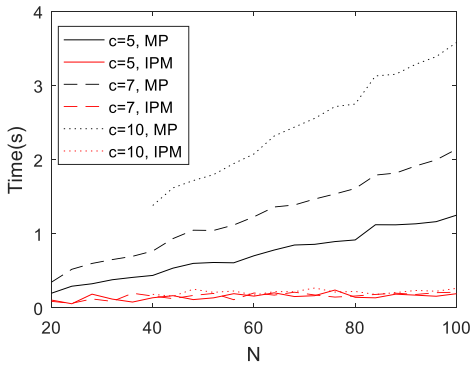


Figure 2.13 The average time taken to run simple networks from Figure 2.12 for increasing N using a message passing algorithm (black) and the interior point method (red) algorithm, for different connectivity's [111].
© Creative Commons Licence, DOI: 10.1016/j.egypro.2016.12.139.

3 Fluctuations

3.1 Introduction

3.1.1 Uncertainty

Electricity grids have always had some element of uncertainty due to volatility in consumer demand but now additional uncertainties from renewable energy sources should be considered. This section looks to understand and accommodate these uncertainties within a power distribution scheme. If the generation or demand of a node is uncertain, predictions are needed to plan for an effective power distribution. However, predictions cannot be deterministic and their full distribution is required. Uncertainty in consumption and demand is a problem since controllable power stations are made to generate exactly the right amount of power, matching the predictions of consumption and power stations generation (point forecasting); if the renewable power stations fall short of these predictions, there may not be sufficient power to satisfy all consumers. Therefore, one way to avoid this is to consider how uncertain the renewable sources are (interval forecasting) and generate extra power at controllable power stations accordingly to compensate for shortcomings; this is called reserve power. Uncertainty about consumer consumption (load) can also be compensated by reserve power, which is able to provide them with extra power if needed. Although this may appear counterintuitive - to generate more power to compensate for uncertain generators creating power - the reserve required to mitigate the effects of uncertainty will be much smaller than the power generated by uncertain generators; any extra reserve required from controllable sources such as fossil fuels will be smaller than what would be required had the renewable generators not been available. Reserve power given at the economic dispatch stage will be lost through heat, or could be stored in batteries at the primary and secondary control stages.

Consumers

Of the 359,905 GW/h of electricity generated in the UK in 2014, around 28,387 GW/h was reused in energy production and a similar amount of 28,651 was lost [112] (Figure 3.1); the remaining 302,867 GW/h were consumed. Figure 3.2 shows the amount used for different purposes, this works out to be 108,324 GW/h domestically, and is expanded to show a pie chart of average usage within the home. Consumers require the ability to use electricity on demand, whenever they see fit and this leads to volatility in demand. There are predictions for the

consumer consumption of energy (load) based on many things; e.g. time of day, week or year, events and weather.

The curve of power consumption over a typical day can be seen in Figure 3.3, and the brown line in Figure 3.4 shows the consumption pattern over a year (this is almost opposite to warmer countries; in winter the UK use extra electricity on heating, whereas in the summer the US use extra electricity on air conditioning).

The error distribution with respect to consumption forecast can be seen in Figure 3.5; according to [113] plotted against a normal distribution of mean $\mu = -0.002$, standard deviation $\sigma = 0.026$. Analysis shows that there is a skewness to the right, $\gamma = 0.715$ and the histogram is narrower, taller and has fatter tails, $\kappa = -4.725$, when compared to the normal distribution. In reference [114] it is suggested that distribution may fit with a Hyperbolic tangent.

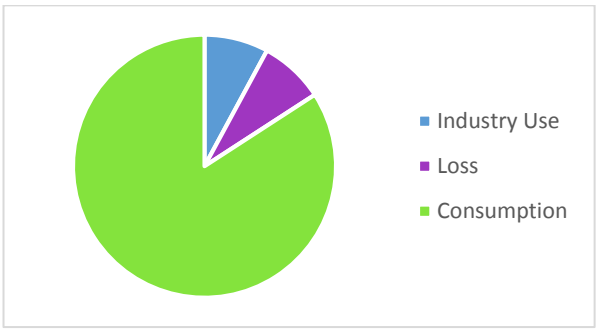


Figure 3.1 Pie chart of the amount of power retained for energy generation, lost and used in consumption [112].

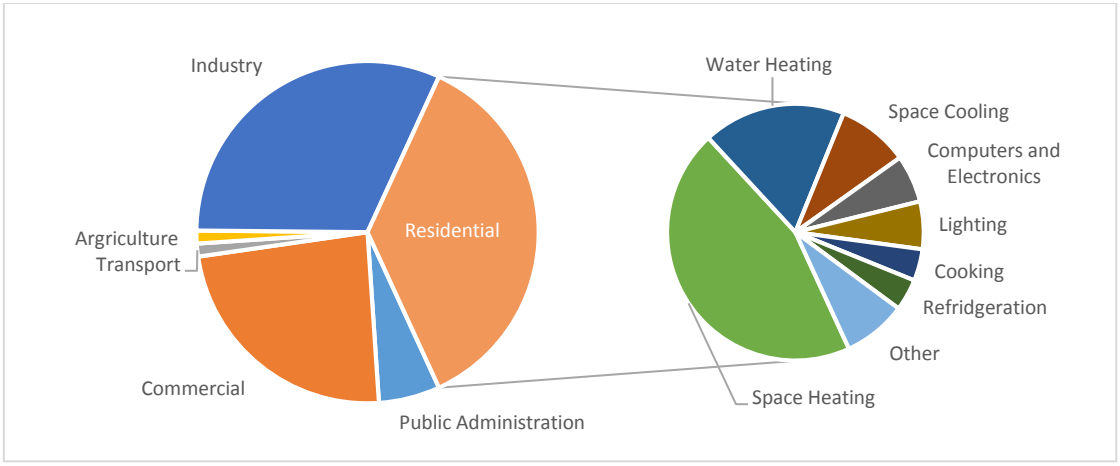


Figure 3.2 Pie chart of what the UK uses in consumption [115].

Figure 3.3 Average curve of power load over one day [116].

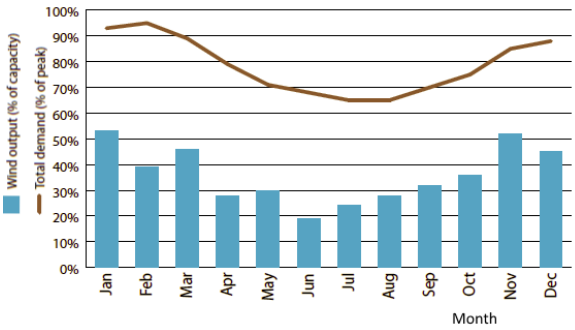


Figure 3.4 Power generation by wind over a year, per month [117]. Wind output as a percentage of capacity (blue bars) and total demand as a percentage of peak (brown line).

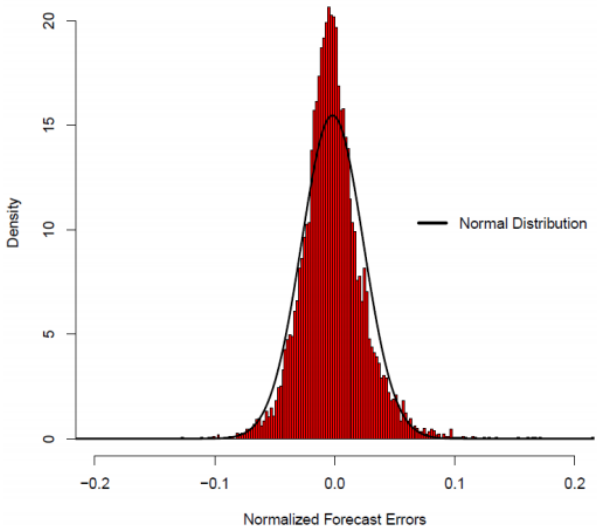


Figure 3.5 Load error distribution compared to a normal distribution (black line) [113]. Reprinted with permission of the National Renewable Energy Laboratory, from http://www.nrel.gov/docs/fy12osti/54384.pdf, accessed on 5th April 2017.

Renewable Energy Generators-Onshore wind turbines

According to [21] there are 5,809 onshore turbines with an operational capacity of 9,571 MW. It is the most commonly used renewable generator source in the UK because of its technological advances and government incentives [118]. Despite the advances in technology, we still must rely on wind predictions to determine our expected generation. The probability distribution of wind is a Weibull distribution according to [119] and Figure 3.9, and Figure 3.8 shows an example of the distribution of wind in Birmingham, UK, which appears to match this model. Although wind does vary in predictive patterns over hours (Figure 3.6), and even months (Figure 3.7 and Figure 3.4), it cannot be predicted with 100% accuracy; for a given time, wind speed is predicted by calculating the raw data from NWP (numerical weather prediction) and considering many factors such as, the placement of the wind turbines within an area, the

surrounding area and its terrain, the amplitude, time of day, dispersion and shadowing effects. Once the wind speed is predicted, this can then be translated into a value for wind power using a power curve, shown in Figure 3.10.

An accurate prediction of the power generated has significant economic benefits as recognised by Hodge and Milligan [120]: "More accurate wind power forecasts can lead to economic efficiency in the unit commitment and dispatch process, as fewer reserves must be kept or deployed to compensate for changes in the wind power output". The problem with volatile energy generators is that one must rely more heavily on reserves to make up for possible lower-than-expected contributions; this increases operational costs which leads to increased final energy prices and environmental impact. According to [40], 10% of annual production is lost by wind power prediction error costs.

The wind speed error distribution according to Lange [121] is Gaussian, [122] also look at the error distribution with respect to forecasted wind speeds and found it to be a truncated normal distribution.

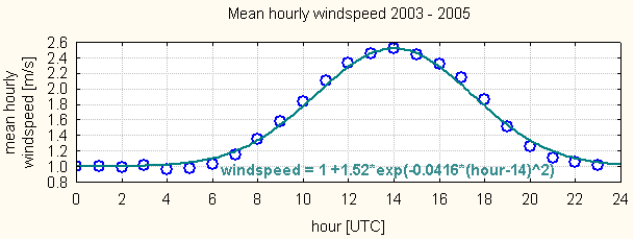


Figure 3.6 Mean wind speed over a 24-hour period [123].

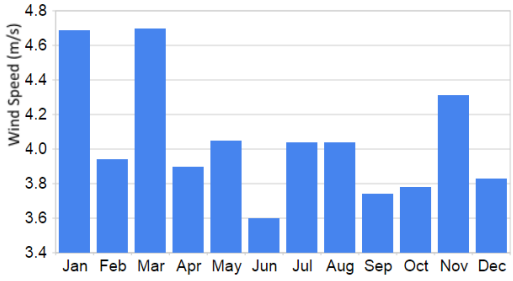


Figure 3.7 Wind speed over a 12-month period in Birmingham averaged from 2000 to 2010 [124].

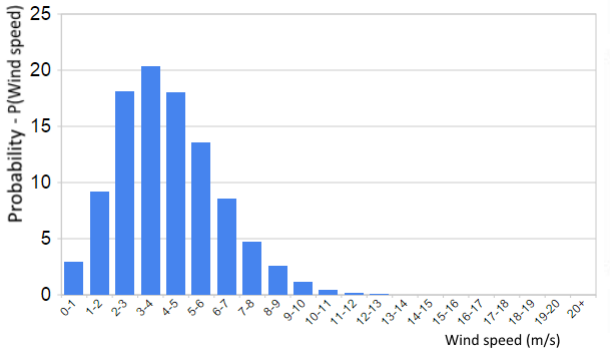


Figure 3.8 Histogram of wind speeds in Birmingham from 2000 to 2010. y-axis is percentage of time, x-axis is meters per second [125].

Figure 3.9 Weibull Wind Distribution [126].

Any error in the estimation of the wind speed will also be transferred to power generation through this power curve. For example, a small mid-speed wind error when relayed through the power curve will amplify the deviation. Alternatively, a small high-speed wind error will translate to a small deviation in wind power output as can be seen in Figure 3.10 and Figure 3.11 which show how the Gaussian error distribution of wind speed may of been skewed by the power curve, and therefore the wind power error distribution may not be Gaussian or Weibull. [127] tells us that a single wind power plant with a forecast of one hour can achieve approximately 15 - 20% mean absolute error relative to the capacity of the plant (load forecast errors are around 1-3%).

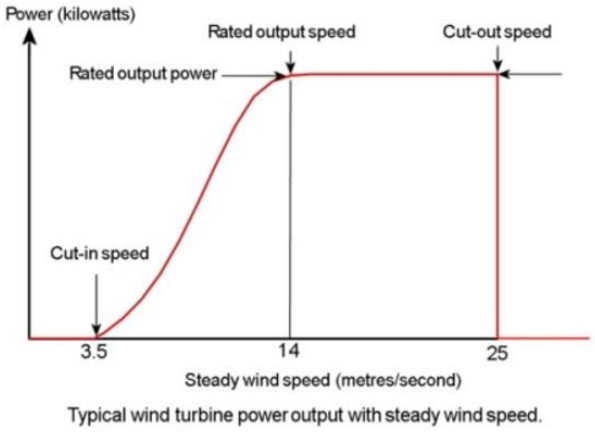


Figure 3.10 Power curve [128]. Representing the power (in kilowatts) output for the wind speed (m/s).

Figure 3.11 How a change in wind speed changes the shape of the power output distribution [121].

Many papers have assumed a Gaussian error distribution [129, 130, 131, 132]. Papers [133, 134] talk of the error distribution with respect to forecast being Beta due to the translation of the wind speed error probability density function through the power curve. They also suggest that a limit of [0,1] is better than the infinite limit of a Gaussian because the power output of a wind turbine is between zero and its maximum capacity (which can be normalised to 1). However, [134] admits that the kurtosis is still too small with a Beta distribution and also highlights that for an accurate confidence interval to be found, the tails need to be the most precise. [120] look at the possibility of the power output distribution being Cauchy, they consider this over many different time intervals and show that it gets closer to the Gaussian the longer the time from forecasting, but for tertiary control which is between 15 minutes and an hour a Cauchy distribution seems to be the most accurate, as the kurtosis seems very large (leptokurtic) at smaller time-scales. According to [120], Cauchy is 16-36% a better fit than Beta, in 95% of the cases in terms of the optimised log likelihood values. Clearly, “If forecast

errors are in principle unavoidable it is at least good to know as precisely as possible when and how they occur" [121].

The scale of the number of wind turbines is also a factor. If there is a large group of wind turbines the error distribution of power generated appears to follow a normal distribution, whereas if it is just the one, this appears to best fit a Cauchy distribution [135].

Renewable Energy Generators-Offshore Wind Turbines

The UK has been the world leader in offshore wind since October 2008, with more installed capacity than any other country [21]. Although less than onshore, the UK has 1,465 turbines and a total operational capacity of 5,097 MW. Industry suggests that by 2020 it will be powering 10% of the UK's electricity annually. The distribution of offshore windfarm production is also Weibull, but with a higher mean, depending on its positioning (Figure 3.12). Most papers talk of wind power error distribution without specifying whether it is on or off-shore, but the wind speed distribution is the same, and the power curve is similar; therefore, we can assume this has the same error distribution as onshore wind power.

Figure 3.12 The Weibull distribution of onshore, and offshore wind power [136].

Renewable Energy Generators-Solar Panels

Solar radiation above the clouds can be confidently predicted, but the atmospheric conditions such as clouds and dust storms make it less predictable at ground level [137]. Average solar power distributed over a day, and a year can be seen in Figure 3.13 and Figure 3.14. Positioning can also be a large factor, for example, Figure 3.15 shows where in the UK gets the highest sunshine duration, and the direction the panels face can also effect the power output. Arthur [138] gives a probability density function of solar irradiance in Ghana, suggesting that the more constant year round probability density function possibilities are exponential, Weibull or Gamma (Figure 3.16).

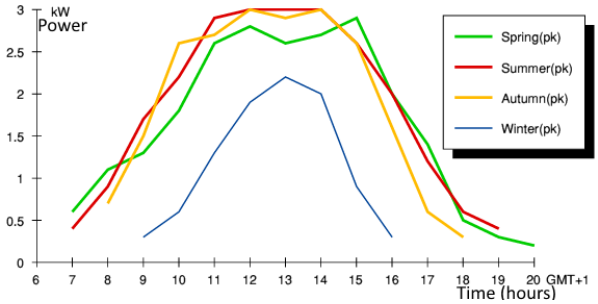


Figure 3.13 Solar Generation over a day for each season [139].

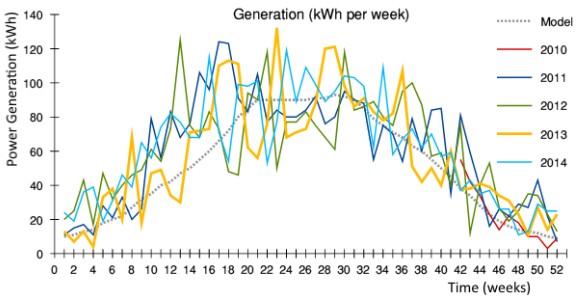


Figure 3.14 Solar Generation over 1 year [139].

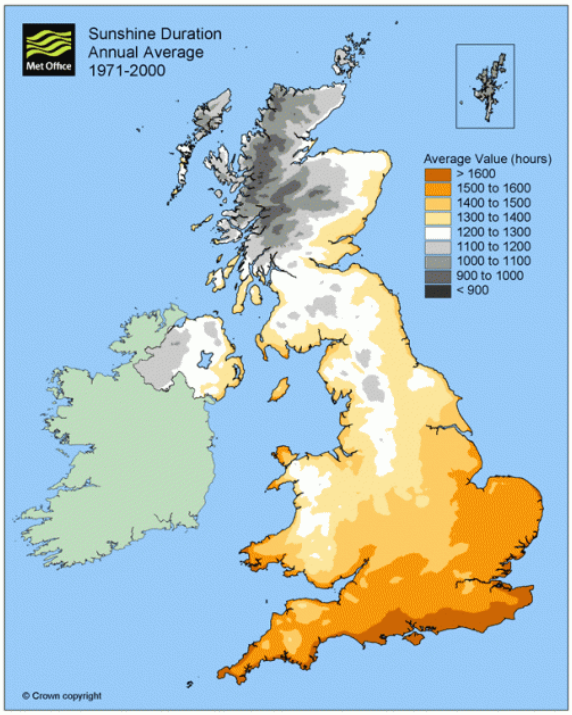


Figure 3.15 Annual average sunshine duration in hours over the UK from 1971-2000 [140]. © Crown copyright, Met Office, [2017].

As solar power generation depends on the clearness of the sky, if the skies are clear the probability of error is small; the error depends on the Clearness index, Figure 3.17. The error distribution of solar power has not been well studied, but [141] suggests it fits a Hyperbolic distribution well.

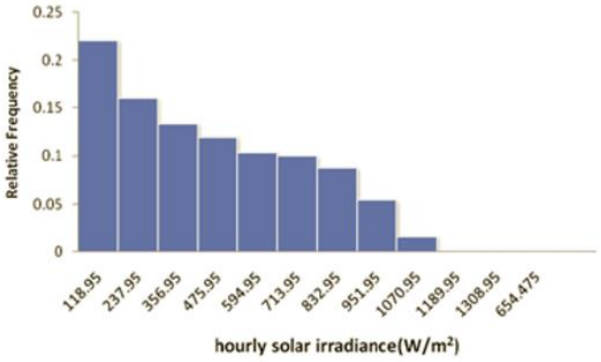


Figure 3.16 Histogram of the probability density function of solar irradiance in Ghana in April [138].

Aston University

Illustration removed for copyright restrictions

Figure 3.17 Clearness index [137]. © Academic Press 2014.

In the future, predictions of wind power, consumers and solar power will become more accurate and uncertainty will become less problematic in the electricity industry. Until then, we will be considering how to maximise the reliability of a power grid that contains volatile generators. This thesis considers uncertainties as Gaussian to best include consumers, wind turbines and solar power distributions, but with the expectation that the message passing method could also consider other distributions.

3.1.2 Fluctuations

As well as uncertainty about the amount of power generated for the next time-interval, *during* each time interval the amount of power produced will also be changing with some uncertainty. We call these fluctuations; the uncontrollable variations around the predicted estimates of generation over each time interval. This is a problem for power grids because every millisecond, power distribution is corrected so that consumers will receive the correct amount of electricity. For a moment during the 15-60 minute interval a renewable source may produce less than its average and a consumer may use more than the average value. Primary and secondary control accounts for this at second to second timescales, but it needs to be considered first within tertiary control so the reserve power will be available. [142] suggests that short term wind fluctuations follow the Weibull distribution, consistent with the long-term distribution of wind. Fluctuations can be dealt with in the same manner as uncertainty, by assigning a distribution of

the expected level of fluctuation; this will provide extra reserve to nodes according to their level of fluctuation to accommodate for smaller generation (or higher demand) during the time frame. Fluctuations in generation can be considered simultaneously with uncertainty in forecasting, assuming they follow the same probability density function. This chapter will assume Gaussian distributions for fluctuations so that they can be encompassed within the uncertainty distribution. An additional problem is that fluctuation within networks results in fluctuations over power lines. If power line limits have already been reached, excess power flowing through lines because of fluctuations may overheat them and could result in tripping.

3.1.3 Problem

An electricity grids biggest priority is the reliability of the network. With fluctuations and uncertainty from renewable generators and consumers increasing, this will become increasingly difficult to ensure. This chapter aims to suggest techniques that will increase the reliability of a network by providing additional reserve to volatile nodes. However, one problem with producing excess power to make uncontrollable sources reliable is that there cannot be excess power in a power grid. This is resolved by losing the power through power loss (heat) and reactive power. These are controlled through changes in frequency, voltage magnitudes and phases. Power loss is inefficient, and economically and environmentally wasteful so a good method of power distribution will look to minimise the amount of power lost while maintaining reliability.

3.1.4 Other Methods

Before introducing the message passing methods we advocate as being highly suitable for the task at hand, we will review existing and suggested methods to address the power distribution in the fluctuative grid. Section 1.6 addresses the existing optimal power flow methods without specifically considering fluctuations; the following methods are adapted from these, or alternative methods which look to include the problem of fluctuations. Efforts already being made to consider the increased fluctuations and volatility of power grids with renewable sources, and how to make this more reliable include:

F1. Chance Constrained Optimal Power Flow (CCOPF) [143] - suggests that the problem of increased volatility in the power grid can be addressed by changing the constraints of the current OPF from hard to “satisfiable within a certain probability”. It requires each constraint to be satisfied within a certain probability according to work done by [144].

Some examples of this can be seen in [145, 143] which specifically concentrate on the fluctuations of wind farms in an electricity grid. This method is relatively new and simulations are not yet entirely compatible with an AC real grid; for example, it is not yet able to consider active power loss. These methods only consider uncertainty heuristically and superficially without inherently incorporating the probabilistic nature within the algorithmic framework. Reference [145] assumes a Gaussian distribution for the fluctuations but also considers how accurate this would be in the case of other distributions and finds that it was least accurate if the real distributions were asymmetric. Reference [143] concentrates on line bandwidths and states that they should be satisfiable within a certain probability; allowing the power flow to exceed the bandwidth limits, but only as long as the breach is not for too long (otherwise overheating and tripping may occur). This raises concerns as to whether repeated overheating levels to maintain the balance in the short term may result in longer term issues. CCOPF uses a sequential linear cutting place algorithm. The method only makes minimal changes to current OPF methods; preliminary results highlight the advantages of real-life implementation costs and usability. Results also show CCOPF to be only slightly slower the standard methods of economic dispatch. And testing showed that it was able to provide cheaper alternatives to standard OPF in cases of high penetration of renewables and works better than adding buffer values to the bandwidths of the line. However, they suggest that power grids cannot handle a penetration of renewable sources higher than 30% without an adjustment to the network, and considers the uncertainty only in the constraints used rather than the main method. It also uses the assumptions that wind farms are far enough away from each other that they can be deemed independent and so correlations are ignored. However, the increase of correlations between micro-grids in networks will increase and will need to be addressed. It is a convex deterministic approach which requires centralised optimisation and cannot be done locally. A related paper [145] develops on this work by removing the small phase angle assumption, which had made it a DCOPF instead of ACOPF, and deals with minimising the probability of the grid losing synchrony. They make the calculations convex by using arcsin and find that additional complexities arise.

F2. Quantum Genetic algorithm (QGA) [146] developed upon the existing OPF method from numbered point A6 has been proposed for dynamic economic dispatch where the quantum probability encoding mechanism and the genetic algorithm increases the chance of finding the global optimum, they also suggest this can be improved by including a chaotic algorithm for the initial condition (CQGA). The method initialises

the population, encodes the power outputs into binary ‘chromosomes’ of information, individually measures the fitness of each object, uses a quantum rotation game to update each object and finally applies adjustments to ensure the solution is not at some local minima. Evolving ‘chromosomes’ allow for multiple states to be passed between generations and this large amount of information allows each update to be well informed. The algorithm is fast and treats wind power generation as a stochastic variable; they suggest that it is able to reduce costs. However, this is a heuristic algorithm and may be deficient in precision, also the binary coding may be oversimplifying the problem.

F3. Distributed and Parallel Optimal Power Flow (DPOPF) [147] suggests using a mixture of recursive quadratic programming and the Lagrange projected gradient method to consider renewable energy at the transmission stage in real time. It sets the output of the renewable sources as fixed power inputs and includes fuel cost in their objective function in an attempt to reduce carbon emissions. The paper pictures the network as a supercomputer due to the processing units and high speed communications between buses, and uses directed bipartite graphs to control the computational synchronization. The method uses a quadratic approximation of the objective function and linearises the constraints, it splits the question into sub-problems and solves them in parallel in two steps, first solving the unconstrained minimisation problem, then dealing with the bounds. They use Armijo’s rule [148] (which aids in finding a suitable step size with an inequality equation based on the convergence theorem) to determine the step size to ensure a convergence and they pass information on the most updated generators in order to reduce computational speed. The method considered is very fast, but the quadratic and linear approximations may result in a suboptimal solution.

F4. Other suggestions include using the Strength Pareto Evolutionary algorithm (SPEA) [149], which creates a hidden structure from the inputs; power demand, wind speed and solar radiance, and outputs of the power generated from wind and solar, power still needed, emissions and costs of generators. These are then used in the economic dispatch calculations. Ref. [150] describes wind as a discretised beta distribution to consider extreme conditions and [151] uses particle swarm optimisation from the numbered OPF method A7 and fuzzy adaptive technique, considers wind using a probability distribution and creates scenarios using a roulette wheel technique. These methods are heuristic and unprincipled, suggesting a compromise in accuracy and completeness.

3.1.5 Aim

The aim of the derivation that follows is to devise an effective, distributive and computationally efficient algorithm to address the power dispatch problem in the presence of fluctuations using a message passing approach. The derivation follows similar steps as Wong and Saad, but now considering that the capacity at each node is probabilistic and drawn from a Gaussian distribution (with a mean, $\bar{\Lambda}$, representing the average expected capacity within a certain time frame, and a standard deviation, σ, representing the estimated variation from the mean, representing the corresponding fluctuations and uncertainty at that particular node over the time frame). The reason this may be more suitable to power grids than existing methods is because the distribution can be considered inherently within the derivation instead of trying to get around them by introducing additional constraints.

The method is based on deriving and approximating the free energy in the presence of uncertainties. To rewrite the free energy in the presence of uncertainties, it is possible to calculate the *average* free energy. The exhaustive method for deriving the average free energy looks at every scenario of possible combinations of actual values of capacity according to their probability distribution, and then takes the average of this with respect to the probability of each combination, this is called quenched averaging, and gives an exact result. However, calculating this can be complicated and it has been found that alternatively, deriving an annealed approximation, where the disorder manifested by the fluctuating generation and demand is treated at the same level as thermal averaging instead of after, results in a good approximate solution without the complex derivation.

3.2 Annealed Approximation

3.2.1 Method

The reason for introducing the annealed approximation is the non-trivial calculations of the quenched average over the disorder, $-T\langle \ln Z \rangle$, are much more complex, where angled brackets represent averaging with respect to the fluctuating production and demand. Moreover, work done on some systems [152] suggests that the annealed approximation gives a good upper bound solution. To find the quenched average of the VFE requires the calculation; $-T\langle \ln Z \rangle$, but in the annealed approximation, is can be replaced by a simpler expression $-T \ln\langle Z \rangle$ at the cost of accuracy.

To implement an annealed approximation, the same zero-temperature VFE as in Equation (49) is used, and A_{ij} and B_{ij} still represent the first and second derivatives of F_{ij}, respectively. Adjustments are now made to the constraint to accommodate the now fluctuating Λ; the new constraint bounds the probability of the total resource at a node being smaller than zero to less than a predetermined value, ρ_j. The probability of an unsatisfied node becomes:

$$\Pr\left(\Theta\left[\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}+\varepsilon_{jk}\right)-y_{ij}+\Lambda_j\right]=1\right)>1-\rho_j \quad \text{where} \quad \Lambda_j=\left(\bar{\Lambda}_j,\sigma_j^2\right), \tag{74}$$

with the possible outcomes of 1 and 0 if the node j is satisfied or unsatisfied respectively.

If we want a node to be satisfied $1-\rho_j$ of the time, assuming an independent distribution of $P(\Lambda_j)$ for each Λ_j, this can be written as:

$$\begin{aligned}\left\langle\Theta\left[\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}+\varepsilon_{jk}\right)-y_{ij}+\Lambda_j\right]\right\rangle_{\Lambda_j}&\\ =\int_{-\infty}^{\infty}\Theta\left[\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}+\varepsilon_{jk}\right)-y_{ij}+\Lambda_j\right]P\left(\Lambda_j\right)\mathrm{d}\Lambda_j&\\ >1-\rho_j,&\end{aligned} \tag{75}$$

where ρ_j is a small predetermined value to determine the case where the node may not be satisfiable (using an assumption that excess power is acceptable and we are only concentrating on meeting demand; as [129] explains there are downsides to excess production such as driving down energy prices, but operators' larger concerns are about meeting consumer demands). The message passing algorithm needs to provide extra reserve to reduce the probability of unsatisfiable needs while still minimising power flow costs.

Given that we are assuming Λ_j to be sampled from a Gaussian distribution, $P\left(\Lambda_j\right)=\frac{e^{-\frac{1}{2\sigma_j^2}\left(\Lambda_j-\bar{\Lambda}_j\right)^2}}{\sqrt{2\pi\sigma_j^2}}$, we can rewrite $P(\Lambda_j)$ in terms of $\bar{\Lambda}_j$ and σ_j^2.

$$\int_{-\infty}^{\infty}\Theta\left[\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}+\varepsilon_{jk}\right)-y_{ij}+\Lambda_j\right]\frac{e^{-\frac{1}{2\sigma_j^2}\left(\Lambda_j-\bar{\Lambda}_j\right)^2}}{\sqrt{2\pi\sigma_j^2}}\mathrm{d}\Lambda_j>1-\rho_j. \tag{76}$$

This can be simplified by defining $\tilde{\Lambda}_j = \frac{\Lambda_j - \bar{\Lambda}_j}{\sqrt{2\sigma_j^2}}$ and rearranging the terms, giving the integral:

$$\int_{-\infty}^{\infty} \Theta\left[\sum_{k\neq i} \mathcal{A}_{jk}\left(y_{jk} + \varepsilon_{jk}\right) - y_{ij} + \tilde{\Lambda}_j\sqrt{2\sigma_j^2} + \bar{\Lambda}_j\right] \frac{e^{-\tilde{\Lambda}_j^2}}{\sqrt{\pi}} \mathrm{d}\tilde{\Lambda}_j > 1 - \rho_j. \quad (77)$$

Using the definition of the complementary error function, $2\int_{-\infty}^{\infty} \Theta(t-a)\frac{e^{-t^2}}{\sqrt{\pi}}dt = 1 - \mathrm{erf}(a) = \mathrm{erfc}(a)$, and using $\mathrm{erfc}^{-1}(1-t) = \mathrm{erf}^{-1}(t)$, the constraint can be rearranged to get:

$$\sqrt{2\sigma_j^2}\mathrm{erf}^{-1}\left(2\rho_j - 1\right) + \sum_{k\neq i} \mathcal{A}_{jk}\left(y_{jk} + \varepsilon_{jk}\right) - y_{ij} + \bar{\Lambda}_j > 0. \quad (78)$$

Now the calculations are derived as before with this new constraint. The optimal ε_{jk} remains unchanged:

$$\varepsilon_{jk} = -\frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}, \quad (79)$$

but the Lagrange multiplier becomes:

$$\mu_{ij} = \min\left[\frac{\sum_{k\neq i} \mathcal{A}_{jk}\left(y_{jk} - \frac{A_{jk} + \phi'_{jk}}{B_{jk} + \phi''_{jk}}\right) + \sqrt{2\sigma_j^2}\mathrm{erf}^{-1}\left(2\rho_j - 1\right) - y_{ij} + \bar{\Lambda}_j}{\sum_{k\neq i} \mathcal{A}_{jk}\left(B_{jk} + \phi''_{jk}\right)^{-1}}, 0\right]. \quad (80)$$

Differentiating F_{ij}^* once to obtain A_{ij} and a second time to obtain B_{ij} gives:

$$A_{ij} \longleftarrow -\mu_{ij}, \quad (81)$$

$$B_{ij} \longleftarrow \frac{\Theta\left(-\mu_{ij} - \epsilon\right)}{\sum_{k\neq i} \mathcal{A}_{jk}\left(B_{jk} + \phi''_{jk}\right)^{-1}}, \quad (82)$$

and the backwards message $y_{jk} \longleftarrow y_{jk} + \varepsilon_{jk}$ remains unchanged.

This allows us to solve a scenario, where the demand and supply fluctuate and where the message passing method will satisfy each node with probability $1-\rho_j$. It gives a node with fluctuating demand an appropriate amount of reserve power. This approximation is good because it allows an accuracy level to be determined, and may be good for calculating the level of reserve controllable power stations should provide.

The annealed approximation has a disadvantage if there is not enough power in the network to satisfy all nodes plus the reserve need; then the program may not converge as it considers the situation unsatisfiable, rather than satisfiable but with a lower confidence level. The outcome of the derivation per node is equivalent to changing the value of Λ, to a new value, $\Lambda + \sqrt{2\sigma^2}\mathrm{erf}^{-1}(2\rho-1)$.

The required reserve value is explained through the Gaussian distribution plot (Figure 3.18). We denote:

$$z = \Lambda + \sigma\sqrt{2}\mathrm{erf}^{-1}(2\rho-1). \tag{83}$$

This gives the value on the x-axis that satisfies the distribution $1-\rho$ of the time, and demonstrates how the annealed inspired method does what was intended. In what follows we will calculate the quenched average which does not require this type of approximation and therefore will be more accurate.

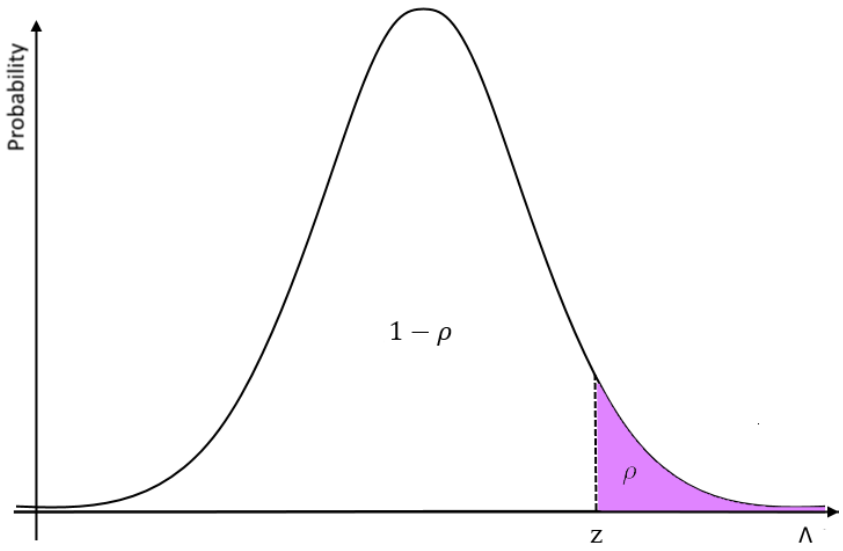


Figure 3.18 The z-value of a one-tailed Gaussian distribution.

3.2.2 Characteristics

We studied the performance of the suggested algorithm on a benchmark network. Figure 3.19 shows an example of a small 9 node IEEE network[6] [153] with a variation of average capacities of $\bar{\Lambda} = 5, -1$ and -2 and standard deviations of $\sigma = 0, 0.5$ and 1. An unsatisfaction rate of $\rho_j = 0.05$ for all nodes is chosen arbitrarily, ensuring a 95% confidence level for the network. The middle diagram shows how the power is optimally deployed over the edges to minimise the distribution cost, $\phi = \frac{y^2}{2}$, while providing each node with sufficient power. The figure on the right shows r, the remaining power, or reserve at each node, in view of each node's level of uncertainty. Comparing the reserve power at each node with their respective standard deviation, we see that consumers with a standard deviation of 0.5 return a reserve of 0.822 MW[7], independent of $\bar{\Lambda}$; this reserve power is for in the case where the consumer requires more power than expected. Similarly, generator 2 (numbered nodes start with node 1 at the top, and increase clockwise) with $\bar{\Lambda} = 5$ and $\sigma = 0.5$ retains a reserve of 0.822 MW to secure the node in case the generator does not produce what was expected (mean value). In the deterministic case the generator at node 2 would have had a reserve of less than 0.822 MW. We see that $\sigma = 1$ returns a remaining power twice that of $\sigma = 0.5$, and any consumers with $\sigma = 0$ have zero reserve, as would happen in an optimisation for all consumers in the deterministic case.

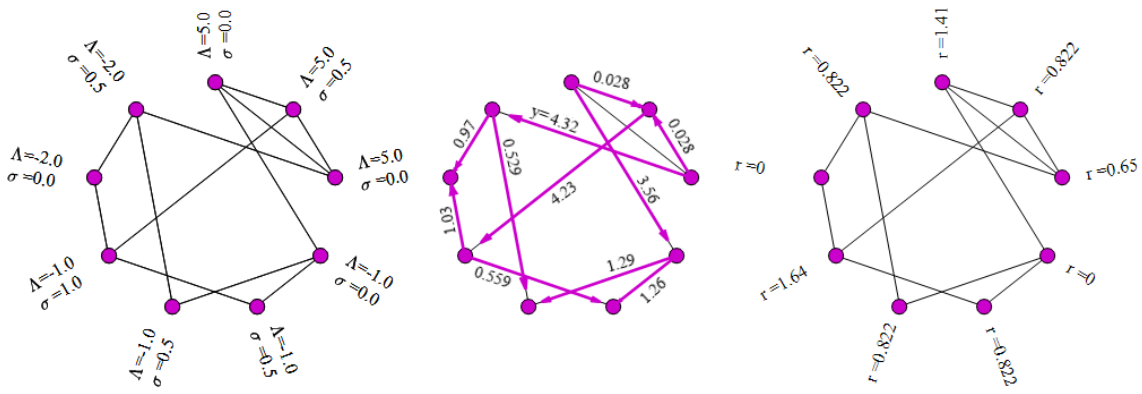


Figure 3.19 Example of power distribution under the annealed approximation and 95% satisfaction requirement on IEEE 9-Bus network [153].

Figure 3.20-Figure 3.26 are from a simple example of a 24-IEEE [154] network with 14 generators and a total expected generation capacity of 20 MW (each with an individual average capacity value of 1 or 2), and 10 consumers with a total expected demand of 10 MW (each

[6] To avoid the complication of leaf nodes (Appendix C), extra edges are added where appropriate to all networks.

[7] We will discuss capacities in terms of MW for easier reading and understanding.

with an average capacity of -1), where $\rho_j = 0.05$ for all nodes. In black we see the level of reserve power given to volatile nodes as their standard deviation increases, and at the red dashed line is where the remaining reserve r reaches the total available power and the program is then unable to converge. In each figure, there is an orange dotted line which represents the necessary resource, z from Equation (83), required for a 95% confidence level with respect to σ. Each blue line denotes the actual unsatisfaction level [155], ρ^*, achieved according to the plotted reserve power; this is calculated by finding the unsatisfaction level of each node (using the nodes reserve power according to its own distribution) and taking the maximum value. We will first consider cases where only one node, consumer or generator is volatile while the others are deterministic. In Figure 3.20 we see that reserve given to a consumer increases linearly as the uncertainty increases, exactly along the revised capacity z. The unsatisfaction level, defining the fraction of cases for which demand would be unsatisfied starts at zero when $\sigma = 0$ and the demand capacity is deterministic; once σ is bigger than 0, the unsatisfaction level is maintained at 0.05 throughout. The program is unable to converge once the reserve required exceeds 10 MW, this is because there is only 10 MW of excess power in the network and higher power allocation would require some load shedding (see Section 4). Figure 3.21 shows a similar example where only one generator node's uncertainty increases. Here we see a difference in that the generator retains unused power in order to minimise distribution cost at low standard deviation; once the power needed to maintain a 95% confidence level (seen by the z-line) exceeds the existing power, the reserve starts to increase linearly with σ. The program stops converging as reserve power reaches 8.1 MW, because the generator was originally providing 1.9 MW, and that power has already been deployed.

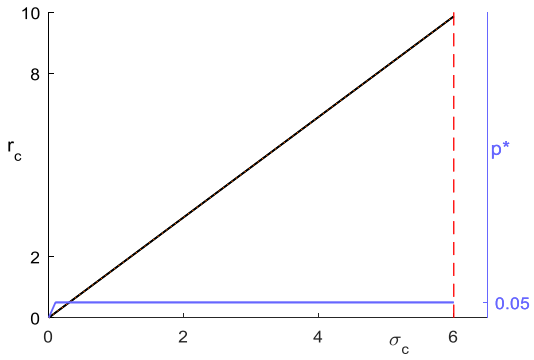


Figure 3.20 The reserve power given to one consumer as its standard deviation is increased in a 24-node network [154] (black). Blue represents the confidence level the reserve power gives. Red shows the standard deviation level for which the program is unable to converge. Subscript c indicates the axis associated with the single consumer.

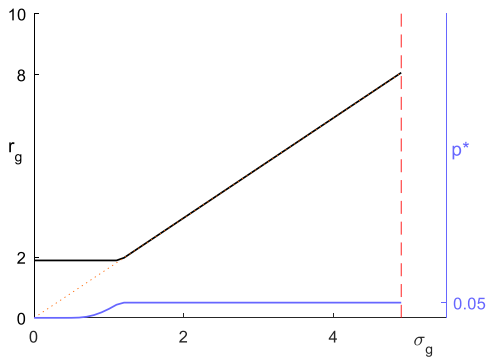


Figure 3.21 The reserve power remaining at one generator as its standard deviation increases in a 24-node network (black). Dotted orange shows the "z-line" from Equation (83). Subscript g indicates the axis associated with the single generator.

In Figure 3.22 all consumers in the network are given an increasing volatility resulting in similar results where r of each node increases at the same rate as when just one volatile node is present. The reason the program does not converge for higher volatility values is because there are now ten consumers requiring excess instead of one, this means the available excess is fully used up at a lower volatility level. The unsatisfaction level here is measured by the consumer with maximum percentile chance of being unsatisfied. Figure 3.23 shows the same behaviour in the case of multiple generators; when all generators have increasing uncertainty, the algorithm stops converging at lower volatility levels as there are 14 generators sharing the available excess which will run out at lower σ levels. It can be seen, that some generators start at $\sigma = 0$ with no available power, this means they would be fully utilised in the deterministic case and the z-line is followed from lower σ levels. Here we observe that the program is unable to converge once all generator reserves reach a focal point on the z-line, this is because there is no longer any spare power available in the network.

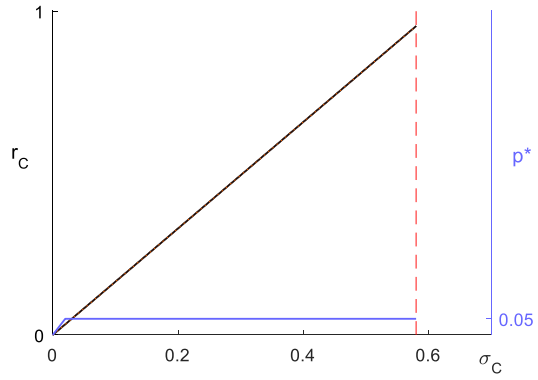


Figure 3.22 The reserve power at multiple consumers as all their standard deviations increase. Subscript C indicates the axis associated with all consumers.

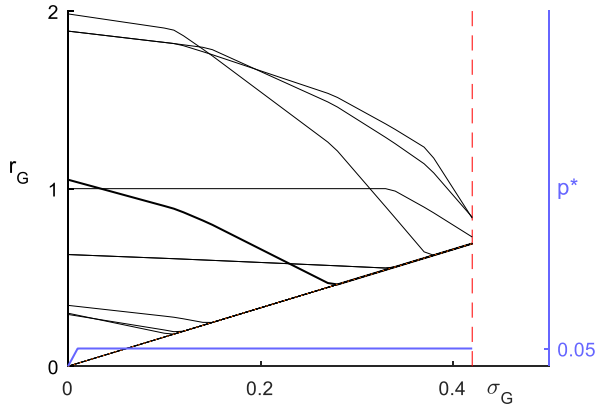


Figure 3.23 The reserve power at all generators as their standard deviations increase. Subscript G indicates the axis associated with all generators.

Figure 3.24 presents the case of deterministic fully utilised generators to demonstrate how ρ^* is lower than ρ for lower standard deviation until the generators start to require extra reserve. These figures also show how the reserve of power at a generator can go up or down; down in order to satisfy the generators who require extra reserve until one by one, their remaining power meets the z-line, where it then goes up as extra reserve is now needed. Figure 3.25 shows an example of the reserve of all nodes in the network as all 24 nodes standard deviations are increased. We see again how all excess power reaches a point, and at that point the network is no longer satisfiable, this happens when all nodes have a reserve of 0.417 MW, which is equivalent to the power available in the network divided by the number of nodes, as expected. The consumers were represented in pink but are only found in the increasing z-line, and may not be obvious.

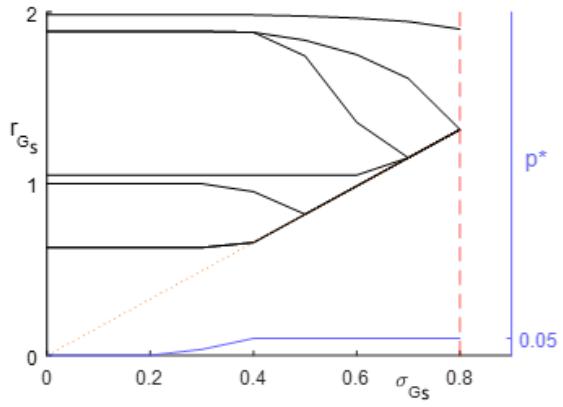


Figure 3.24 The reserve power at all generators that were not fully utilised at $\sigma = 0$, as their standard deviation increases. Subscript G_S indicates the axis associated with these generators.

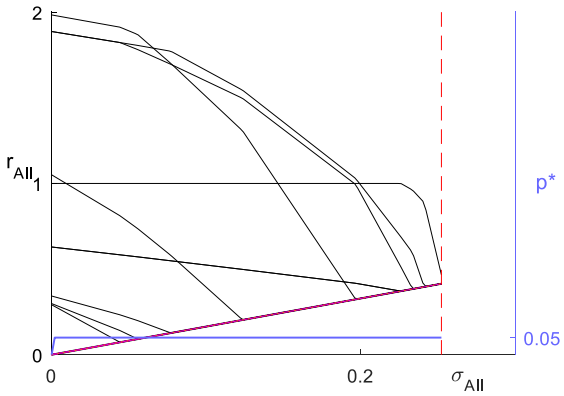


Figure 3.25 The reserve at all 24 nodes in a 24-node IEEE synthetic network with an increasing standard deviation. Pink represents the reserve at consumer nodes, black represents the reserve at generators (all consumers and fully utilised generators at $\sigma = 0$ follow the z-line for all values of σ).

Finally, Figure 3.26 shows an example of some nodes with fixed sigma at 0, 0.2, 0.25, 0.5 and 1, and some nodes have an increasing sigma. It can be seen, that the reserve at nodes with fixed standard deviations remains constant whether consumer or generator, reserve at generators with no uncertainty decreases to zero (or until they reach their minimum reserve according to their value of sigma; like the generator assigned $\sigma = 0.5$) and nodes with increasing standard deviation strictly follow the z-line once it intersects the reserve power the node has. All figures show how every node strictly maintains a 95% confidence level by following the z-line once their extra reserve is depleted.

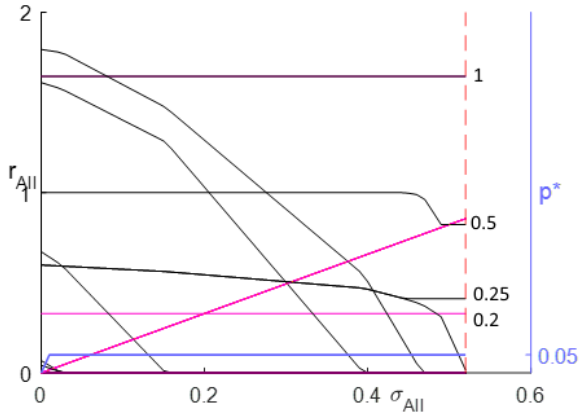


Figure 3.26 The reserve power per node in a network where some nodes have increasing standard deviations, and some have fixed standard deviations of 0, 0.2, 0.25, 0.5 and 1 (labelled). Pink represents the reserve at consumer nodes, black represents generators. Nodes with fixed standard deviations maintain a corresponding reserve value and generators with no uncertainty use their power until they have no reserve. The network is unable to find a solution once all available reserve (reserve not specifically used for protecting an uncertain node) has been provided.

3.3 Quenched Averaging

3.3.1 Method

To more accurately take the average expectation without approximations, quenched averaging is considered, as we want the fluctuations to be incorporated accurately into the distribution of the network's properties. Quenched averaging works by assuming a general temperature, and averaging over all possible combinations of capacity values, per their individual probability distributions, and the probability of each combination; only once the average is taken, we then fix the temperature to zero. It is more accurate than taking the average possible combinations before considering the disorder, but calculations are more complicated. Quenched averaging of the free energy is calculated by $\langle F \rangle = -T\langle \ln Z \rangle$. The expected VFE becomes:

$$\left\langle F_V(y_{ij}|\mathbf{T}_j)\right\rangle_{\Lambda_j} = \int_{\{y_{jk}|\sum_{k\neq i} \min y_{jk}-y_{ij}+\Lambda_j\}} \left[\sum_{k\neq i}\left[F_V\left(y_{jk}|\mathbf{T}_k\right) + \phi(y_{jk})\right]\right] P\left(\Lambda_j\right)\mathrm{d}\Lambda_j. \tag{84}$$

The expected VFE should be optimised with respect to all variables under the constraints. These calculations will find the optimal F_{ij}^* and then take the quenched average. Starting with F_{ij}^* in Equation (85), notice that the non-averaged F_{ij} is expanded around the derivatives of the expected VFE, denoted $A^{<>}$ and $B^{<>}$. The rationale for this is that we want to optimise the expected VFE given the current values, and if we calculate iteratively the expected VFE derivatives, these are the available messages received from the rest of the graph.

$$F_{ij}^* = \frac{1}{2}\sum_{k\neq i} \mathcal{A}_{jk} \frac{\mu_{ij}^2 - \left(A_{jk}^{<>} + \phi'_{jk}\right)^2}{B_{jk}^{<>} + \phi''_{jk}}. \tag{85}$$

As with Equation (55), μ_{ij} is:

$$\mu_{ij} = \min\left[\frac{\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk} - \frac{A_{jk}^{<>}+\phi'_{jk}}{B_{jk}^{<>}+\phi''_{jk}}\right) - y_{ij} + \Lambda_j}{\sum_{k\neq i}\mathcal{A}_{jk}\frac{1}{B_{jk}^{<>}+\phi''_{jk}}}, 0\right], \tag{86}$$

and for brevity we define $x \equiv y_{ij} - \sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk} - \frac{A_{jk}^{<>}+\phi'_{jk}}{B_{jk}^{<>}+\phi''_{jk}}\right)$.

To find $A_{ij}^{<>}$ and $B_{ij}^{<>}$ the average expected VFE is needed.

$$\langle F_{ij}^*\rangle_{\Lambda_j} = \frac{1}{2}\sum_{k\neq i}\mathcal{A}_{jk}\frac{\langle\mu_{ij}^2\rangle_{\Lambda_j} - (A_{jk}^{<>}+\phi'_{jk})^2}{B_{jk}^{<>}+\phi''_{jk}}. \tag{87}$$

$A_{jk}^{<>}$, $B_{jk}^{<>}$, ϕ'_{jk} and ϕ''_{jk} are not functions of Λ_j and so unaffected by the averaging operation. In order to differentiate $\frac{\partial\langle F_{ij}^*\rangle_{\Lambda_j}}{\partial y_{ij}}$, $\langle\mu_{ij}^2\rangle_{\Lambda_j}$ needs to be calculated, which will be different from $\langle\mu_{ij}\rangle_{\Lambda_j}^2$ in general. As x is the point of change in $\min[0,*]$, this allows us to limit the values of Λ_j between $[-\infty, x]$, and for $\Lambda_j = [x,\infty]$, $\mu_{ij} = 0$. This can then be rewritten as:

$$\langle \mu_{ij}^2 \rangle_{\Lambda_j} = \frac{1}{\sum_{k\neq i} \mathcal{A}_{jk} \left(\frac{1}{B_{jk}^{<>} + \phi_{jk}''} \right)^2} \left[\tilde{\Lambda}_2 - 2x\tilde{\Lambda}_1 + x^2 \tilde{\Lambda}_0 \right]. \tag{88}$$

Considering the integral limits, and defining:

$$\tilde{\Lambda}_0 = \int_{-\infty}^{x} P(\Lambda_j) \mathrm{d}\Lambda_j = \frac{1}{2} \left[\mathrm{erf}\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j}} \right) + 1 \right], \tag{89}$$

$$\tilde{\Lambda}_1 = \int_{-\infty}^{x} \Lambda_j \, P(\Lambda_j) \mathrm{d}\Lambda_j = \frac{\bar{\Lambda}_j}{2} \left[\mathrm{erf}\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j}} \right) + 1 \right] - \frac{\sqrt{2\sigma_j}}{2\sigma_j \pi} e^{-\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j}} \right)^2}, \tag{90}$$

$$\begin{aligned} \tilde{\Lambda}_2 = \int_{-\infty}^{x} \Lambda_j^2 \, P(\Lambda_j) \mathrm{d}\Lambda_j \\ = \left(\frac{\bar{\Lambda}_j^2 + \sigma_j^2}{2} \right) \left[\mathrm{erf}\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j}} \right) + 1 \right] \\ - \frac{e^{-\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j}} \right)^2}}{\sqrt{\pi}} \left(\sigma_j^2 \left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j^2}} \right) + \sqrt{2} \bar{\Lambda}_j \sigma_j \right), \end{aligned} \tag{91}$$

when $P(\Lambda_j) = \mathcal{N}(\bar{\Lambda}_j, \sigma_j)$.

The differentiation of F_{ij}^* with respect to y_{ij} gives:

$$A_{ij}^{<>} \leftarrow \frac{1}{2} \frac{\left(\mathrm{erf}\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j^2}} \right) + 1 \right) (x - \bar{\Lambda}_j) + \frac{2\sigma_j^2 e^{-\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2\sigma_j^2}} \right)^2}}{\sqrt{2\pi\sigma_j^2}}}{\sum_{k\neq i} \mathcal{A}_{jk} \left(\frac{1}{B_{jk}^{<>} + \phi_{jk}''} \right)}, \tag{92}$$

which in this case is equivalent to $-\langle \mu_{ij} \rangle_{\Lambda_j}$, and the second differentiation gives:

$$\overset{<>}{ij} \leftarrow \frac{1}{2}\frac{\operatorname{erf}\left(\frac{x-\bar{\Lambda}_j}{\sqrt{2\sigma_j^2}}\right)+1}{\sum_{k\neq i}\mathcal{A}_{jk}\left(\frac{1}{B_{jk}^{<>}+\phi''_{jk}}\right)}. \tag{93}$$

A backwards message with $y_{jk}+\varepsilon_{jk}$ or a forwards' message, where the fluctuations are incorporated within the $A_{ij}^{<>}$ and $B_{ij}^{<>}$ values, assume the expressions:

$$y_{jk} \leftarrow y_{jk} - \frac{A_{jk}^{<>}+\phi'_{jk}+\langle\mu_{ij}\rangle}{B_{jk}^{<>}+\phi''_{jk}}, \tag{94}$$

or,

$$y_{ij} \leftarrow \frac{B_{ij}^{<>}y_{ij} - A_{ij}^{<>} - B_{ji}^{<>}y_{ji} + A_{ji}^{<>} - \phi'_{ij} + \phi''_{ij}y_{ij}}{B_{ij}^{<>}+B_{ji}^{<>}+\phi''_{ij}}. \tag{95}$$

Using these derivations, we can compute a scenario where the demand and supply fluctuate and the algorithm provides more power to nodes with high uncertainty.

3.3.2 Characteristics

Figure 3.27 shows an identical optimisation problem as in Figure 3.19, now solved using the quenched equations, here it can be noticed that the reserve given to consumers with uncertainty is negligible, in fact only the generators maintain some reserve, and this appears to be due to minimising distribution costs.

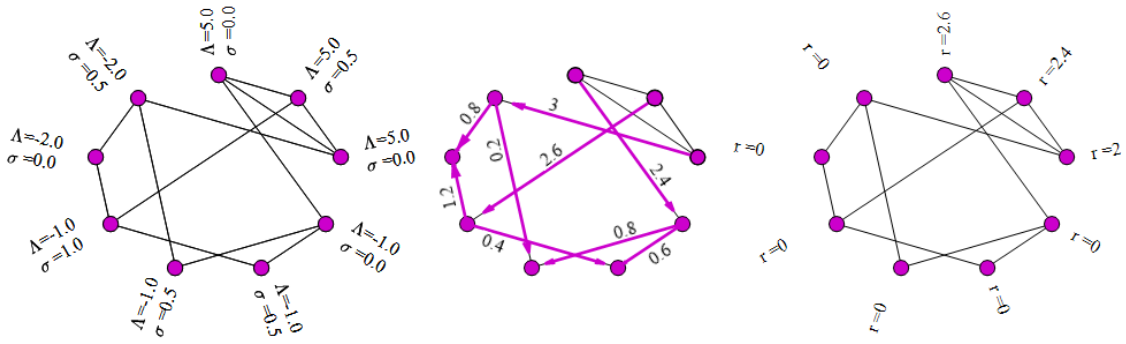


Figure 3.27 Example of power deployed to satisfy a simle 9-bus network calculated using quenched averaging (to the nearest 1 decimal place).

Figure 3.28 - Figure 3.40 show the amount of reserve power given to nodes as their uncertainty increases, in a 24-node synthetic IEEE network [154] (the same network and conditions as those used in the section on annealed averaging) calculated using quenched averaging. Figure 3.28 shows the case when one consumer's standard deviation is increased, we see that the reserve power increases very little at low volatility (Figure 3.29), reaching an almost linear increase at higher values. Comparing this increase with the z-line which all annealed approximations followed, we see that the quenched averaging reserve power rises slower, and gives a much higher probability of the node being unsatisfied.

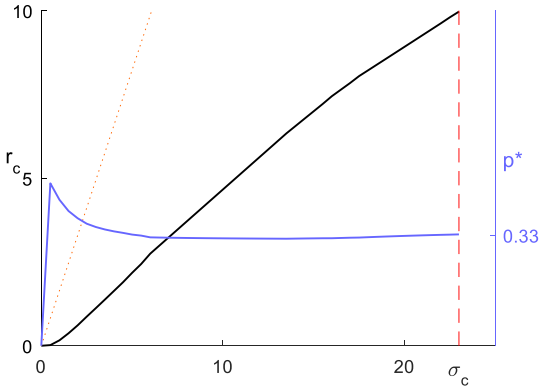


Figure 3.28 The excess given to a consumer of mean capacity $\bar{\Lambda} = -1$, as the standard deviation of the corresponding Gaussian probability function increases.

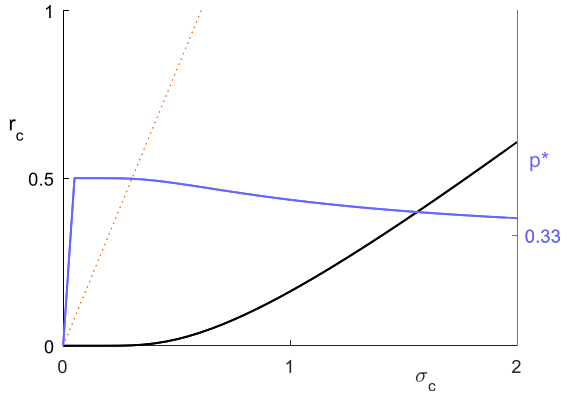


Figure 3.29 The initial rise as seen in Figure 3.28 magnified.

Further work shows that the capacity value of the consumer determines the initial increase of reserve (Figure 3.30); a higher absolute capacity value results in a slower initial increase, this is due to proportionality between the standard deviation and the mean value. It cannot be proposed that this is due to the capacities proximity to becoming unsatisfied, because all consumers have zero reserve as sigma equals zero. Figure 3.31 gives the extended curves of Figure 3.30, it shows how the algorithm assigns reserve in reaction to large sigma, and shows a linear increase for all capacity variations, each with the same gradient, but a higher negative capacity results in the curve being slightly to the right, with a slightly higher unsatisfaction level. For an increase in sigma, $\bar{\Lambda} = 0$ has a linear reserve increase which results in a constant unsatisfaction level of $\rho^* = 0.285$; still much higher than annealed. All other capacities start with a $\rho^* = 0.5$ when sigma is small; because negligible values of reserve are given when sigma has some level of uncertainty. The unsatisfaction curves tend toward 0.285 (this is not a fixed value; it depends on factors such as wealth of the network, connectivity, proximity to available resource, etc.). We propose that the difference in initial increase is due to the proportionality between $\bar{\Lambda}$ and σ, the increase in reserve after this appears to be linear because, as observed within annealed approximations, the more uncertain the node the more reserve it requires and the value of $\bar{\Lambda}$ becomes irrelevant as $\bar{\Lambda} \ll \sigma$.

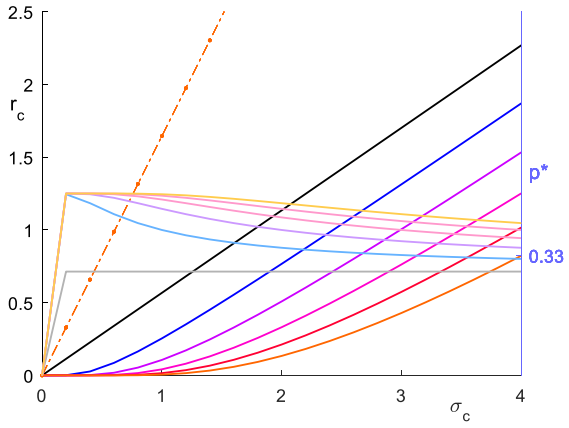


Figure 3.30 The excess given to one consumer as the standard deviation of the corresponding Gaussian probability function increases for $\bar{\Lambda} = 0$ (black), -1 (blue), -2 (purple), -3 (pink), -4 (red), -5 (orange). Pale equivalent colours represent the corresponding percentile provided by the excess. Orange dashed line represents the z-line.

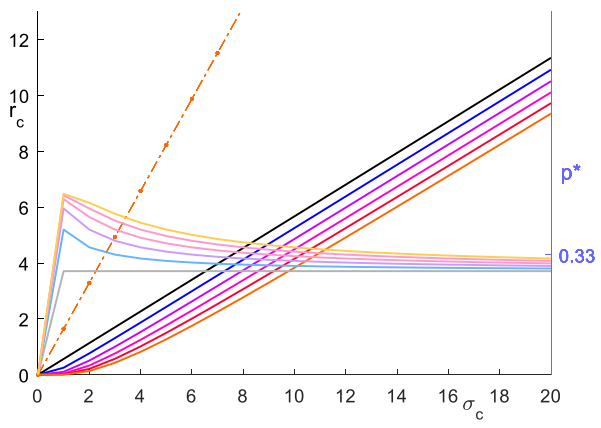


Figure 3.31 Extended curves of Figure 3.30, conditions the same as, showing higher values of sigma.

Figure 3.32 shows a network with one consumer, and six generators with ample excess reserve that are four or more generations away from the consumer (the shortest path between the generators and the consumer is 4 nodes away). The black line represents the increase in reserve given to the consumer when there is one additional generator with sufficient power connected as a 1st generation neighbour to the consumer. Blue shows the same example, but the generator is a 2nd generation neighbour (there is one node between them), purple shows a generator which is a 3rd generation neighbour and pink uses the 4th generation neighbours already there. The graph shows how when the consumer has a generator close by, it receives an almost linear reserve for increasing sigma. When the consumer must rely on generators further away, it is given less reserve, decreasing the confidence of the network. This shows that the quenched averaging algorithm considers the distribution cost objective function within its decision of how much reserve power to provide to each uncertain node; so, when the cost of distribution is higher, the priority of providing reserve to uncertain nodes is slightly lowered, resulting in a more gradual increase.

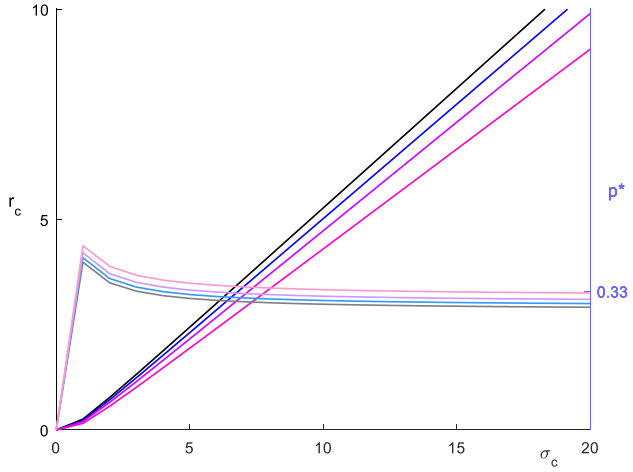


Figure 3.32 The excess given to one consumer as the standard deviation of the corresponding Gaussian probability function increases. There are six generators with large amounts of power, connected by a minimum of 4 neighbouring nodes away. In black, there is an additional generator with large amounts of power directly neighbouring the consumer, blue shows the additional generator connected one neighbour away from the consumer, purple shows a generator connected 2 neighbours from the consumer and red shows no additional generator. The lighter shade of each colour corresponding percentile.

Figure 3.33 shows an example of a network with one consumer, there is one generator that directly neighbours the consumer node and six others which have large amounts of power, but are connected by a minimum of 4 neighbouring nodes away. The neighbouring generator has power 40 (black), 10 (blue), 5 (purple), 2 (pink) and 1 (red). The lighter shade of each colour represents the respective confidence level. To minimise distribution cost, if possible, the

neighbouring generator provides most of the reserve power sent to the uncertain consumer (the consumer has a connectivity 2, some power is passed from the further generators along the path not connected to directly to the closer generator). Once the reserve power required at the consumer reaches the available power at the neighbouring generator, if the consumer continues to require more power, this will be at a far higher distribution cost. From Figure 3.32 we know that a higher distribution cost results in a gentler gradient increase in reserve over increasing sigma, and this gradient decrease can be seen in the red, pink and purple curves once the required reserve power exceeds the available reserve at the neighbouring generator. The jump before this change in gradient is a result of the necessary redistribution in power flow over the network, the increase in distribution cost and the reweighting between distribution cost and unsatisfaction levels. It can be seen for each initial capacity of the neighbouring generator, the jump in reserve happens at each point where the generators reserve has been used up, and a more gradual gradient is adopted; this results in a drop in the unsatisfaction percentile at these points. Note: not all redistribution within the network or fully utilised generators results in these jumps; fully utilised close-by generators sometimes result in a more gradual increase, due to the increase in distribution cost when having to take power from further generators, but without a discontinuity, and occasionally the redistribution of power over edges does not affect the reserve at uncertain nodes at all.

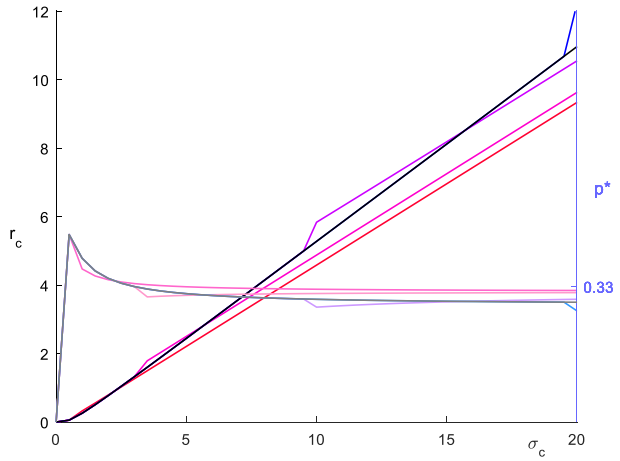


Figure 3.33 The excess given to one consumer as the standard deviation of the corresponding Gaussian probability function increases. There is one generator that is neighbouring the consumer node and the six other generators with large amounts of power, connected by a minimum of 4 neighbouring nodes away. The neighbouring generator has power 40 (black), 10 (blue), 5 (purple), 2 (pink) and 1 (red). The lighter shade of each colour is the corresponding percentile.

Figure 3.34 shows the reserve given to one uncertain generator. An increase in reserve starts at lower volatility levels than the in the annealed approximation case, but rises much slower throughout. Figure 3.35 shows the reserve of each consumer in the network as all their standard

deviations increase together. We see that this reacts differently to the annealed approximation allocation because they all increase at different rates, this is due to the algorithm weighing the cost of distribution against the probability of the constraint being unsatisfied (unlike with annealed, where the constraint is strictly obeyed without being affected by the transportation cost). In this example the variation in increase rates between consumers is due to each nodes proximity to available reserve. Interestingly, a jump can be seen between $\sigma = 1.6$ and 1.7; this appears to be due to a redistribution in power flow over the network, similar to cases from Figure 3.33. Further work suggests the redistribution is triggered by one edge reversing its power flow, affecting most edges; 20% significantly (Figure 3.36). The redistribution causes minimal change in distribution cost (Figure 3.37), but allows additional reserve at the node. It is important to mention that this jump and redistribution of power is not in real time, and that it only refers to how the network would react at different levels of sigma, which would usually remain constant in a real-time power network; therefore, any concern about discontinuity problems are irrelevant.

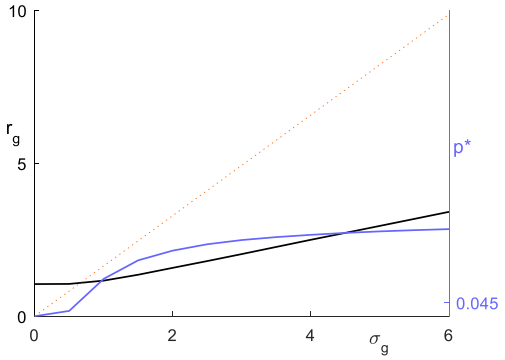


Figure 3.34 The excess given to a generator of mean 1.5, as the standard deviation of its Gaussian probability distribution increases.

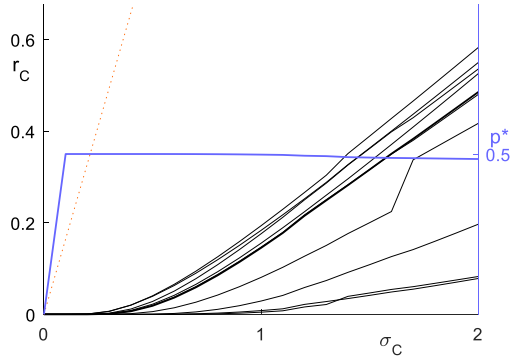


Figure 3.35 The excess given to all consumers, as their standard deviation increase uniformly.

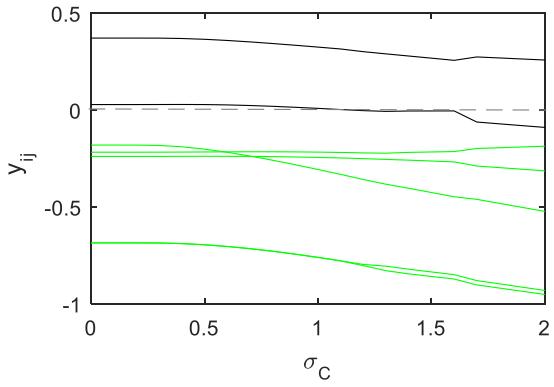


Figure 3.36 The edges effected by the redistribution at $\sigma = 1.6$. Black edges indicate those connecting to the jumping node, green represents other edges that are most effected by the redistribution.

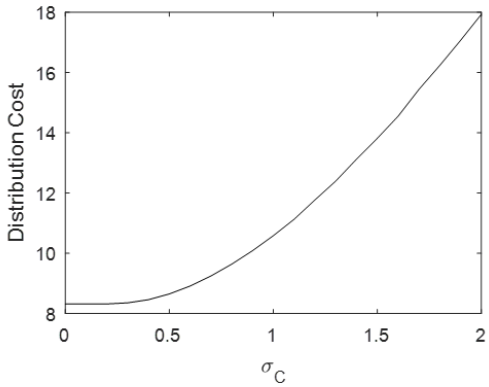


Figure 3.37 The distribution cost of the network as sigma increases for the case in Figure 3.35.

Figure 3.38 shows an example where all generators have increasing standard deviation; they initially split into two sections of increasing and decreasing reserves; this happens when the generators with little or no reserve power require a positive reserve as sigma increases, and the closest generators with available reserve provide this. This is shown in Figure 3.39, where there are two nodes starting with no reserve connected explicitly to another with reserve; it can be seen that as both generators with zero reserve at $\sigma = 0$ are increased, the reserve at the other decreases to accommodate this. Figure 3.40 also shows this, where the reserve of consumers and generators starting at low reserve increase as sigma increases, using the power of closely connected generators that do have available reserve.

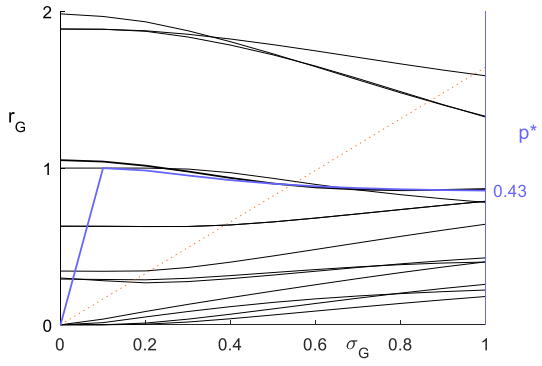


Figure 3.38 The excess given to all generators as their standard deviations are increased uniformly from 0 to 1.

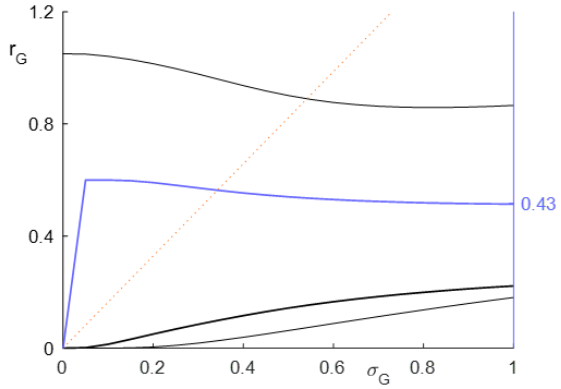


Figure 3.39 The excess at three closely connected generators from Figure 3.38 and their corresponding reserve.

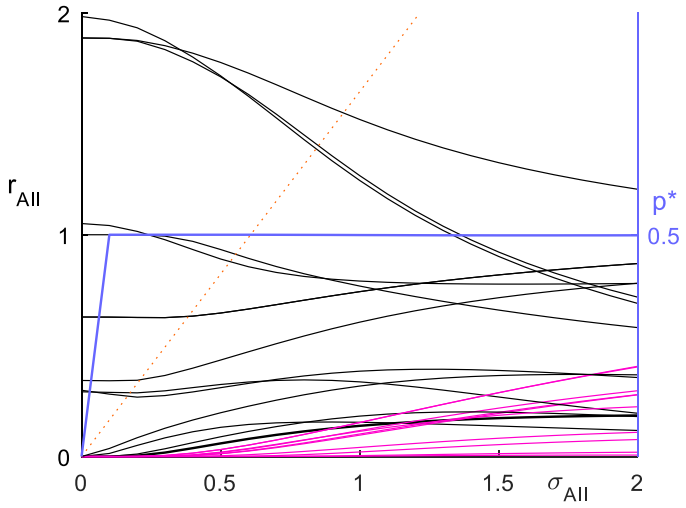


Figure 3.40 The excess given to all nodes as their standard deviations are increased uniformly from 0 to 2.

Figure 3.41 - Figure 3.45 show how reserve power in the network reacts for larger values of σ. (Note: These are for test and understanding purposes only. All capacities in these examples are between -1 and 2, therefore standard deviation values beyond 2 show a large amount of volatility, uncharacteristic of any uncertainty expected to be experienced from renewable energy sources.) At the point $\sigma > 2$, all nodes require extra levels of reserve; this is where the network reserve dynamics change, independent of original gradients and reserve. In Figure 3.41 where all consumers have increasing sigma, an obvious change in reserve occurs around $\sigma = 5$; this is the point where all available resource from deterministic generators have been depleted. We see that as this point approaches there are a lot of discontinuities and unexpected changes in gradient; presumably because the tight margins of available resource result in regular redistributions of the network. Once available power is gone, the gradients become smooth. At this point for all three figures, we see that once reserve is depleted the nodes tend to random reserve values between 0 and 2; this appears to be irrespective of consumer or generator nodes, and reserve values at $\sigma = 0$ or $\sigma = 1$. Figure 3.44 shows that the increase in sigma continues to increase the distribution cost of the network, even for large sigma values. The initial steady distribution cost as sigma increases is due to the slow initial reaction of the reserve according to σ as seen in Figure 3.29. Figure 3.45, which displays when all nodes have increasing sigma, shows how the reserve at high sigma appears to be dependent on the connectivity of each node, where nodes with less connections are given higher levels of reserve. One possible reason for this, is that each node is iteratively asking for more and more power, and with lower connectivity nodes there is a smaller effect of diffusion of responsibility, and so the lower connected nodes receive more power.

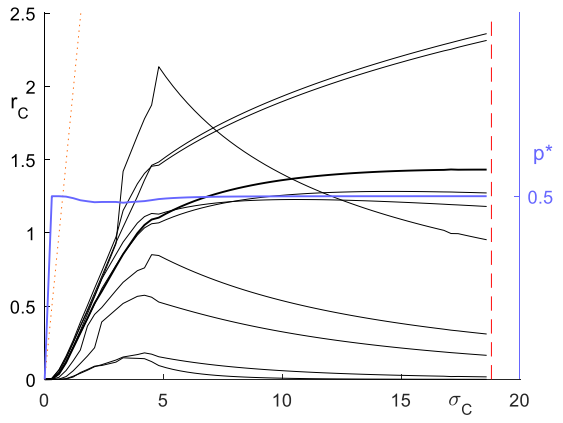


Figure 3.41 The excess given to all consumers as their standard deviations are increased uniformly from 0 to 20.

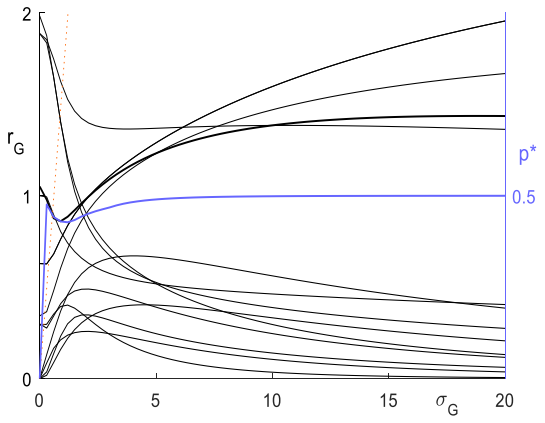


Figure 3.42 The excess given to all generators as their standard deviations are increased uniformly from 0 to 20.

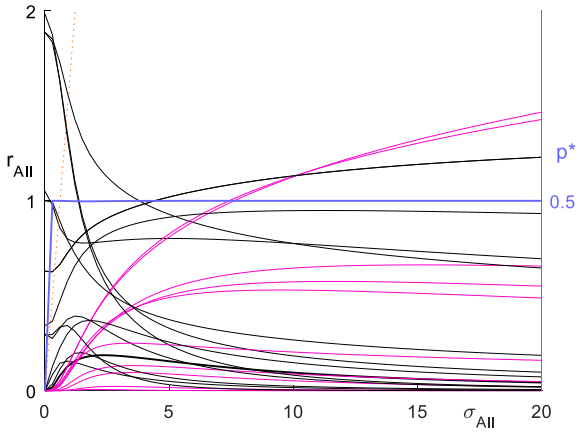


Figure 3.43 The excess given to all nodes as their standard deviations are increased uniformly from 0 to 20.

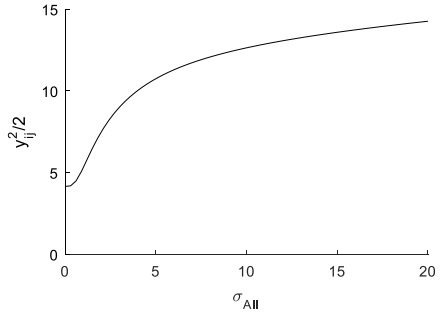


Figure 3.44 The distribution cost when all nodes in the network have their standard deviations increased uniformly from 0 to 20, as seen in Figure 3.43.

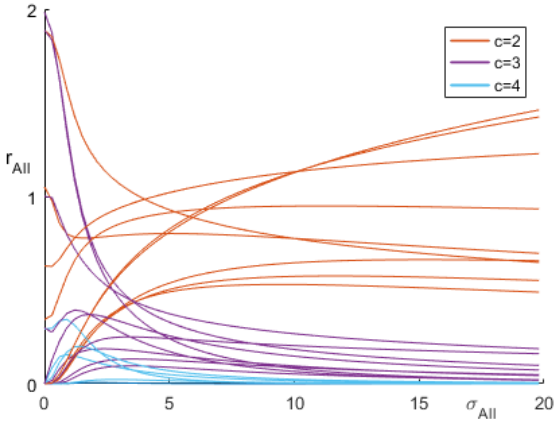


Figure 3.45 The reserve power given to all 24 nodes in a network as their standard deviations increase simultaneously (from Figure 3.43), coloured in terms of their individual connectivity values.

The advantages of using quenched averaging is that the algorithm inherently considers the uncertainty within the network; however, we can see throughout these results that this type of averaging allows for the network to be satisfied with only a 50-72% confidence level in most cases; which makes it arguably a worse choice for increasing the reliability of power grids than the annealed approximation. We speculate that the reasoning behind such small reserve is that the function we minimise is the expected VFE; it includes a hard constraint which outputs 0 if the node is unsatisfied and 1 if it is. It therefore may be biased towards a small range of values that dominate the expected VFE due to the probability of occurrence coupled with a high probabilistic weight, possibly due to high cumulative transportation costs, at the expense of unsatisfied nodes with a lower probabilistic weight.

3.4 Quenched Averaging with Additional Resource

3.4.1 Method

Due to the small excess suggested by the quenched averaging approach for a fluctuating node, we have taken advantage of the benefits of combining the quenched averaging approach with the extra resource suggested by the annealed approximation. More specifically, we will be continuing to use the equations obtained from quenched averaging, but adjust the resource Λ_j to $\Lambda_j + \sqrt{2\sigma_j^2}\text{erf}^{-1}(2\rho_j - 1)$. This will allow fluctuations to be inherently considered, with the additional ability to predetermine the level of confidence required.

3.4.2 Characteristics

Looking at similar situations as Figure 3.20 - Figure 3.25 with lines and colours denoting the same things, Figure 3.47 - Figure 3.52 show the equivalent conditions, using a quenched with additional resource (QAR) algorithm. Figure 3.46 shows an example of power grid optimisation using QAR on a small 9-node network. The results appear to be equivalent to those of Figure 3.19, indicating that the same level of reliability is achieved as in annealed averaging, presumably because of the significant resource allocated by it.

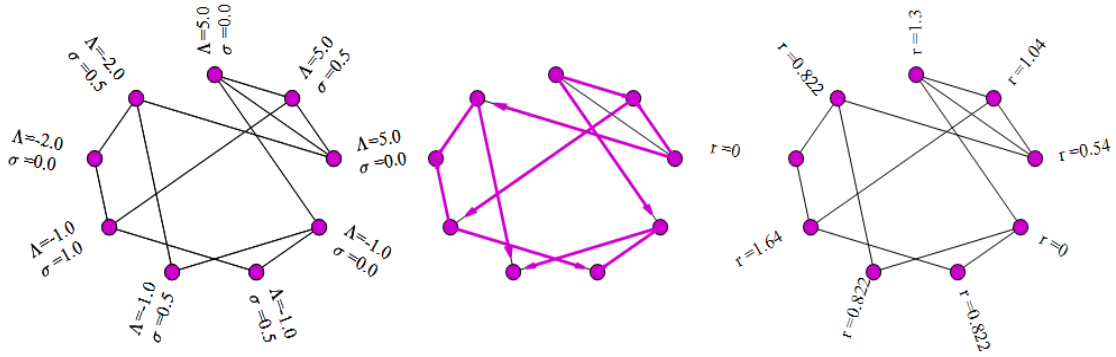


Figure 3.46 An example of power grid working on 9-node network for quenched averaging with soft constraints.

All nodes with increasing sigma are given a reserve at or over the z-line. In Figure 3.47 and Figure 3.48 we see that the reserve sits slightly above the dotted line representing the z-line that annealed followed, giving a stronger confidence than the predetermined $\rho = 0.05$ required.

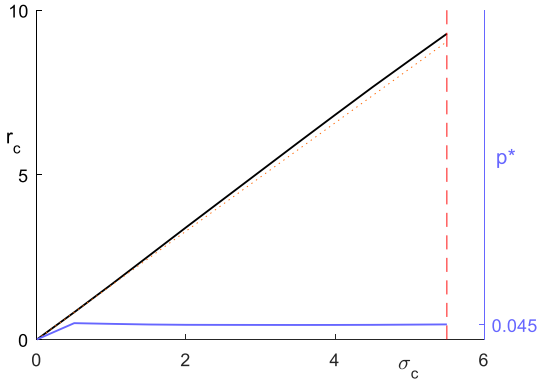


Figure 3.47 The excess given to a consumer of mean -1, as the standard deviation of its Gaussian probability function increases using a combined quenched averaging technique and additional resource.

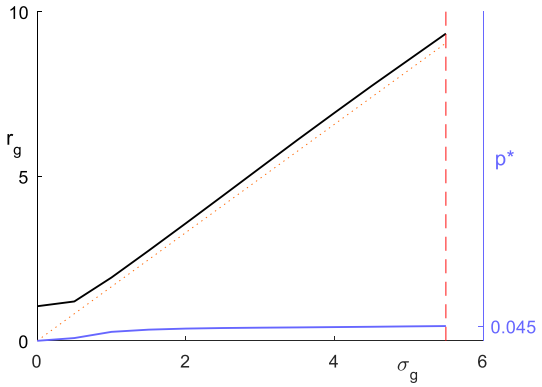


Figure 3.48 The excess given to a generator of mean 1.5, as the standard deviation of its Gaussian probability distribution increases using a combined quenched averaging technique and additional resource.

Figure 3.49 is almost identical to its annealed equivalent, presumable because the extra reserve is minimised due to the high levels of volatility throughout the network. In Figure 3.49 - Figure 3.52 the graphs appear very similar to annealed results, but smoother in places. The smooth curves follow from the quenched averaging's capability to acknowledge the presence of uncertainty within the network, and therefore provides further reserve where appropriate. Figure 3.51 also demonstrates how increasing sigma can sometimes result in discontinuous curves due to redistribution of flow over edges throughout the network.

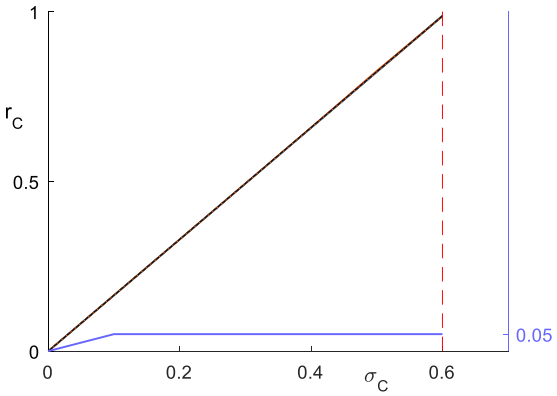


Figure 3.49 The excess given to all consumers as their standard deviations increase using a combined quenched averaging technique with extra resource.

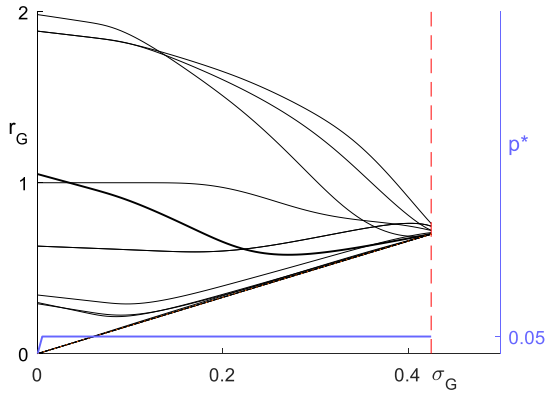


Figure 3.50 The excess given to all generators as their standard deviations increase using a combined quenched averaging technique with extra resource.

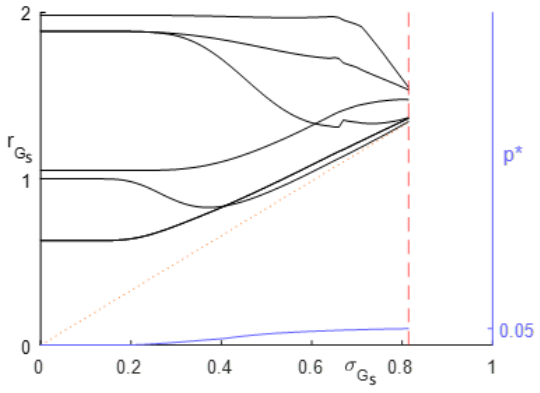


Figure 3.51 The excess given to generators, with available reserve at $\sigma = 0$, as their standard deviations increase using a combined quenched averaging technique with extra resource.

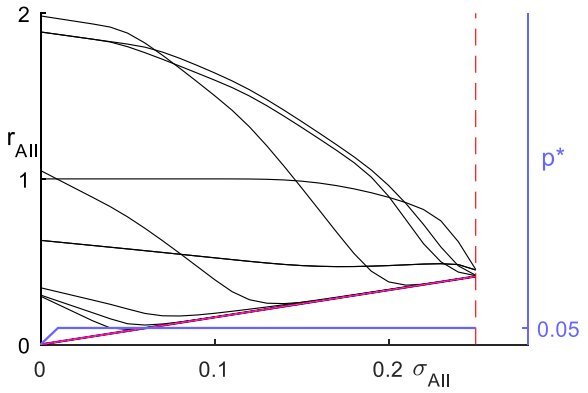


Figure 3.52 The excess given to all nodes in the network as their standard deviations increase using a combined quenched averaging technique with extra resource.

Looking in more detail at how quenched averaging with a fixed confidence level compares with annealed approximation and quenched averaging separately, Figure 3.53 shows in blue the quenched averaging result, in comparison to the yellow line which represents the annealed approximation result. The orange line represents the reserve given by QAR minus the annealed approximation result. We can see how the extra reserve power given to a volatile node using QAR minus the annealed approximation result and is much smaller than quenched averaging result with no additional resource. This verifies that the QAR technique is not merely the addition of the reserves given from quenched averaging and annealed averaging, but that it acknowledges that each distribution is asking for a specific confidence level, and aims to satisfy this in the presence of the uncertainty and the situation of the network (topology, distribution costs, available reserve, etc.).

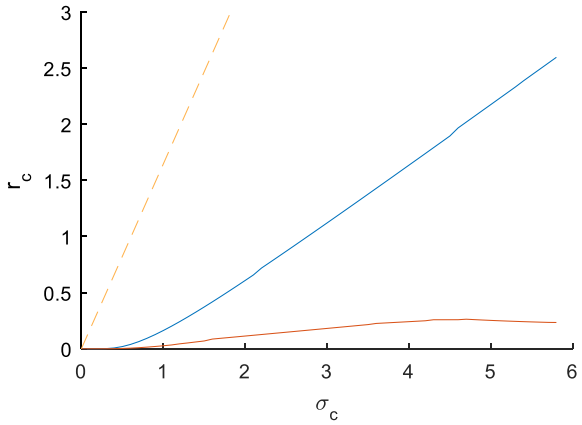


Figure 3.53 Reserve given to a volatile consumer when distributed using annealed averaging (dashed yellow), quenched averaging (blue) and the difference between the reserve given through QAR and annealed averaging (solid orange).

Using the QAR method allows us to inherently consider fluctuations with a predetermined value of confidence, as the figures show; the QAR method will be used for the following graphs.

3.5 Results

Figure 3.54 shows an example of the IEEE 118-Bus [156] with normally generated capacities (about $\mathcal{N}(1,1)$) and fixed standard deviation of 0.5. Three algorithms have been applied to optimise the power flow: based on message passing without considering fluctuations (Section 2.3); with soft constraints (Section 3.2) where $\rho = 0.3$; and with soft-constraints and message passing that incorporates fluctuations (Section 3.4) where $\rho = 0.3$. The size of the red dots signifies the overall capacity deficit for each node over a 60-minute time window when every second the *actual* capacity of each node was randomly generated from its own probability distribution $\mathcal{N}(\bar{\Lambda}_j, 0.25)$. Incorporating soft-constraints with message passing that considers fluctuations (QAR) gave the most reliable results. The soft-constraints were set to satisfy the network with 70% confidence for demonstration only; real networks would require a much higher confidence level. The figure highlights the advantage of incorporating fluctuations in the optimisation algorithm. Its related computational cost was negligible.

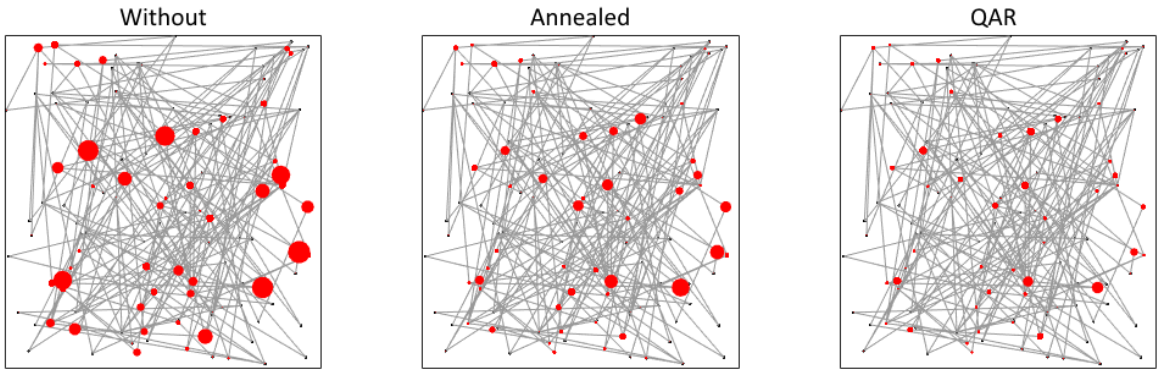


Figure 3.54 The size of the red dots indicate constraint violations (the sum of the capacity deficit of each node when capacities were chosen from the random variable using $\mathcal{N}(\bar{\Lambda}_j, 0.25)$ every second over a 60-minute time window) when network capacities are fluctuating (mean capacities were chosen from a normal distribution $\bar{\Lambda} = \mathcal{N}(1,1)$ and each node had a standard deviation of $\sigma = 0.5$). Power flow was calculated without considering fluctuations (left); when fluctuations are considered using the annealed approximation with $\rho = 0.3$ (centre); when we consider fluctuations within derivations using QAR with $\rho = 0.3$ (right) [157].

Figure 3.55 shows the level of extra resource given to each node according to its own sampled variance from the Gaussian distribution $\sigma \sim \mathcal{N}(\sigma^m, 0.25)$, where the mean standard deviation value is $\sigma^m = 0$ (blue), 1.5 (orange), 3 (yellow). The black dashed line represents the extra resource a node should receive according to the confidence level, ρ. Blue '×' symbols show network solution values when abundance of resource is available and overall volatility is low. We see that excess at some nodes is high, and all nodes are above or on the line. Orange '•' symbols show that as the overall volatility increases the generators' excess decreases since more resource is required to satisfy a given threshold. Yellow '○' symbols represent a case where the network is unable to satisfy the soft constraints of each node due to insufficient power, the algorithm nevertheless aims to satisfy each node appropriately per its volatility. The QAR algorithm addresses fluctuations in generation/demand and even when soft constraints are unsatisfiable it appears to provide an appropriate optimal solution, where nodes with higher volatility receive more power reserve. When power is available we see that all volatile nodes are given reserve equal to or higher than the calculated z-line, and any extra reserve above this is distributed with some correlation to volatility, increasing the reliability of the system. The inset shows the excess given to each node per its standard deviation, considering networks with different average capacity values, showing that the overall excess in the network increases with the average capacity. Note: All networks are synthetic randomly generated random regular graphs of size $N = 100$, $c = 6$, and averaged from 10 reiterations, unless stated otherwise.

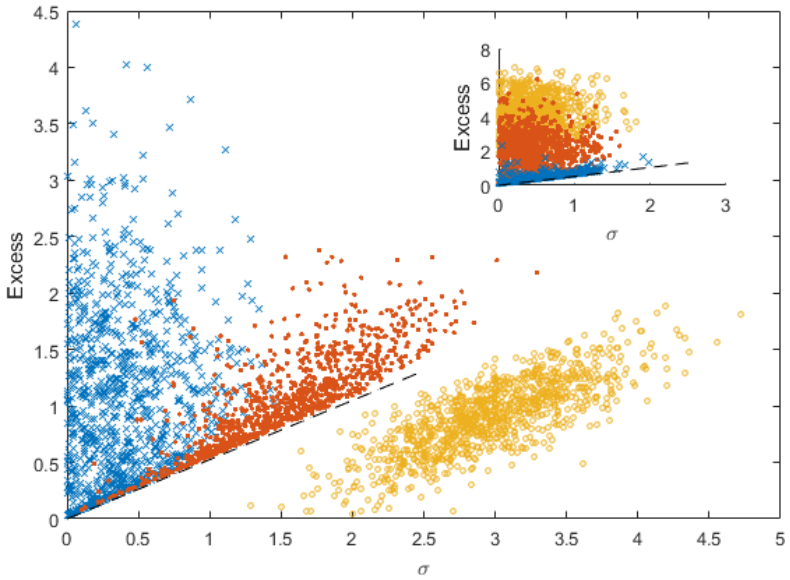


Figure 3.55 Excess given to each node against its standard deviation σ, drawn from a truncated (the positive part of a) Gaussian of standard deviation 0.5 and means 0 (blue ×), 1.5 (orange •), 3 (yellow ○); for ten randomly generated 100-node networks of connectivity 5, randomly connected, with Gaussian average capacities ($\mathcal{N}(1,1)$). Inset shows the case where average capacities are randomly generated from a Gaussian of mean 0 (blue ×), 2 (orange •), 4 (yellow ○), and variance 1 [157].

Figure 3.56 shows a scatter diagram of the reserve power left at each node according to their individual ρ values which are randomly sampled from a uniform distribution between 0.05 and 0.2, denoted; $\mathcal{U}(0.05,0.2)$. Each network consists of 50% consumers with $\bar{\Lambda} = -1$ and 50% generators with $\bar{\Lambda} = 1.2$ (This will be denoted[8] as $\mathcal{F}(-1,1.2)$) and sigma is fixed at the values $\sigma \sim \mathcal{F}(\sigma^m)$ where $\sigma^m = 0.025$ (blue), 0.05 (orange) and 0.075 (yellow). We see that the minimum reserve power decreases as ρ increases; decreasing the level of reliability of the network. This graph demonstrates how the network can assign individual values of ρ to different nodes; essentially assigning the level of importance of individual nodes being satisfied. We see how individual nodes that require less reliability receive less reserve (this could be an alternate method to our suggested load shedding approach in Section 4). It can also be seen that for larger individual σ_j values, the minimum reserve power increases; providing extra security for more volatile nodes. The extra reserve given to nodes with higher values of σ can be seen in Figure 3.57, and to satisfy this constraint we see an overall decrease in generators' available reserve.

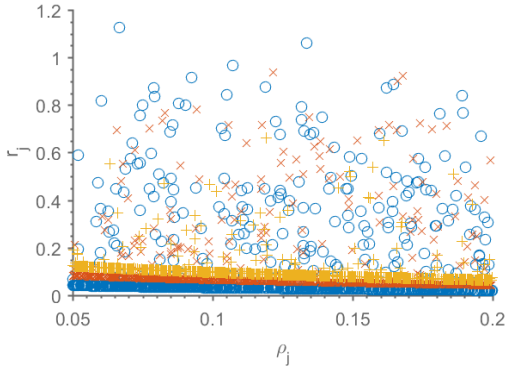


Figure 3.56 The reserve power at all nodes from ten 100-node networks of random regular topology of $c = 6$ and $\bar{\Lambda} = \mathcal{F}(-1,1.2)$ as ρ increases where $\sigma = 0.025$ (blue), 0.05 (orange) and 0.075 (yellow).

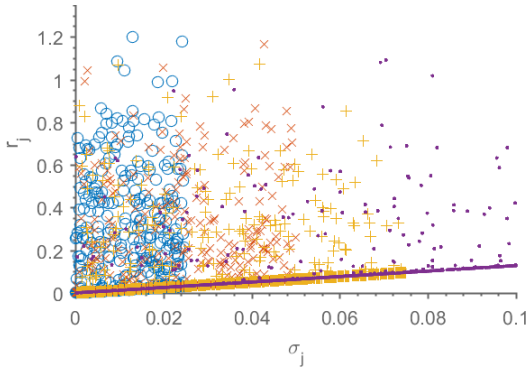


Figure 3.57 The reserve power at all nodes according to each nodes sigma value, from ten 100-node networks with a random regular topology of $c = 6$ and capacity, $\bar{\Lambda} = \mathcal{F}(-1,1.2)$, where σ is randomly chosen from the distribution $\mathcal{U}(0,\sigma^+)$ and $\sigma^+ = 0.025$ (blue), 0.05 (orange), 0.075 (yellow) and 0.1 (purple).

Figure 3.58 shows a Weibull best fit distribution of power sent, $P(S)$, from generators in a 100-node network with random regular connectivity 6 and capacities sampled from $\mathcal{F}(-1, 1.2)$ in a network that considers fluctuations. We see how the increased value of the fluctuations σ results in a higher average amount of power sent and a higher majority of maximised capacity

[8] $\mathcal{F}(\bullet)$ denotes a dirac delta distribution, equally partitioned for any values inside the function.

generators. This is because as the volatility of the network increases, more reserve power is required to increase security and so the maximum level of more generators is reached. Figure 3.59 shows the same distribution network but demonstrates the reserve power given to each consumer as a best fit beta distribution, it shows how for low levels of uncertainty the reserve at all consumer nodes is low, but for higher volatility among consumers, reserve is increasing.

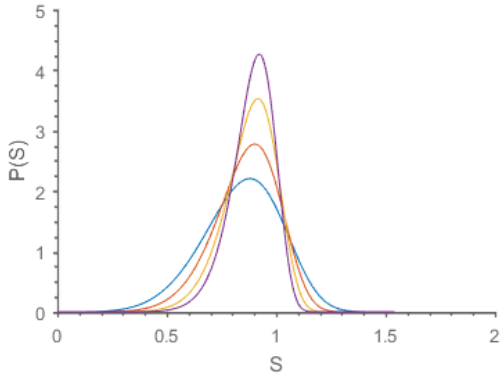


Figure 3.58 The best fit of power distribution sent, S, from generators, where $\sigma = \mathcal{U}(0, \sigma^+)$, and $\sigma^+ = 0.025$ (blue), 0.05 (orange), 0.075 (yellow) and 0.1 (purple).

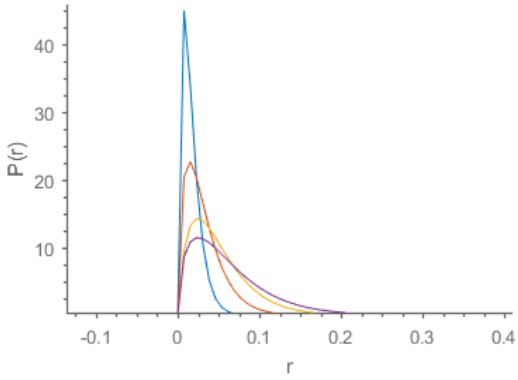


Figure 3.59 The best fit of reserve power distribution at consumers, where $\sigma = \mathcal{U}(0, \sigma^+)$, and σ^+=0.025 (blue), 0.05 (orange), 0.075 (yellow) and 0.1 (purple).

We tested ten networks of $N = 100$ and plotted histograms of the distribution of power flow along edges, y, for an increasing value of degree connectivity c, as shown in Figure 3.60. A half normal distribution curve was drawn from these values in Figure 3.61 which clearly shows how the distribution of power along edges becomes sharper as c increases. This is because although the network aims to minimise distribution costs, the chance of consumers being connected to sufficient reserve power is smaller and the characteristic path length will be longer, so requested power must travel further distances to satisfy consumers.

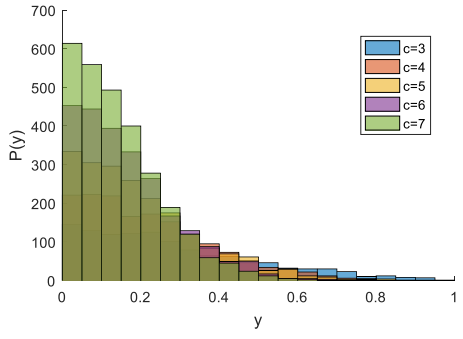


Figure 3.60 Histogram of power distribution over edges for different connectivity values c and system size $N = 100$. The higher the connectivity, the smaller current values are [158].

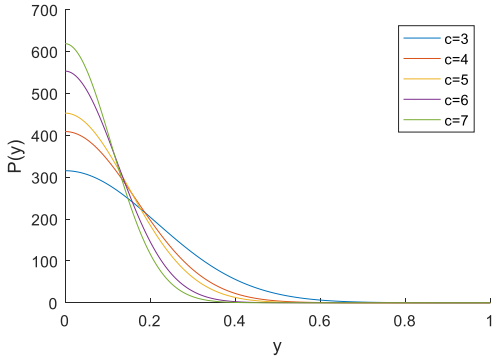


Figure 3.61 Half-normal distribution best fit curve of histogram in Figure 3.60; $N = 100$ [158].

Figure 3.62 also shows that when volatile nodes have a high connectivity, the electrical grid has an overall reduced distribution cost. This is because a volatile node is more likely to be closer to a node with the power it requires, reducing the expected flow over edges. This could be useful to understand at the design stage of power grids, as it suggests the advantages of introducing renewable sources in highly connected areas. Figure 3.63 shows that the reserve given to a volatile node decreases as the shortest path between the volatile node and available generators increases. This demonstrates that the excess given to a volatile node is largely effected by the distribution cost to reduce the costs incurred by moving resource from far off locations at the expense of providing less additional reserve. It may suggest the necessity of weighting the distribution cost so that it has a smaller impact on adjusting the reliability of the network.

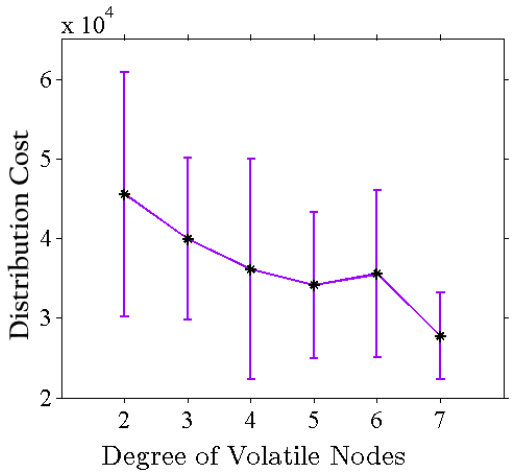


Figure 3.62 The distribution cost when uncertainties are put on 4 nodes each with a degree 2, 3, 4, 5, 6 and 7, in an 80-node network [159]. © Copyright Clearance Centre [2016] IEEE.

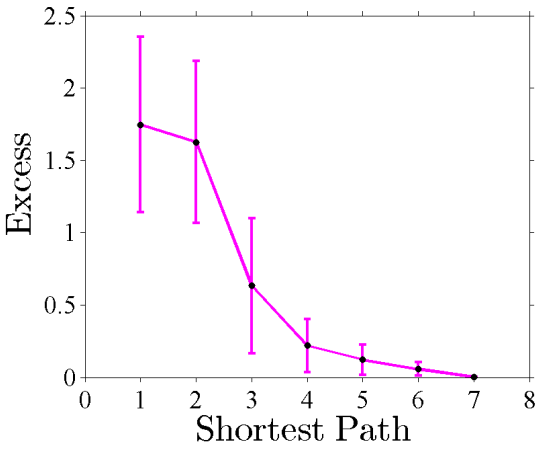


Figure 3.63 Excess provided to a volatile consumer node as the generators move on average further from the node, increasing the transportation length and therefore the distribution cost [159]. © Copyright Clearance Centre [2016] IEEE.

Using the QAR algorithm, we calculate the effect of uncertainties on the quadratic distribution cost. Figure 3.64 considers five cases of different fractions of volatile nodes in the grid the *overall* σ value is distributed between. It shows that the distribution of volatility makes a huge difference to the distribution cost and the importance of having smaller and more spread out renewable sources throughout the grid instead of having one large farm concentrated in one place. This could be beneficial for the design and structure decisions for future grids. It is known that a decentralised grid also increases its robustness against structural perturbations [160]. The graph does show however, that if all nodes share the overall volatility, this does not give the lowest costs, of the cases available; numerically, 75% shows the lowest cost.

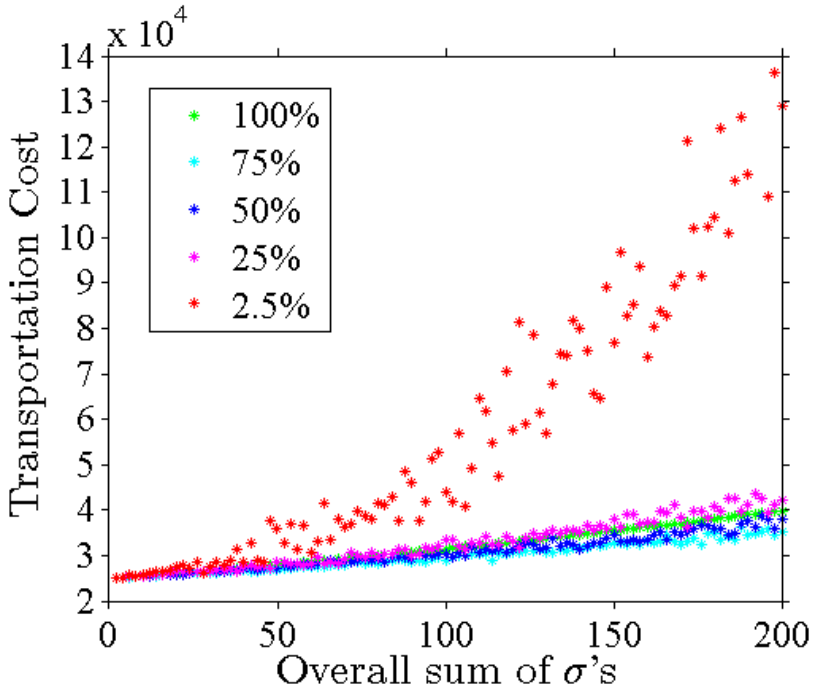


Figure 3.64 Distribution cost as the standard deviation σ increases using QAR averaging, when the overall standard deviation σ is distributed between 100%, 75%, 50%, 25% or 2.5% of the nodes (in a 40-node network) [159].

Figure 3.65 indicates that as the average capacity of a network increases, the number of edges used decreases until all edges are unused (as there are no more consumers). We can also see that for higher volatility, edges become fully utilised for lower mean values due to the need for additional resource across the network. The fraction of unused edges appears to follow a cumulative Gaussian distribution and the average capacity increases. The inset demonstrates the effect of the overall standard deviation on the number of edges used, showing that the more volatile a network (higher standard deviation), the more edges are used.

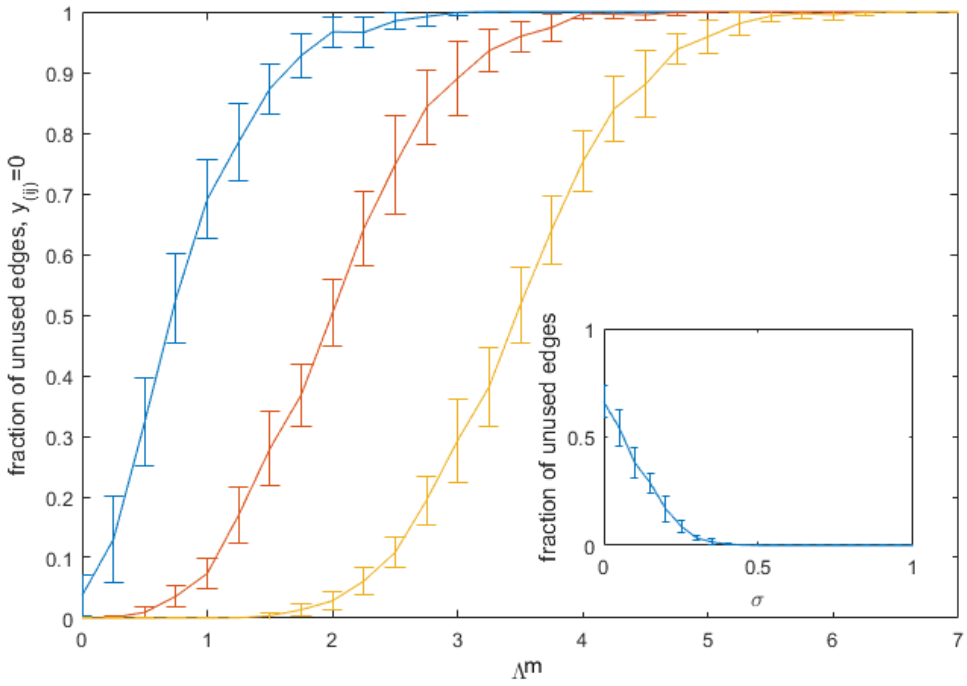


Figure 3.65 How increasing average capacity, Λ^m, in a randomly generated 100-node network effects the fraction of unused edges; when standard deviation at every node is 0 (blue, left), 0.25 (orange, centre) and 0.5 (yellow, right), with average capacities randomly drawn from $\mathcal{N}(\Lambda^m, 1)$. Inset shows how increasing standard deviation (with average capacities 1) effects the fraction of unused edges. Error bars indicate the deviation from the mean fraction of unused edges when ran ten times for randomly generated networks of given size, connectivity and capacity distribution [157].

It can be seen in Figure 3.66 and Figure 3.67 how the increase in the networks random regular connectivity results in a large decrease in power distribution cost (loss). From Figure 3.66 we see how increasing σ marginally increases the distribution cost, and Figure 3.67 shows how smaller values of ρ give a higher distribution cost. Both graphs demonstrate that at higher connectivities, the values of σ and ρ have less effect on the distribution cost; we assume that this is because the effects of high connectivity negate those of σ and ρ.

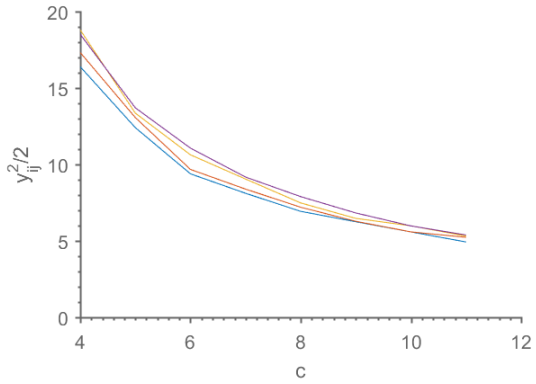


Figure 3.66 The distribution cost as the random regular connectivity (degree) increases, where $\sigma = \mathcal{U}(0, \sigma^+)$, and $\sigma^+ = 0.025$ (blue), 0.05 (orange), 0.075 (yellow) and 0.1 (purple) and average capacities are fixed at $\mathcal{F}(-1,1.2)$.

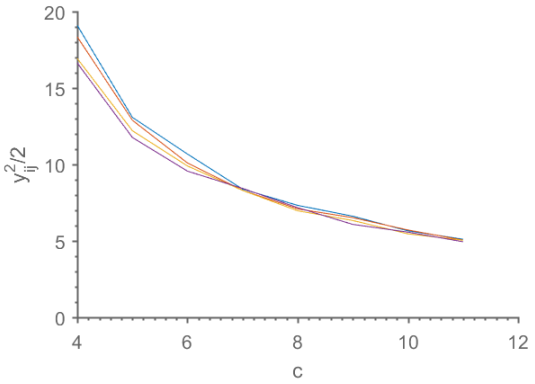


Figure 3.67 The distribution cost as the random regular connectivity (degree) increases, where $\rho = 0.05$ (blue), 0.1 (orange), 0.15 (yellow) and 0.2 (purple), fixed. And average capacities are fixed at $\mathcal{F}(-1,1.2)$.

Figure 3.68 shows how increasing the value of ρ results in a lower excess of power given to consumers and that a low σ value also decreases the reserve given to nodes. The line graph in Figure 3.69 shows how this affects the distribution cost and that because less reserve at uncertain nodes is required as ρ increases, there is less need for high power flow and therefore a decrease in distribution cost, it is also obvious that a lower σ decreases the distribution cost further due to a smaller requirement of reserve at volatile nodes. The non-monotonic decrease in the curves is because transportation cost depends on the individual topology of each network and so has a high variation in values.

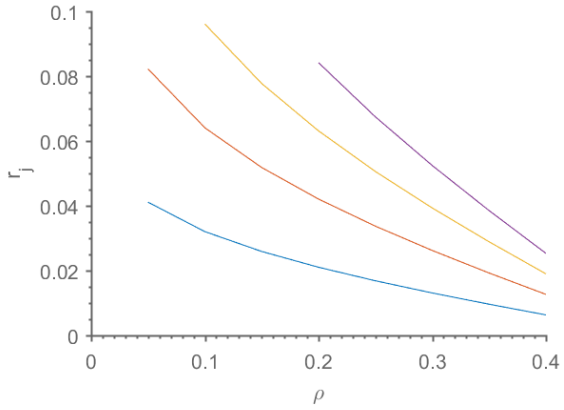


Figure 3.68 The average reserve power given to nodes as the percentile ρ increases and $\sigma = \mathcal{F}(0.025$ (blue), 0.05 (orange), 0.075 (yellow), 0.1 (purple)).

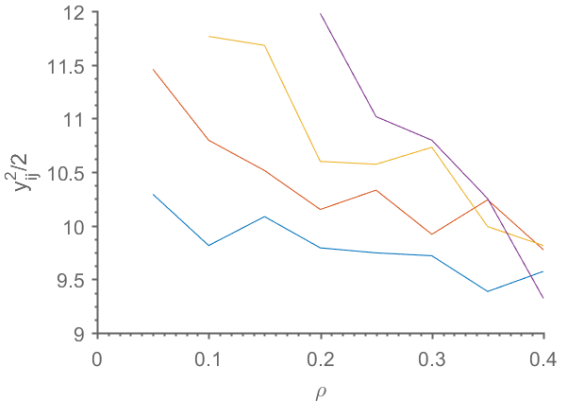


Figure 3.69 The distribution cost as the percentile ρ increases and $\sigma = \mathcal{F}(0.025$ (blue), 0.05 (orange), 0.075 (yellow), 0.1 (purple)).

Both Figure 3.70 and Figure 3.71 have been plotted using a bounded line, indicating the minimum to the maximum values in shaded colour, and the average as a solid line. We see in Figure 3.70 where σ is increasing equally for all nodes, how the reserve power given to consumers increases due to higher uncertainty. The graph also shows high reserve for lower values of ρ to increase the confidence of the network. When the σ values are uniformly chosen between $\mathcal{U}(0, \sigma)$ in Figure 3.71 we see that the reserve is also uniformly distributed between the mean from Figure 3.70 and 0. This is more representative of a real network as the uncertainties may be varied. Where the lines in Figure 3.70 end is when the network is unable to satisfy all nodes due to insufficient levels of reserve, whereas for distributed levels of uncertainty allow the network to continue satisfying nodes due to lower levels of reserve.

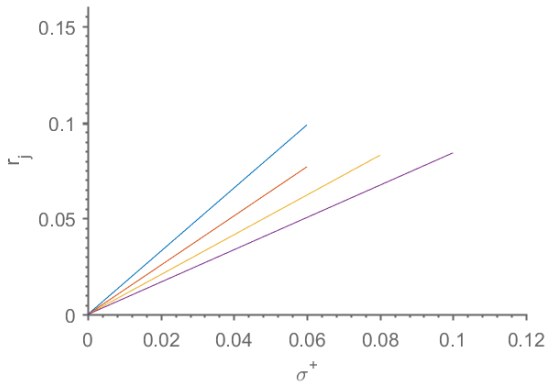


Figure 3.70 The reserve power given to consumers as fixed σ increases for $\rho = 0.05$ (blue), 0.1 (orange), 0.15 (yellow) and 0.2 (purple). Average capacities are chosen from $\mathcal{F}(-1,1.2)$.

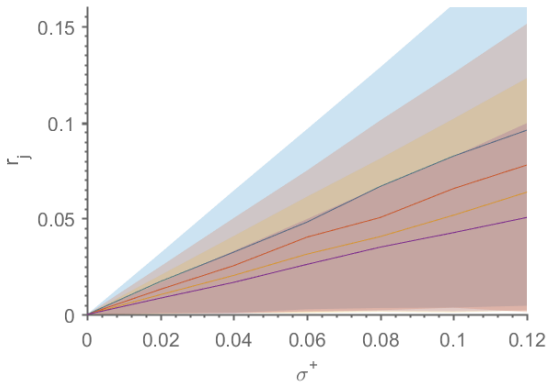


Figure 3.71 The reserve power given to consumers as σ increases uniformly $\mathcal{U}(0, \sigma^+)$ for $\rho = 0.05$ (blue), 0.1 (orange), 0.15 (yellow) and 0.2 (purple). Average capacities are chosen from $\mathcal{F}(-1,1.2)$.

3.6 Discussion

From the results we have found that the quenched averaging method does not provide enough power to increase the reliability of the network as it stands, the setup based on annealed averaging works well and the quenched averaging method with added resource (QAR) works better than both methods. A higher uncertainty results in higher reserve being given to nodes as does reducing the percentile value and the shortest paths. The excess power given to volatile nodes does however result in higher power flow over edges, resulting in more power being lost through power loss on edges and the necessity for loss in real time through the primary and secondary stages of distribution control (Section 1.3) for unused power reserves.

The distribution of fluctuating nodes may be different to the Gaussian distribution considered in the thesis; the main concern is what difference this makes. Most important for power grids is the confidence level being over 90% (literature suggests that a 90-99% confidence interval is necessary [161, 162]). The power given to a Gaussian wind generator in order to maintain a ρ confidence would be:

$$z_{Gauss} = \mu + \sigma\sqrt{2}\mathrm{erf}^{-1}(2\rho - 1), \tag{96}$$

and if the node had a Cauchy distribution this would be:

$$z_{Cauchy} = z_0 + b\tan[\pi(\rho - 0.5)], \tag{97}$$

where $0 \leq \rho \leq 1$.

Between $\rho = 0.9$ and $\rho = 1$ the Gaussian has a much smaller z value where two types of Gaussian have been matched with the Cauchy distribution as shown in Table 2 (Previously in [120] Cauchy distribution was suggested because of the leptokurtic nature of the distribution). Table 2 shows a Gaussian distribution 1 where $\mu = z_0$ and $\sigma = b$, and Gaussian distribution 2 where $\mu = z_0$ and the heights of both distributions are equal. We see from both Gaussian distributions that if a Cauchy distribution is a more appropriate fit of the error forecast at the tails, the z-value of Gaussian distributions would not be enough. This shows that we may be severely overestimating the amount of power the uncertain wind-farm type node can confidently share with the grid, this should be a consideration for any future work. However,

the estimations mentioned earlier have been focussing on closely matching the whole distribution of the wind-farm rather than the tails, and more research on which distribution best fits the tails of a wind-farm error forecast would be necessary. One additional flaw of both distributions is their infinite tails, which is unrepresentative of fluctuating sources which have a minimum and a maximum a production value.

Table 2. Cauchy distribution with $z_0 = 0.5$ anb $b = 0.05$ against Gaussian distribution 1 which has $\mu = 0.5$ and $\sigma = 0.05$ and Gaussian distribution 2 with $\mu = 0.5$ and $\sigma = 0.0625$.

		Confidence Interval			
		90%	95%	97.5%	99%
z-value	Cauchy	0.653	0.815	1.135	2.091
	Gaussian 1	0.564	0.582	0.597	0.616
	Gaussian 2	0.5801	0.602	0.622	0.645

Exploring this point further, as seen in Figure 3.10, there is also the possibility that when winds get too high the wind turbines are turned off to prevent them from breaking (this maximum capacity value is different depending on the capacity of the turbine). If the z-value was higher than the cut-out speed then the turbine generation capacity could be considered as zero but this does not discount the event of fluctuations exceeding capacity and the generator cutting out during the time-interval. This factor should also be considered in future work.

To solve the problem of power distribution in the presence of edge-bandwidth and fluctuations, an extra constraint would need to be added to the VFE calculation similar to that of Equation (62). To include strict bandwidth limits within a volatile grid using message passing algorithms would require bounding the modified power flow, $y + \varepsilon$ according to any expected fluctuations, as this cannot exceed bandwidth in either direction over 15-60 minute the time frame.

We have suggested increasing the reliability of networks with uncertain generators by increasing reserve power at the volatile nodes. Another approach to reducing the impact of renewable generators in a power network could be to consider a heterogeneous set of controllable demand, which is usually uncontrolled but could be designed to be more controllable. The ability to control demand would reduce the strain at peak times, and could be used to mitigate the fluctuations of renewable sources. For example, in the residential scenario,

around half of a household's power goes on heating and refrigeration. These processes are not specifically time dependant and often only work in stages[9]; they could be turned off or down for short periods of time to make demand more controllable according to uncontrolled generation. If for example each household's refrigeration and space heating was controllable and set to be off when demands are high or fluctuations are high, for a maximum of 12 minutes an hour. That means a group of 5 houses with a rolling power pause could reduce the groups residential load at peak flow by 10%. This could be implemented in some industries and residential areas; it requires housing cooperation and intelligent systems within smart grids to implement this.

Other considerations are that an increase in the number of microgenerators in the network will result in correlated fluctuations, and the ongoing developments in smart grids may see a point where consumers also have the capability to become generators when they produce more power than used. Electric vehicle charges could be controlled or turned off according to fluctuations or advanced battery storage may be the solution to mitigating fluctuating renewable generation altogether.

[9] Turning the heating on does not mean the heating is on 100% of the time, the water in a radiator is heated to a temperature, the heating systems are turned off while hot water remains in the radiator, and once the radiator temperature drops past a certain level, the heaters are turned back on again. This concept is similar in fridge freezers.

4 Minimising Load Shedding

4.1 Introduction

4.1.1 Background

At all times in a power grid, demand and consumption must balance; if there are large imbalances, this can lead to instability in the grid, large fluctuations in voltages and grid failure. Using operating systems close to their maximum capacity does increase the risk of failure, as the difference between generation and consumption becomes small, this is referred to as "tight" and action must be taken to avoid instability. Typically, electricity grids work close to capacity in order to minimise losses. Load shedding is the situation where power demand exceeds generation and it is impossible to satisfy all consumers. It is very rare for highly developed countries to experience this as demand is continuously monitored and generators are more reliable and have surplus capacity; they are much more common in developing countries. Eskom, the main supplier of South Africa electricity [163] explains how to prevent widespread blackouts or damage to equipment when generated power is unable to provide the power demanded. They list four main stages:

- *Stage 1* – Power utilities can make agreements with industrial consumers to turn down their equipment at the times when demand is high. Sometimes these are mandatory but usually they are agreed through reduced energy prices. This can only be done when demand exceeding generation has been predicted or scheduled as it takes a minimum of 2 hours to implement.
- *Stage 2* – The available generation is shared amongst consumers; this is called scheduled load shedding. It can result in low voltages and brownouts (where lights dim or flicker).
- *Stage 3* – Rolling load shedding is where electricity is sequentially cut in different regions, in order to reduce the impact on consumers. This is usually a predetermined plan that is implemented when necessary.
- *Stage 4* – In the rare case when load shedding is needed outside of the published schedules, emergency switching is used to protect the network. Blackouts can occur from power stations tripping, and can last from seconds to weeks, as rebooting the stations usually requires tapping into power sources from other countries.

If there is a blackout, some institutions have backup generators, such as hospitals, sewage treatment, telecommunication and mines but it is still preferable to protect critical needs through judicious distribution.

A study on load shedding for distributed stream processing looks at a distributed approach based on metadata aggregation and propagation [164]. Stream processing and power flow have many similar properties and considerations, including large flow (power flow and input streaming rates), bandwidths and the possibility of bottlenecks[10]. It aims to maintain a quality of service while managing large scale input rates which may drive the system into overload. They aim to combat this by dropping tuples (a data structure consisting of multiple parts) to reduce load. The stream processing system, uses an offline system to predetermine the best plan of load shedding in advance so that it can be quickly invoked if necessary; similar to how power grids plan stage 3 of load shedding. It suggests "Advanced Planning with FIT" (Feasible input table), where one generates a table of feasible combinations for a specific node, and propagates it to its parent node. The parent considers all the tables from its descendants it maps them from its outputs to its own inputs, merges the tables into one, removes the infeasible combinations and sends this new table to its parents. This continues until convergence, enabling it to shed load according to itself and its descendants, and the results are promising. This method closely follows that of discrete message passing; however, their work only currently considers tree-like structures. The algorithms' computational complexity appears to increase approximately linearly with the system size, and this demonstrates how other probabilistic propagation methods have already been suggested as promising methods for load shedding on large scale networks.

4.1.2 Problem

Although it is rare to experience blackouts in well-established electricity grids, the increased contribution of power from fluctuating renewable generation is likely to "tighten" the distance-to-failure gap, making power outages more likely. This chapter considers stage 2 – where consumers share the available generation – in an efficient way using message passing.

[10] A bottleneck is when the bandwidth of an edge is too small to fit enough resource through, potentially leaving some nodes unsatisfied.

4.2 Load Shedding

4.2.1 Method

Due to the fluctuating nature of renewable sources and their increasing contribution to the power grid the networks reliability is expected to decrease. Therefore, we need to consider the scenario of demand exceeding generation; this occurs when the sum of all node capacities in the network is negative, $\sum_{a\in N} \Lambda_a < 0$. In this event, the hard constraint (5) is unsatisfiable and so we add a variable ζ_j to each node j to represent each nodes deficit, which is minimised through the additional objective function, - $\frac{\zeta_j^2}{2}$ (This work has developed upon work done in [109] and [165]). The constraint can be rewritten as:

$$\sum_{k\neq i} \mathcal{A}_{jk} y_{jk} - y_{ij} + \Lambda_j + \Theta(-\Lambda_j)\Theta\left(-\sum_{a\in N} \Lambda_a\right)\zeta_j \geq 0, \tag{98}$$

where the first step function allows for only consumers to lose power (generators should not lose power in the event of load-shedding) and the second step function sums all the capacities in the network ensuring that the constraint is only softened by ζ_j if and only if overall demand exceeds overall generation. The new objective function is specific to node j and the related tree but not directly related to y_{jk} or messages from descendants, allowing it to be considered only once, outside of the sum over all descendants.

$$F_{ij} = \sum_{k\neq i} \mathcal{A}_{jk}\left[(A_{jk} + \phi'_{jk})\varepsilon_{jk} + \frac{1}{2}(B_{jk} + \phi''_{jk})\varepsilon_{jk}^2\right] + \alpha_j \frac{\zeta_j^2}{2}, \tag{99}$$

Where F_{ij} is the expanded VFE message to be passed to node i, similar to Equation (4), where α_j is a predetermined weighting variable which allows one to incorporate the importance of maintaining high levels of power at each consumer node into the calculations, the importance value. The first step function in Equation (98) can be replaced by giving generator nodes very large importance values to prevent generator-load-shedding; we will use this for simplicity in the results section.

Minimising with respect to deficit variable, ζ_j, gives:

$$\zeta_j = -\frac{\Theta(-\Lambda_j)\Theta(-\sum_{a\in N}\Lambda_a)\mu_{ij}}{\alpha_j}. \tag{100}$$

Plugging this into Equation (99) and minimising with respect to the Lagrange multiplier gives:

$$\mu_{ij} = \min\left[\frac{\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk} - \frac{A_{jk}+\phi'_{jk}}{B_{jk}+\phi''_{jk}}\right) - y_{ij} + \Lambda_j}{\sum_{k\neq i}\mathcal{A}_{jk}\left(B_{jk}+\phi''_{jk}\right)^{-1} + \frac{\Theta(-\Lambda_j)\Theta(-\sum_{a\in N}\Lambda_a)}{\alpha_j}}, 0\right], \tag{101}$$

leading to the messages to the ancestor:

$$A_{ij} \longleftarrow -\mu_{ij}, \tag{102}$$

$$B_{ij} \longleftarrow = \frac{\Theta(-\mu_{ij}-\epsilon)}{\sum_{k\neq i}\mathcal{A}_{jk}\left(B_{jk}+\phi''_{jk}\right)^{-1} + \frac{\Theta(-\Lambda_j)\Theta(-\sum_{a\in N}\Lambda_a)}{\alpha_j}}, \tag{103}$$

where step functions are dealt with using *if* loops within codes, and the unchanged backwards message to the descendants:

$$y_{jk} \leftarrow y_{jk} - \frac{A_{jk}+\phi'_{jk}+\mu_{ij}}{B_{jk}+\phi''_{jk}}. \tag{104}$$

With these equations, our corresponding programs will remain unchanged for all networks, unless the overall consumption is more than the overall production; the algorithm will then minimise the deficit according to each nodes importance level, alongside the cost of power loss.

It may be possible to modify these equations to say that if load shedding is necessary, the cost of power loss can be ignored. These would look like:

$$\begin{aligned} F_{ij} = \sum_{k \neq i} \mathcal{A}_{jk} \Bigg[& \left(A_{jk} + \Theta\left(\sum_{a \in N} \Lambda_a \right) \phi'_{jk} \right) \varepsilon_{jk} \\ & + \frac{1}{2} \left(B_{jk} + \Theta\left(\sum_{a \in N} \Lambda_a \right) \phi''_{jk} \right) \varepsilon_{jk}^2 \Bigg] + \alpha_j \frac{\zeta_j^2}{2} \\ & + \mu_{ij} \left(\sum_{k \neq i} \mathcal{A}_{jk} (y_{jk} + \varepsilon_{jk}) - y_{ij} + \Lambda_j \right. \\ & \left. + \Theta(-\Lambda_j) \Theta\left(- \sum_{a \in N} \Lambda_a \right) \zeta_j \right). \end{aligned} \tag{105}$$

However, with a high importance factor α in comparison to the distribution costs, these additions may be unnecessary. The difficulty in introducing these hard constraints is that if a descendant node passes the message $B_{jk} = 0$ (indicating it has spare resource) and if the recipient is a generator or a node not experiencing deficit, the denominators of the next messages to ancestors will be zero, resulting in unnecessary infinite messages.

4.2.2 Characteristics

To demonstrate the effects of the algorithm we first tested it on a simple synthetic 14-bus IEEE benchmark network [154] with two generator nodes of fixed capacities $\Lambda = 4$, with the remaining nodes defined as consumers of fixed capacities $\Lambda = -2$, resulting in an overall power deficit of -16 MW. Figure 4.1 shows pie charts where the colour of each slice represents the different nodes and their importance weights I (also denoted as α) are given on the circumference which are varied to show the effects of importance values. The pie chart represents the total amount of deficit within the network and how it is divided among the nodes. The algorithm does not include the step function which disallows generators from being negative, this is to better demonstrate the effects of the importance factor α. Any slices detached indicate a node which still has positive power (surplus at a generator). We see that when importance weights for all nodes are small (top left) the distribution cost part of the objective function becomes relatively more important and some power remains at the generators at the expense of extra deficit at consumers. The top right figure shows how when the importance of all nodes is large and equal the power deficit is equally divided and effects of minimising power loss are negligible. Bottom left is an example of a case where one node is less important than

others and the algorithm assigns the majority of the power deficit to it. Bottom right shows that alternatively, if one node is much more important that others the power deficit assigned to it is negligible. These graphs show the algorithms capability to prioritise nodes in a load shedding situation and to minimise the overall effects of load shedding, it highlights that all deficit needs to go somewhere, and how it can be distributed among nodes.

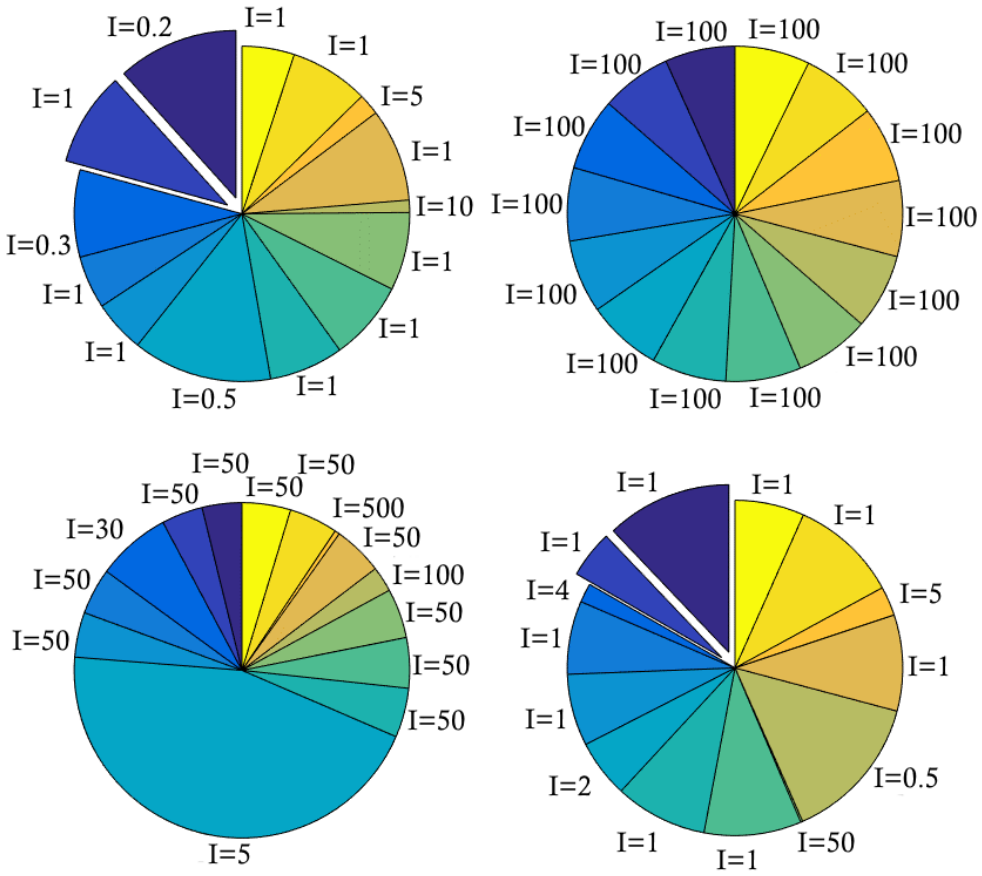


Figure 4.1 Pie charts with colours representing different nodes; I values represent the importance value α for each node. Examples are taken from a 14-Bus synthetic IEEE benchmark network. There are two generator nodes with a capacity $\Lambda = 4$, and the remaining consumer nodes have $\Lambda = -2$. Slices detached from the chart indicated power surplus at the node. (In this case generators were allowed to be negative, load shedding was not exclusive to consumers) [111]. , DOI: 10.1016/j.egypro.2016.12.139.

4.3 Load Shedding with Fluctuations

4.3.1 Method

Load shedding can also incorporate the quenched averaging with soft constraints (QAR) calculations, to consider both load shedding and fluctuations. This gives a VFE of:

$$\begin{aligned}\left\langle F_V(y_{ij}|\mathbf{T}_j)\right\rangle_{\Lambda_j} = \int P(\Lambda_j)\mathrm{d}\Lambda_j \min\Bigg[&\sum_{k\neq i}\mathcal{A}_{jk}\left[F_V\left(y_{jk}\middle|\mathbf{T}_k\right)+\phi(y_{jk})\right]\\&+\mu_{ij}\Bigg[\sum_{k\neq i}\mathcal{A}_{jk}y_{jk}-y_{ij}+\Lambda_j+\sqrt{2\sigma_j^2}\,\mathrm{erf}^{-1}(2\rho_j-1)\\&+\Theta\left(-\sum_{a\in N}\Lambda_a\right)\zeta_j\Bigg]+\alpha_j\frac{\zeta_j^2}{2}\Bigg].\end{aligned} \tag{106}$$

Minimising with respect with ζ_j allows us to find:

$$\zeta_j = -\frac{\Theta(-\sum_{a\in N}\Lambda_a)\mu_{ij}}{\alpha_j}, \tag{107}$$

and take the average of the VFE:

$$\langle F_{ij}^*\rangle_{\Lambda_j} = \frac{1}{2}\sum_{k\neq i}\mathcal{A}_{jk}\,\frac{\langle\mu_{ij}^2\rangle_{\Lambda_j}-\left(A_{jk}^{<>}+\phi'_{jk}\right)^2}{B_{jk}^{<>}+\phi''_{jk}}+\frac{1}{2}\left\langle\frac{\Theta(-\sum_{a\in N}\Lambda_a)\mu_{ij}^2}{\alpha_j^2}\right\rangle_{\Lambda_{j\cdot}} \tag{108}$$

Calculations remain similar to those previously; μ_{ij}^2, before finding the average, gives:

$$\mu_{ij}^2 = 0 \text{ or } \frac{\Lambda_j^2-2x\Lambda_j+x^2}{\left(\sum_{k\neq i}\frac{1}{B_{jk}^{<>}+\phi''_{jk}}+\frac{\Theta(-\sum_{a\in N}\Lambda_a)}{\alpha_j}\right)^2}, \tag{109}$$

where $x \equiv y_{ij}-\sqrt{2\sigma_j^2}\,\mathrm{erf}^{-1}(2\rho_j-1)-\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}-\frac{A_{jk}^{<>}+\phi'_{jk}}{B_{jk}^{<>}+\phi''_{jk}}\right)$. If we rewrite $\Theta(-\sum_{a\in N}\Lambda_a)$ as $\Theta\left(-\sum_{\substack{b\neq j\\ b\in N}}\Lambda_b-\Lambda_j\right)$, where b indicates all nodes in the network apart from j, we can assume all other node capacities to be deterministic at that point, but still considering

the Gaussian probability distribution Λ_j at node j. We now require the average of the square of μ_{ij}:

$$\langle \mu_{ij}^2 \rangle_{\Lambda_j} = \begin{cases} 0, & \text{if } \Lambda_j - x \leq 0, \\ \dfrac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2} \displaystyle\int \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j, & \text{if } -\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b - \Lambda_j < 0, \\ \dfrac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2 + \dfrac{1}{\alpha_j}} \displaystyle\int \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j, & \text{if } -\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b - \Lambda_j \geq 0. \end{cases} \tag{110}$$

As Λ_j is a probability distribution, we can write this as an integral, Appendix D shows how $\langle \mu_{ij}^2 \rangle$ varies, depending on whether x or $-\sum_{b \neq j} \Lambda_b$ is larger. The derivatives then become:

$$A_{ij}^{<>} \leftarrow -\langle \mu_{ij} \rangle_{\Lambda_j} = \begin{cases} \dfrac{1}{2} \dfrac{\left(\operatorname{erf}\left(\dfrac{x - \bar{\Lambda}_j}{\sqrt{2}\sigma_j} \right) + 1 \right)(x - \bar{\Lambda}_j) + \dfrac{2\sigma_j^2}{\sqrt{2\pi\sigma_j^2}} e^{-\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2}\sigma_j^2} \right)^2}}{\Theta(-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j)}}{\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk}^{<>} + \phi''_{jk}} + \dfrac{}{\alpha_j}}, & \text{if } x \leq -\sum_{\substack{b \neq j \\ b \in N}} \bar{\Lambda}_b, \\ \dfrac{1}{2} \dfrac{\left(\operatorname{erf}\left(\dfrac{-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j}{\sqrt{2}\sigma_j} \right) + 1 \right)(x - \bar{\Lambda}_j) + \dfrac{2\sigma_j^2}{\sqrt{2\pi\sigma_j^2}} e^{-\left(\frac{-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j}{\sqrt{2}\sigma_j^2} \right)^2}}{\Theta\left(-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j \right)}}{\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk}^{<>} + \phi''_{jk}} + \dfrac{}{\alpha_j}} + \\ \dfrac{1}{2} \dfrac{\left(\operatorname{erf}\left(\dfrac{x - \bar{\Lambda}_j}{\sqrt{2}\sigma_j} \right) - \operatorname{erf}\left(\dfrac{-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j}{\sqrt{2}\sigma_j} \right) \right)(x - \bar{\Lambda}_j) + \dfrac{2\sigma_j^2}{\sqrt{2\pi\sigma_j^2}} \left(e^{-\left(\frac{x - \bar{\Lambda}_j}{\sqrt{2}\sigma_j^2} \right)^2} - e^{-\left(\frac{-\sum_{b \neq j} \bar{\Lambda}_b - \bar{\Lambda}_j}{\sqrt{2}\sigma_j^2} \right)^2} \right)}{\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk}^{<>} + \phi''_{jk}}}, & \text{if } x > -\sum_{\substack{b \neq j \\ b \in N}} \bar{\Lambda}_b, \end{cases} \tag{111}$$

$$B_{ij}^{<>} \leftarrow \begin{cases} \frac{1}{2}\frac{\operatorname{erfc}\left(\frac{x-\bar{\Lambda}_j}{\sqrt{2}\sigma_j}\right)+1}{\sum_{k\neq i}\mathcal{A}_{jk}\ \frac{1}{B_{jk}^{<>}+\phi''_{jk}}+\frac{\Theta(-\sum_{\substack{b\neq j\\ b\in N}}\bar{\Lambda}_b-\bar{\Lambda}_j)}{\alpha_j}}, & \text{if } x \leq -\sum\limits_{\substack{b\neq j\\ b\in N}}\bar{\Lambda}_{b,} \\ \frac{1}{2}\frac{\operatorname{erf}\left(\frac{x-\bar{\Lambda}_j}{\sqrt{2}\sigma_j}\right)-\operatorname{erf}\left(\frac{-\sum_{\substack{b\neq j\\ b\in N}}\bar{\Lambda}_b-\bar{\Lambda}_j}{\sqrt{2}\sigma_j}\right)}{\sum_{k\neq i}\mathcal{A}_{jk}\ \frac{1}{B_{jk}^{<>}+\phi''_{jk}}+\frac{\Theta(-\sum_{\substack{b\neq j\\ b\in N}}\bar{\Lambda}_b-\bar{\Lambda}_j)}{\alpha_j}}, & \text{if } x > -\sum\limits_{\substack{b\neq j\\ b\in N}}\bar{\Lambda}_{b,} \end{cases} \tag{112}$$

and,

$$y_{jk} \leftarrow y_{jk} - \frac{A_{jk}+\phi'_{jk}+\langle\mu_{ij}\rangle_{\Lambda_j}}{B_{jk}+\phi''_{jk}}. \tag{113}$$

4.4 Results

As the average importance value increases, we can see from the histograms in Figure 4.2 how the mean proportion of power sent from each generator increases around some normal distribution. This can be seen clearly when each histogram is fitted with a normal distribution (Figure 4.3) and the distributions are compared. The mean and standard deviation increases for higher importance costs, and this increase in sent power is due to the importance cost superseding the distribution cost, as more power is rearranged within the network to achieve a low load shedding cost.

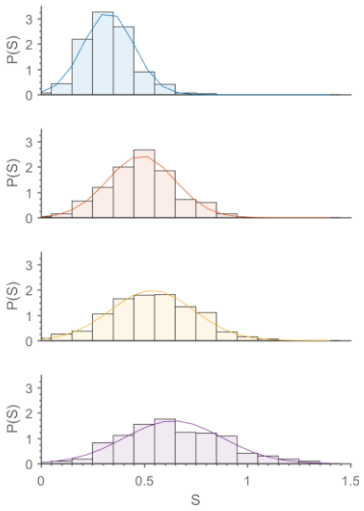


Figure 4.2 The histograms and their best fit normal distribution for the probability distribution of the power sent, S, from all generators. As explained in Figure 4.3.

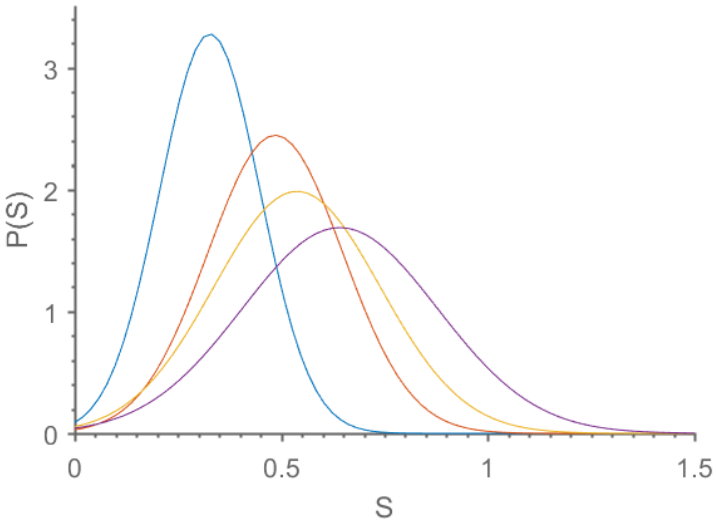


Figure 4.3 The normal distributions of power sent, S, from generators as the fixed importance value increases from 0.25 (blue), 0.5 (orange), 0.75 (yellow) to 1 (purple), tested on multiple 100-node networks which have a random regular connectivity of 6 and with randomly chosen capacities from the distribution $\Lambda = \mathcal{U}(-1,0.5)$ where half of nodes are consumers, and half generators.

Next we evaluate how the importance value and the distribution of importance values affect the nodes. It can be seen from Figure 4.4 how increasing the fixed importance value results in nodes having a smaller, less negative, distribution of deficit with smaller variance. This is clear in Figure 4.5 when the importance values are chosen randomly from the distribution $\mathcal{U}(0, \alpha^+)$, where α^+ is increasing from 1 to 4, that the deficit of the nodes appears to stay to the left of a negative hyperbolic function highlighting how higher importance nodes are prioritised. The values in Figure 4.6 are randomly chosen from the distribution $\mathcal{U}(\alpha^-, \alpha^+)$ where α^+ is the same as in Figure 4.5, but where the difference between α^+ and α^- is always 1, the results show how

the function always maintains a positive correlation between α_j and r_j, but when the average importance is smaller the distribution of deficit is much larger and has a more obvious curvature. The much larger distribution in deficit for small importance values is due to the distribution cost contribution to the objective function; with the smaller values of α, the distribution cost is more heavily weighted and so less power is sent to minimise deficit. Also for all networks, the nodes with higher importance are prioritised and are provided with more power to reduce their deficit to minimise the objective function.

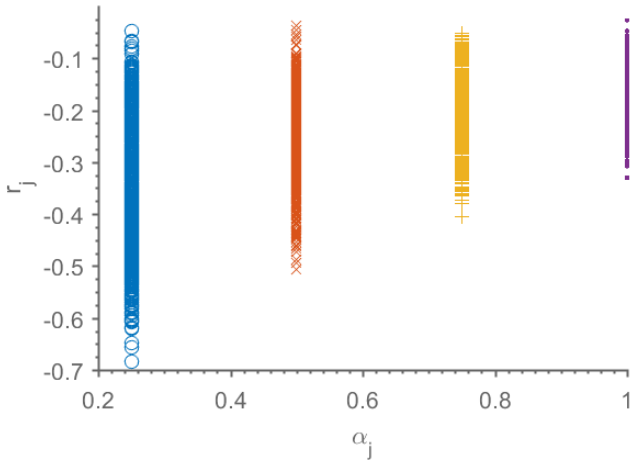


Figure 4.4 Scatter diagram plotting each individual node's importance value and deficit where the importance values were chosen from the distribution $\mathcal{F}(\alpha^F)$, where $\alpha^F = 0.25$ (blue ○), 0.5 (orange ×), 0.75 (yellow +) and 1 (purple •).

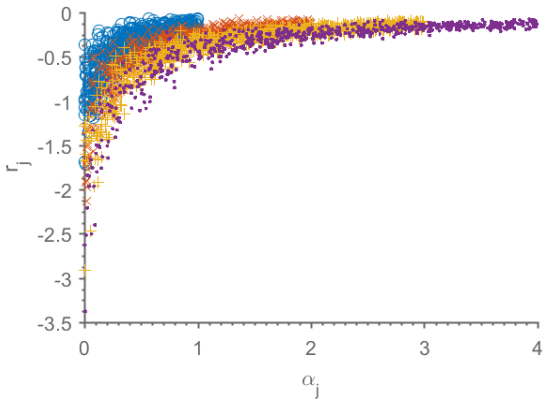


Figure 4.5 Scatter diagram plotting each individual node's importance value and deficit where the importance values were chosen from the distribution $\mathcal{U}(0, \alpha^+)$, where $\alpha^+ = 1$ (blue ○), 2 (orange ×), 3 (yellow +) and 4 (purple •).

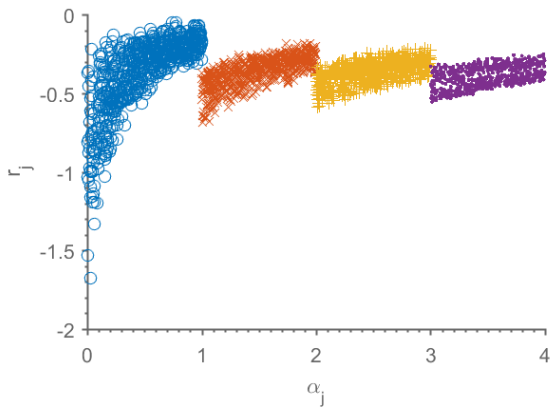


Figure 4.6 Scatter diagram plotting each individual node's importance value and deficit where the importance values were chosen from the distribution $\mathcal{U}(\alpha^-, \alpha^+)$, where $\alpha^+ = 1$ (blue ○), 2 (orange ×), 3 (yellow +) and 4 (purple •), and $\alpha^- = \alpha^+ - 1$.

It is seen in Figure 4.7 how the deficit of the nodes changes depending on a correlation between the importance of a node, and the original capacity of the node. Importance values are assigned using a negative correlation; $\alpha = \mathcal{C}(-1, \Lambda)$[11] in blue, and a positive correlation; $\alpha = \mathcal{C}(+1, \Lambda)$ in orange, with respect to each consumers' absolute capacity value. Illustrating in blue circles how when small absolute capacities have a large α, even before the algorithm is ran, the shape of the capacities will be similar, and so distribution cost will be low to achieve a minimised load shedding curve. However, when large α values are on nodes with large absolute capacities, to reduce the deficit on the very low capacities will require a lot more power flow and so the distribution cost will carry a much higher weight; resulting in a more uniform final deficit among nodes. These figures demonstrate how important nodes are prioritised and how significant the accurate allocation of importance values is.

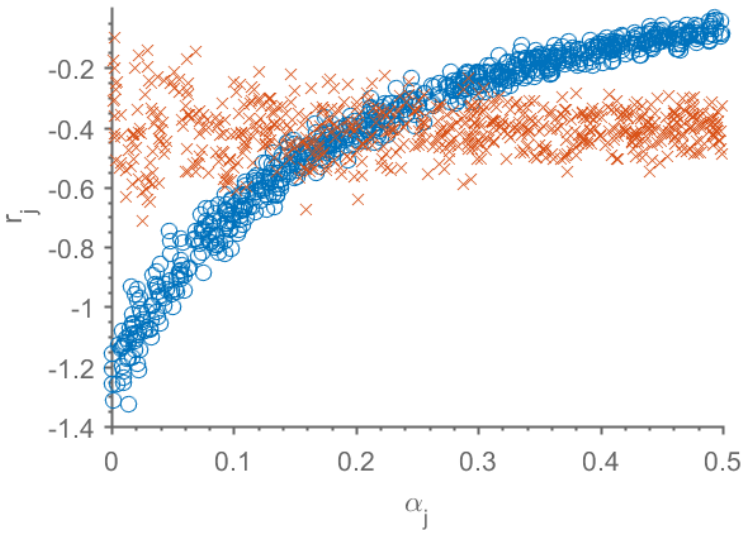


Figure 4.7 Scatter diagram plotting each individual node's importance value and deficit where the importance values were chosen from the distribution $\mathcal{C}(-1, \Lambda)$ (blue ○) and $\mathcal{C}(+1, \Lambda)$ (orange ×) and the correlation is with respect to the absolute capacity of each consumer.

[11] $\mathcal{C}(a, b)$ implies a negative correlation with respect to the distribution b if $a = -1$, and a positive correlation if $a = +1$.

Figure 4.8 shows an example with fluctuations which is similar to the deterministic case in Figure 4.6, where now $\sigma = 0.5$ for all nodes. In the deterministic case we see how all nodes result in a value of deficit (negative reserve), lower deficit for higher importance nodes, whereas in the volatile case some nodes with higher importance values actually receive positive reserve values at the cost of less important nodes becoming more negative. For higher overall importance weights the network appears to return to a similar pattern where all nodes have deficit. The figure demonstrates how the algorithm considers the importance weights together with volatile nodes, allowing more important volatile nodes to receive reserve, with the understand that they may use more or provide less than expected, but will still have a low deficit in comparison to less important nodes.

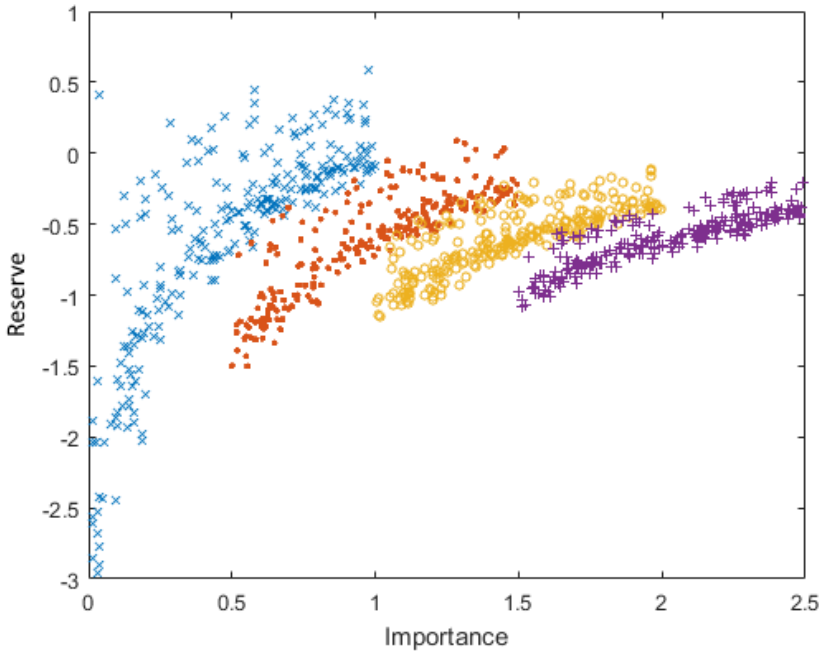


Figure 4.8 Scatter graph showing examples of importance values selected randomly from a distribution $\mathcal{U}(\alpha^-, \alpha^+)$ where $\alpha^+ = 1$ (blue ×), 1.5 (orange •), 2 (yellow ○) and 2.5 (purple +), and $\alpha^- = \alpha^+ - 1$. Each symbol represents a node in a 100-node random regular connected network of $c = 6$, plotted according to its power deficit (or negative reserve) and the importance weight it was given. One fifth of nodes are given mean capacity $\bar{\Lambda} = 1$ and the remaining consumers have $\bar{\Lambda} = -1$, all variances are fixed at $\sigma = 0.5$ [111]. DOI: 10.1016/j.egypro.2016.12.139.

It can be shown how load shedding affects the nodes according to their capacities. In Figure 4.9 and Figure 4.10 we see the deficit left at the nodes according to their initial capacity. We see how most generators have $r = 0$ as they have been depleted, and nodes starting with a lower capacity end with a lower deficit; the positive correlation between capacity and deficit decreases as the importance value increases due to the distribution cost being gradually ignored. An example of distribution cost being favoured and allowing some generators to maintain some of their reserve power, can be seen marked in blue in Figure 4.9. In Figure 4.10 examples where the importance values were chosen either positively (orange ×) or negatively (blue ○) correlated with their absolute capacities are shown. With negative correlation, we see a steeper increase in residual deficit for consumers with lower capacities as they have a low importance value and associated distribution cost required to provide power is low. For the positive correlation, the spread of the deficits appears almost unaffected by the value of Λ since the greater the demand the greater the importance and so lower capacities require a high level of power flow which increases the distribution cost and, due to weighting, restricts power flow before a fully minimised load shedding can be achieved. The correlation of importance and capacities is a major factor to consider because in real networks it may be found that larger consumers require a higher importance (similar to the orange curve), in this case we see quite an even distribution of deficit among nodes, which may even be smaller at lower capacities for larger importance values.

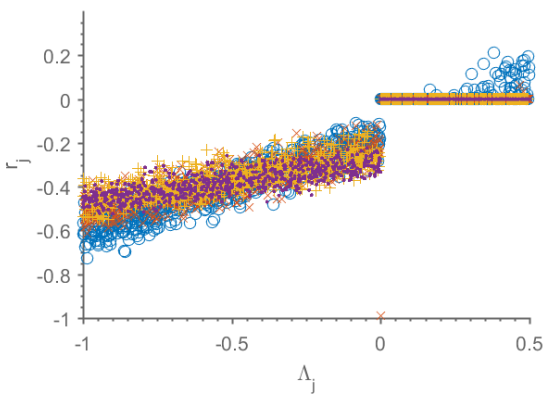


Figure 4.9 The deficit remaining at nodes according to their own capacity. From a 100-node network, randomly regularly connected by 6 nodes each, and with capacities randomly chosen from the distribution $\mathcal{U}(0.5, -1)$. Importance values were randomly chosen from a distribution $\mathcal{F}(\alpha^F)$, where $\alpha^F = 0.25$ (blue ○), 0.5 (orange ×), 0.75 (yellow +) and 1 (purple •).

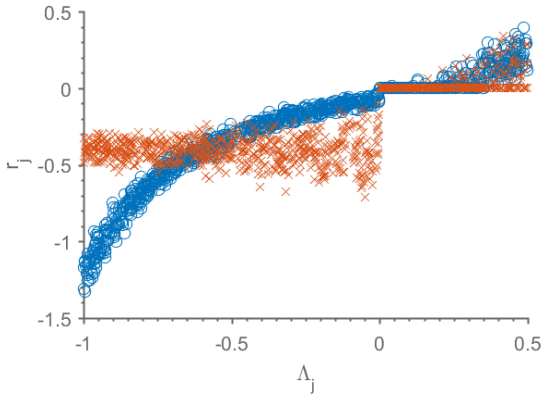


Figure 4.10 The deficit remaining at nodes according to their own capacity. Conditions are the same as in Figure 4.9 but importance values were randomly chosen from a distribution $\mathcal{C}(-1$ (blue ○) or 1 (orange ×), Λ).

Figure 4.11 shows histograms of the power distribution supplied over the edges once a solution is found. It can be seen, that when importance is zero, minimising power loss is the most important objective and no power is distributed allowing all consumers to remain their original negative capacity. As the importance increases the power distribution histogram tends towards a bi-modal distribution with means at 0.043 and 0.33; representing the power sent to minimise the deficit of neighbouring nodes. The peaks presumably characterise more and less central edges with high and low power flow, respectively. They become less emphasised and more spread out as resources primarily aim to reduce load shedding irrespective of power loss, equalising the deficit. Each graph also includes the average power loss per solution (PL), which increases as importance values increase, while the corresponding network load shedding cost (LS) decreases. When there is inevitable failure in real power networks due to power deficit, reducing the deficit at the most important nodes becomes a priority over distribution costs. For higher values of α the distribution remains the same because the distribution cost has already became negligible.

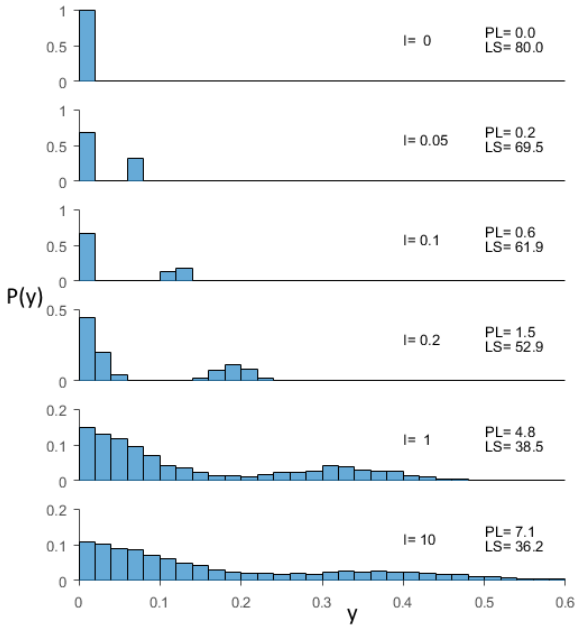


Figure 4.11 The average, normalised distribution of the power values at each edge after convergence. The importance given to each node in the 100-node network, is indicated above each of the histograms and capacities are $\Lambda = \mathcal{F}(-1, 1)$ where 80% are consumers and 20% are generators [111]. © Creative Commons Licence, DOI: 10.1016/j.egypro.2016.12.139.

Adjusting the topology of the network so that the random regular fixed connectivity increases from $c = 4$ to 11, in Figure 4.12 we can see that a high connectivity decreases the load shedding cost, where load shedding costs are calculated by $\frac{\alpha_j r_j^2}{2}$. We also observe in Figure 4.13 that the distribution costs are decreased for increasing connectivity too, and a smaller importance value decreases the distribution costs further because the distribution cost is prioritised over importance cost, which can be seen in Figure 4.12 where smaller α gives a higher importance cost at high connectivities. The figures show the advantages of a high connectivity and that at low connectivity it appears that lower importance values result in lower load shedding costs; this may be due to the α value being included within the calculations.

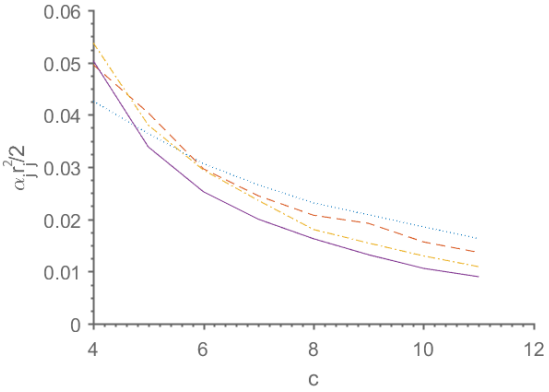


Figure 4.12 The importance cost of the network as the random regular connectivity is increased from 4 to 11 on ten networks with 100-nodes given capacities from the distribution $\mathcal{U}(-1,0.5)$. Blue indicates fixed importance value of 0.25, orange is 0.5, yellow 0.75 and purple 1.

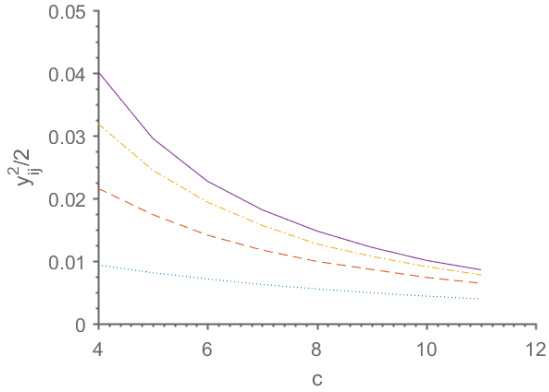


Figure 4.13 The distribution cost of the network as the random regular connectivity is increased from 4 to 11 on ten networks with 100-nodes given capacities from the distribution $\mathcal{U}(-1,0.5)$. Blue indicates fixed importance value of 0.25, orange is 0.5, yellow is 0.75 and purple is 1.

Looking at how the scale of capacity deficit in the network effects the optimisation with respect to importance (Figure 4.14[12]), we see that both distribution and average of the deficit increase (less power at consumer nodes) as the network becomes poorer in resource. It can also be shown how at the point where overall excess is 0 (in this case where half the nodes have $\Lambda^- = -1$ and the other half have $\Lambda^+ = 1$), the deficit jumps to 0; this is where load shedding is no longer necessary, the function $\Theta(-\sum_{a\in N}\Lambda_a)$ becomes 0 and the variable ζ_j vanishes. Figure 4.15 shows a bounded line graph where the mean is represented with the solid lines and the shaded portion is from the minimum value of power sent, to the maximum. The graph shows that the importance value makes a large difference to the size of the jump from load shedding to zero

[12] Figure 4.14 to Figure 4.17 do allow generators to have deficit.

deficit, where smaller importance values result in a much larger jump; this is due to the influence of the distribution costs (which in Figure 4.16 can be seen to increase at the point where load shedding is no longer required). Interestingly in Figure 4.17 we see that when the importance values are high, the average and maximum amount of power sent is higher than for lower importance values, but that the minimum values of the power provided do not appear to be affected by either the wealth of the network or the importance of nodes. We suspect that this is because of generators that are far from consumers and are (distribution) costly to send high values of power.

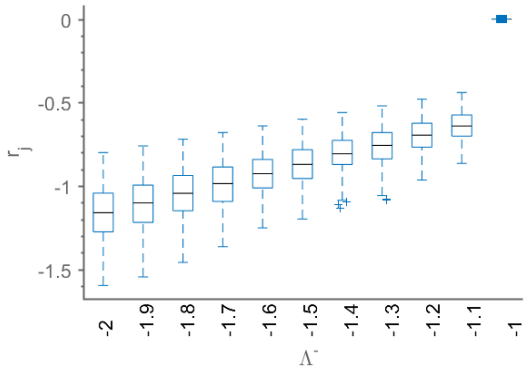


Figure 4.14 A box plot of the distribution of deficit, r, given to consumers as the lower value Λ^- increases, when capacities are chosen from the fixed distribution $\Lambda = \mathcal{F}(\Lambda^-, 1)$. Plotted from ten 100-node networks with a random regular connectivity of 6, and each consumer node has a fixed importance value of 0.25.

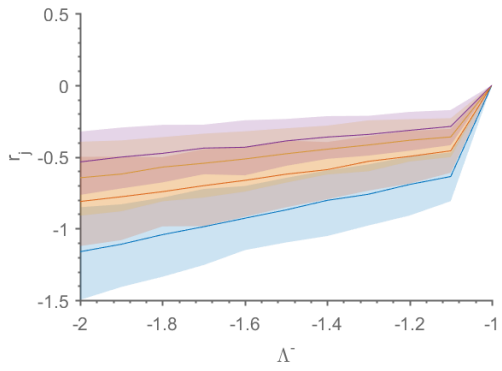


Figure 4.15 A borderline plot of the distribution of deficit, r, given to consumers as the lower value Λ^- increases, when capacities are chosen from the fixed distribution $\Lambda = \mathcal{F}(\Lambda^-, 1)$. Plotted from ten 100-node networks with a random regular connectivity of 6, and each consumer node has a fixed importance value of 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

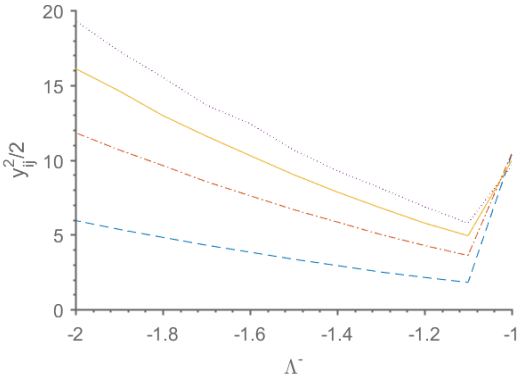


Figure 4.16 The average distribution cost, as the lower value Λ^- increases, when capacities are chosen from the fixed distribution $\Lambda = \mathcal{F}(\Lambda^-, 1)$. Plotted from experiments on ten 100-node networks with a random regular connectivity of 6, and each consumer node has a fixed importance value of 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

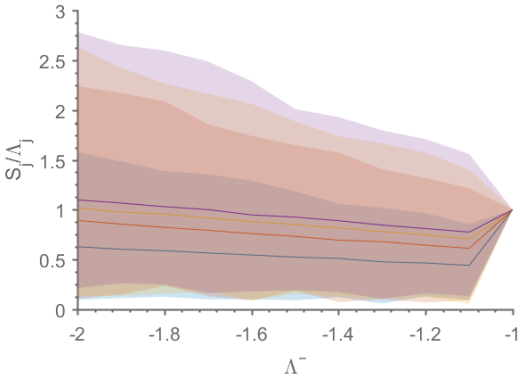


Figure 4.17 A borderline plot of the distribution of power provided by generators, as the lower value Λ^- increases, when capacities are chosen from the fixed distribution $\Lambda = \mathcal{F}(\Lambda^-, 1)$. Plotted from experiments on ten 100-node networks with a random regular connectivity of 6, and each consumer node has a fixed importance value of 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

Increasing the system size N does not appear to change the distribution of deficit at the nodes (Figure 4.18), although importance values still make a difference on its variance. From Figure 4.19 we can see how the power provided from all nodes does not appear to change much as N increases. But the maximum power increases, and this can be seen more clearly in Figure 4.20, where the variance of the power provided from just generators does increase with the system size. This may be due to the graph becoming sparser (because connectivity is fixed and not increasing proportionally with N) this decrease in the proportion of connectivity in terms of N, increases the distribution cost and requires closer generators to give more, and further away generators to give less, increasing the variance. This load shedding algorithm allows consumers to become more negative if necessary to reduce the deficit at other nodes. Further work could look to avoid power flow from customer nodes; as in real-life (excluding micro-grids) a consumer cannot give away power and increase its deficit.

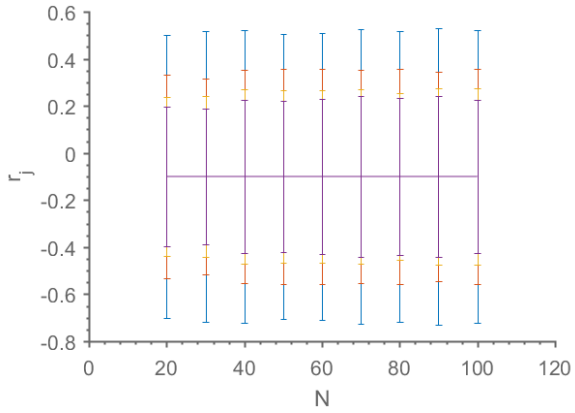


Figure 4.18 The distribution of deficit given to each node as N increases, where importance values are fixed at 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

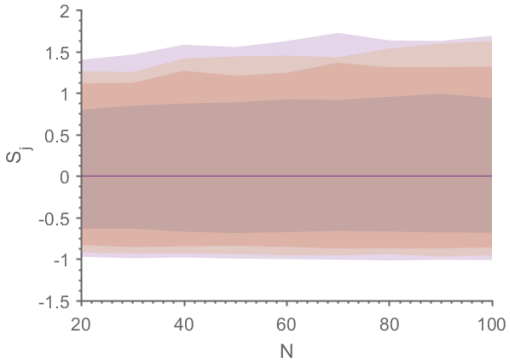


Figure 4.19 The distribution of power sent, S, from all nodes as N increases, where importance values are fixed at 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple). (Consumers receiving power is demonstrated with a negative S value.

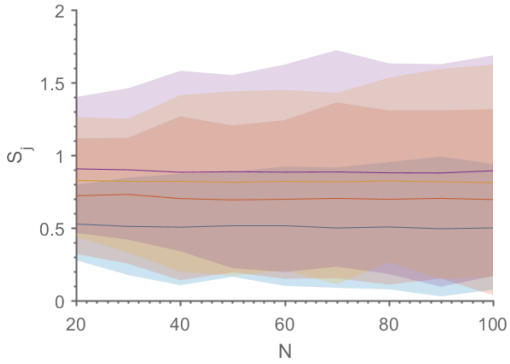


Figure 4.20 The distribution of power provided from just generators as N increases, where importance values are fixed at 0.25 (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

When increasing the importance values given to nodes, the distribution of the values may make a difference. To start with we look at fixed importance values and see in Figure 4.21 a hyperbolic-like increase of the deficit given to nodes as the importance value increases. In Figure 4.22 where generators are included it can also be seen that positive power is also present for smaller importance values due to the minimisation of distribution cost, demonstrating the importance of properly chosen importance values. We can see from Figure 4.23 and Figure 4.24 how increasing the importance values results in an increase in distribution costs and a decrease in load shedding costs, and that a lower connectivity gives higher costs for both. We assume a higher load shedding cost at low connectivity because the distribution cost is larger and so restricts the power flow from adjusting deficit.

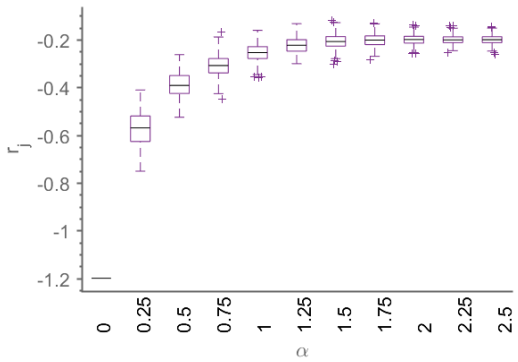


Figure 4.21 The distribution of consumer deficits on multiple 100-node network when the capacities are randomly chosen from the distribution $\mathcal{F}(-1.2,1)$ on a random regular network with connectivity 6 and increasing fixed importance values from 0 to 10.

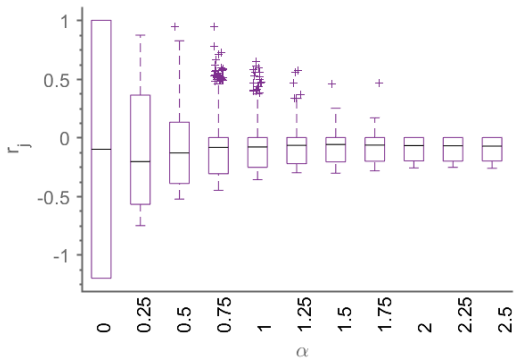


Figure 4.22 The distribution of deficits on all nodes, including generators, as importance values increase. Conditions as in Figure 4.21.

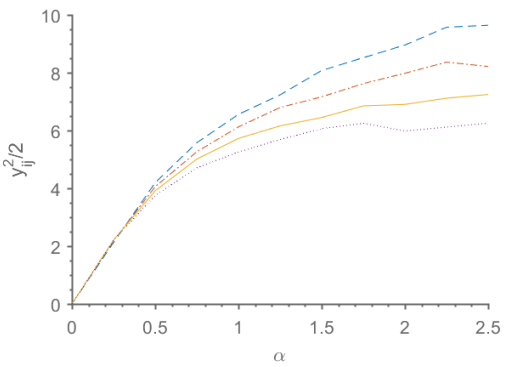


Figure 4.23 The increase in distribution costs as importance increases, with coloured lines indicating a random regular connectivity of 6 (blue), 7 (orange), 8 (yellow) and 9 (purple).

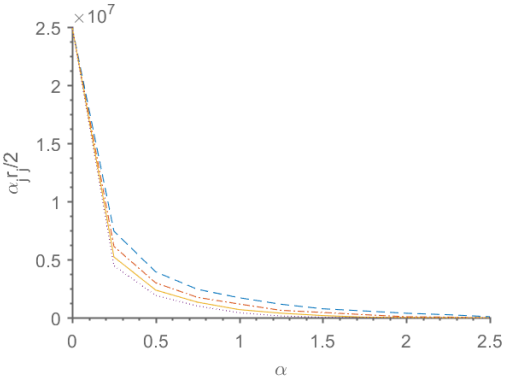


Figure 4.24 The decrease in importance costs with the increase in the importance value, with coloured lines indicating a random regular connectivity of 6 (blue), 7 (orange), 8 (yellow) and 9 (purple).

A similar figure demonstrates a box plot of the currents, y, increasing logarithmically as the average α values are increased (Figure 4.25), these results demonstrate the values of α at which load shedding becomes a higher priority than distribution cost; in this case we could suggest that by $\alpha = 1$ load shedding has a higher priority objective function, however this would vary for different sizes, topologies, connectivities, wealth and demand of a network. On a network with 40-nodes, Figure 4.26 shows the change in distribution of power along edges, y, as the standard deviation of the importance value is increased, where the importance value given to each node is drawn from the normalised Gaussian distribution: $\mathcal{N}(\alpha^m, \alpha^s)$. We see that as the importance values become more distributed among nodes the higher the average value of y is; the variance of the power over edges is also increased. The high α^m values do not vary much from each other, but small α^m (blue) does have a smaller average y, as expected.

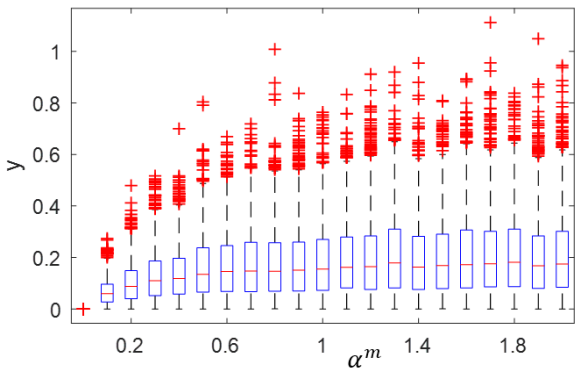


Figure 4.25 The distribution of power flow over edges y as the uniform importance value increases for $N = 40$ networks [158].

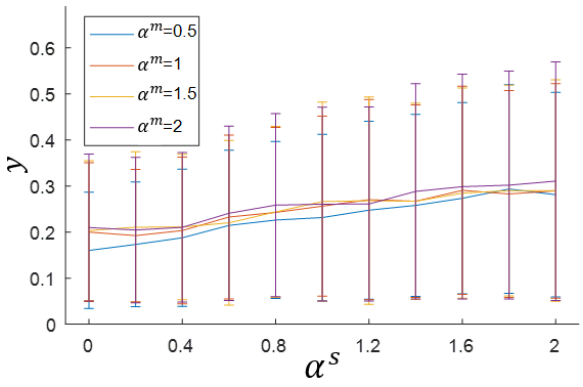


Figure 4.26 The distribution of power flow over edges y for the increase in the standard deviation of importance for mean importance of $\alpha^m = 0.5, 1, 1.5$ and 2. The horizontal line describes the mean of the of y, and the vertical lines describe the standard deviation [158].

Figure 4.27 gives the average power through edges for networks with different average capacities. The capacities were drawn from a normal distribution of unit variance $\mathcal{N}(\Lambda^m, 1)$. Stars show results from 40-node networks and circles indicate 100-node networks (colour is meaningless). The average power over an edge appears to decrease rapidly as Λ increases past 0, this is due to a reduced demand at nodes. Below 0, but higher than -1 the average y appears to remain high, this is to provide power where possible from producers. But when $\Lambda < -1$, the majority of nodes will be consumers and so any passed power will be purely to rearrange deficit, this is where changes in Λ^m appear to have no correlation with the average power flow, this is an unrealistic situation however as it illustrates a network with no generators.

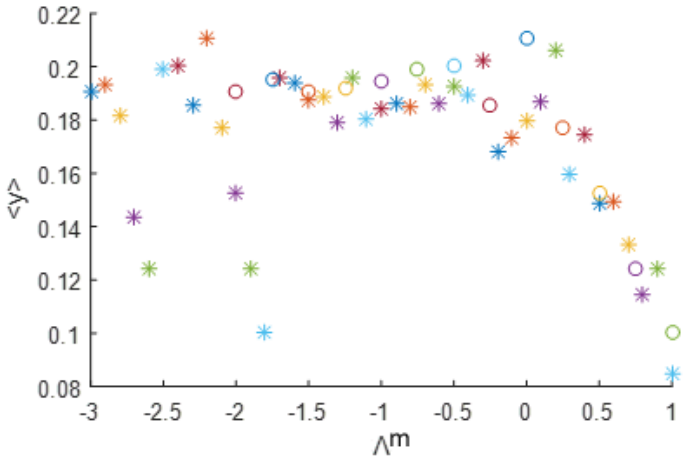


Figure 4.27 The effect of the average capacity Λ^m on the average value of y (capacity was drawn randomly from a Gaussian distribution with mean Λ^m and standard deviation of 1). Stars $*$ represent results from 40-node networks, ○'s represent results from 100-node networks [158].

4.5 Discussion

We presented a power flow optimisation method based on message passing, focussing on how to effectively manage an event of inevitable failure due to excess demand.

The formulation addresses the minimisation of load shedding for reducing the impact on consumers; specifically, *only* in the case when demand exceeds generation.

We have shown how the weighting of the nodes has a large impact on the remaining deficit at each node; a higher importance value reduces the deficit at the corresponding node and a very high importance value will reduce the deficit further to negligible levels. However, increasing the importance value of all nodes is not advantageous because the deficit must be spread evenly

somehow so it is distributed between them; although high overall importance values does allow the network to ignore distribution cost due to the need to satisfy their demands at all (distribution) costs. If the importance values are distributed among the nodes, the respective distribution of deficit at those nodes appears to follow a rectangular hyperbolic-like function, always favouring the more important nodes. We saw that correlation between importance values and starting capacities resulted in very different curves of power deficit, and this should be considered when thinking of the security of the network.

A discontinuous point was seen when the network changes from being satisfiable to unsatisfiable, needing to shed load, this jump reduces the power sent and therefore the pressure on the edges in the network and should be seriously considered for real networks as the change may either result in a relieved, safer network due to reduced pressure, or the sudden change over edges and at consumers and generators may be problematic (softer constraints or load shedding with fluctuations may mitigate this problem).

A problem observed was that although power was being provided to minimise load shedding costs, sometimes the power was sent from consumers. In a real network without micro-grids, this would not be possible and should be considered in future research. Additionally, we see that when importance values are low, that some generators may retain power; this is due to the balancing between minimising load shedding and distribution costs. This should not happen unless a fixed minimum value (P_{min}, which is explained in Section 5) has been placed on the generators and so further work should introduce a change of constraint that requires generators to have zero excess power (or P_{min}) during load shedding scenarios, and limit consumers to ensure they cannot provide power to other consumers through an additional constraint. Without a constraint on generators requiring them to send all their power, and as the scenario where load shedding is necessary is much more serious than the need to minimise distribution costs, we suggest that importance values should always have a large value (unless agreements with a consumer have been made, and cost is predominant) as very small importance values result in a network where necessary power does not flow through edges.

Results show that an overall increase in importance weight increases the average power flow along edges and the distribution cost, but minimised the load shedding costs and the distribution of the deficit over nodes. We also see how importance weighting on volatile networks can results in positive reserve at high importance nodes to minimise possible deficit during fluctuations, resulting in less important nodes with more negative reserves to compensate for this.

To develop this algorithm further, one could set minimum reserve values at the generators to protect them from shutdown in case of excessive demand; as shut down increases the time for the network to be up and running again. This could be done using the P_{min} value as mentioned in the next chapter. Additionally, preventing the consumer nodes from having a lower capacity than their used power would increase the practicality of the algorithm.

Future work on minimising load shedding costs could include turning the access of consumers off one at a time to accurately represent real/rolling blackouts as described in stages 3 and 4 at the beginning of the chapter. Also, an existing method to reduce the risk of load shedding necessity is the $N-1$ contingency. This is where a power grid always produces an extra amount of power that could be used if any generator or power line was disconnected from the system, this means that the excess contingency reserve would be equivalent of the largest generator capacity. We could adapt our approach to consider a contingency practise such as this, but it may be advantageous to also include the probability of each failure. Further work could also prepare for $N-2$ contingencies, which are also suggested within OPF; these could utilise joint probabilities to include the probability of a pair of generators or lines tripping at the same time (the probability of targeted attacks taking two of the largest generators, or the probability of a local disaster where two generators within close proximity are affected).

5 Minimising Generation Costs

5.1 Introduction

Although minimising the cost of power losses is one of the main objective functions of a power flow program, a more important cost to be minimised is the generation cost. This chapter will focus on ways of minimising the economic cost of power production at different power generating units. Each power station has a different cost of production as described in Table 3:

Table 3. Report for congress on power plants, their characteristics and costs [166].

Technology (1)	Developer Type (2)	Non-Fuel O&M Cost (3)	Fuel Cost (4)	SO_2 and NOx Allowance Cost (5)	CO_2 Allow. Cost (6)	Prod. Tax Credit (7)	Total Operating Costs (8)	Capital Return (9)	Total Annualized $/Mwh (10)
Coal: Pulverized	IOU	$5.57	$11.13	$0.61	$0.00	$0.00	$17.31	$45.79	**$63.10**
Coal: IGCC	IOU	$5.46	$10.41	$0.10	$0.00	$0.00	$15.97	$67.02	**$82.99**
NG: Combined Cycle	IPP	$2.57	$30.57	$0.14	$0.00	$0.00	$33.27	$28.50	**$61.77**
Nuclear	IOU	$6.13	$5.29	$0.00	$0.00	($3.18)	$8.23	$74.99	**$83.22**
Wind	IPP	$6.67	$0.00	$0.00	$0.00	$0.00	$6.67	$74.07	**$80.74**
Geothermal	IPP	$13.69	$0.00	$0.00	$0.00	$0.00	$13.69	$45.54	**$59.23**
Solar: Thermal	IPP	$13.71	$0.00	$0.00	$0.00	$0.00	$13.71	$86.61	**$100.32**
Solar: Photovoltaic	IPP	$4.17	$0.00	$0.00	$0.00	$0.00	$4.17	$251.24	**$255.41**

The age and position of each generator also affects the production cost due to maintenance and transportation of fuel. This variation in cost is a large factor in deciding which power sources to turn on, up, down and off at unit commitment and the economic dispatch stages of power distribution. Although some of these costs are not specific to the cost of immediate power, they need to be negated somewhere. Some of these costs are fixed and some are variational. The cost curve, which is the cost at the generator as a function of the power produced, is shown in Figure 6.1 [167]:

Figure 5.1 Power output cost curve [167].

In the interval between 0 and MW_{min} it is uneconomical or impractical to produce power, and MW_{max} is the maximum the generator can produce ($MW_{min,}$ P_{min} and P_j will be used interchangeably, denoting the lowest power the generator can output. MW_{max} and Λ are also interchangeable, as are G_{max} and G_j). It has been suggested [167] that the curve is quadratic but could be modelled using a linear approximation; some also use piecewise linear models [168].

This chapter will consider and compare three different ways of considering the generation cost:

- Linear – much research has employed this simplified objective function as it results in a faster convergence and since some of the optimisation methods may be unable to consider non-linear functions.
- Negative Quadratic – the reason for trialling this model is to consider the jump between 0 and MW_{min}. This model allows for power to be produced within this gap, but the generator sends no power at no cost in this range (a way of simplifying unit commitment decisions). The negative quadratic outputs a steep cost between 0 and MW_{min} to encourage power stations to generate higher powers and to simulate start-up costs. However, between MW_{min} and MW_{max} the shape of the cost curve is unrepresentative.
- Positive Quadratic – this is a more representative model of the shape of the cost curve, but does not consider the jump from the generator being off to being on with a minimum MW. It therefore only considers economic dispatch, and not unit commitment; allowing us to assume that the generator is on and producing a minimum of MW_{min}.

5.2 Linear Cost Curve

5.2.1 Method

This simple method of describing the cost curve has been used by many, it can consider the MW_{min} and MW_{max} values and the associated costs but is unable to capture the shape of the cost curve.

Figure 5.2 shows the linear approximation of the cost curve, where the x-axis indicates the power entering the node, $\sum_{k \neq i} y_{jk} - y_{ij}$ (or negative sent power).

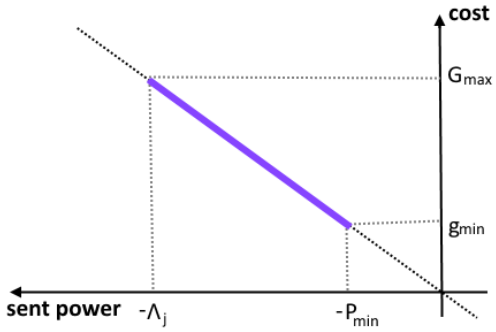


Figure 5.2 An experimental linear cost curve model.

To model this within the suggested message passing equations, one variable must be forfeit. As the message passing method relies on differentiating the VFE to find the optimal solutions, any constants within the equation will be lost. Therefore, the cost line will be considered as passing through the y-axis at $x = 0$. Because of this, only three points on the line can be predetermined; for example, if we are to assign the maximum point to be at $(-\Lambda_j, G_{max})$ this gives us the line $y = -\frac{G_{max}}{\Lambda_j} x$, although we can still set at $(-P_{min}, g_{min})$ the MW_{min} value, the cost at this point will automatically be $y = \frac{G_{max}}{\Lambda_j} P_{min}$. Alternatively, the minimum point can be set and the maximum cost will automatically be $y = \frac{g_{min}}{P_{min}} \Lambda_j$, we will use the first.

For later derivations of the other models we are required to change to vector notation, and for consistency we will start this here. In this notation $\boldsymbol{y}$ is a vector of all y_{jk} values for descendants and $\mathbf{1}$ signifies the one vector of length $c - 1$. In addition to the original constraint that all consumers should be satisfied, there are now three more conditions which need considering:

- The generation cost objective function is to be only applied to generators. Therefore, we require all nodes with a non-positive capacity to have a generation cost, $G_j = 0$, where G_j reflects the G_{max} cost when the generator has been fully utilised.
- We do not want any generators to give more power than they can produce. This can be seen with the original constraint (in matrix form):

$$(\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij} + \Lambda_j \geq 0, \tag{114}$$

where $\boldsymbol{\varepsilon}$ is the vector of ε_{jk} variables (vector summations are only with respect to descendants).

- Finally, we do not want generator nodes to receive power (the new cost functions could encourage power to flow towards generators for a profit), so we introduce:

$$(\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij} \leq 0. \tag{115}$$

This constraint can also be adjusted to consider P_{min} (which is negative for generators).

$$(\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij} + P_{min} \leq 0. \tag{116}$$

We can now consider the linearised VFE about the current values:

$$\begin{aligned} F_{ij} = \boldsymbol{\varepsilon}^T \boldsymbol{A} + \frac{1}{2}\boldsymbol{\varepsilon}^T \boldsymbol{B}\boldsymbol{\varepsilon} - \frac{G_j}{\Lambda_j}\left((\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij}\right) \\ + \mu_{ij}\left[(\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij} + \Lambda_j\right] \\ + G_j \gamma_{ij}\left[(\boldsymbol{y} + \boldsymbol{\varepsilon})^T \mathbf{1} - y_{ij} + P_j\right], \end{aligned} \tag{117}$$

where G_j is zero for all consumers, and a predetermined, individual generation cost value for each generator. There is also a second Lagrange multiplier, γ_{ij}, which indicates that generators must provide power values higher than the P_{min} (or P_j); by setting all consumers to $G_j = 0$ cost, the constraint is only enforced on generators.

The vector $\boldsymbol{A}$ represents the first derivative with respect to each descendant's power flow:

$$\boldsymbol{A} = \left[\frac{\partial(F_V + \phi)}{\partial y_{j1}} \quad \frac{\partial(F_V + \phi)}{\partial y_{j2}} \quad \cdots \quad \frac{\partial(F_V + \phi)}{\partial y_{jK}}\right], \tag{118}$$

where $K = c - 1$, the distribution cost ϕ, remains an objective function and the matrix $\boldsymbol{B}$ gives the second derivative with respect to descendant's power flows:

$$B = \begin{bmatrix} \frac{\partial^2(F_V+\phi)}{\partial y_{j1}^2} & \frac{\partial^2(F_V+\phi)}{\partial y_{j1}\partial y_{j2}} & \cdots & \frac{\partial^2(F_V+\phi)}{\partial y_{j1}\partial y_{jK}} \\ \frac{\partial^2(F_V+\phi)}{\partial y_{j2}\partial y_{j1}} & \frac{\partial^2(F_V+\phi)}{\partial y_{j2}^2} & & \frac{\partial^2(F_V+\phi)}{\partial y_{j2}\partial y_{jK}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2(F_V+\phi)}{\partial y_{jK}\partial y_{j1}} & \frac{\partial^2(F_V+\phi)}{\partial y_{jK}\partial y_{j2}} & \cdots & \frac{\partial^2(F_V+\phi)}{\partial y_{jK}^2} \end{bmatrix}. \quad (119)$$

Minimising the VFE with respect to the Lagrange multipliers gives:

$$\mu_{ij} = \min\left[\frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j + \left(-\boldsymbol{A} + \frac{G_j}{\Lambda_j}\mathbf{1}\right)^T \boldsymbol{B}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{B}^{-1}\mathbf{1}}, 0\right], \quad (120)$$

$$\gamma_{ij} = \max\left[\frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j + \left(-\boldsymbol{A} + \frac{G_j}{\Lambda_j}\mathbf{1}\right)^T \boldsymbol{B}^{-1}\mathbf{1}}{G_j\mathbf{1}^T\boldsymbol{B}^{-1}\mathbf{1}}, 0\right], \quad (121)$$

which when calculated gives $\mu_{ij} = f(\gamma_{ij})$, and $\gamma_{ij} = g(\mu_{ij})$; both functions are zero *only* if they are obeyed, and as they are non-overlapping constraints, we can assume that if either function is non-zero the other must be zero, allowing us to nullify each one in the other's equation. However, this is not the case for the derivatives of both functions.

Minimising the VFE again to find the optimal messages to the ancestor node gives the messages:

$$A_{ij} = \begin{cases} \frac{G_j}{\Lambda_j} - (\mu_{ij} + G_j\gamma_{ij}), & \text{in case 1,} \\ \frac{G_j}{\Lambda_j} - 2(\mu_{ij} + G_j\gamma_{ij}), & \text{in case 2,} \\ \frac{G_j}{\Lambda_j}, & \text{in case 3,} \end{cases} \quad (122)$$

and:

$$B_{ij} = \begin{cases} \dfrac{1}{\mathbf{1}^T \boldsymbol{B}^{-1}\mathbf{1}}, & \text{in case 1,} \\ \dfrac{4}{\mathbf{1}^T \boldsymbol{B}^{-1}\mathbf{1}}, & \text{in case 2,} \\ 0, & \text{in case 3,} \end{cases} \tag{123}$$

where the cases are defined in Table 4. Although we can assume each Lagrange multiplier to be zero if the other is non-zero, the same cannot be assumed about its derivative. In blue are the cases of inconsistent values, and in orange is the case that only happens if $MW_{min} = MW_{max}$. These cases can be better understood with Figure 5.3.

Table 4. The possible cases of A_{ij} and B_{ij}.

	$\mu_{ij} < 0, \frac{\partial \mu_{ij}}{\partial y_{ij}} \neq 0$	$\mu_{ij} = 0, \frac{\partial \mu_{ij}}{\partial y_{ij}} \neq 0$	$\mu_{ij} > 0, \frac{\partial \mu_{ij}}{\partial y_{ij}} = 0$
$\gamma_{ij} < 0, \frac{\partial \gamma_{ij}}{\partial y_{ij}} = 0$	Case 1	Case 1	Case 3
$\gamma_{ij} = 0, \frac{\partial \gamma_{ij}}{\partial y_{ij}} \neq 0$	Case 2	Case 2	Case 1
$\gamma_{ij} > 0, \frac{\partial \gamma_{ij}}{\partial y_{ij}} \neq 0$	Case 2	Case 2	Case 1

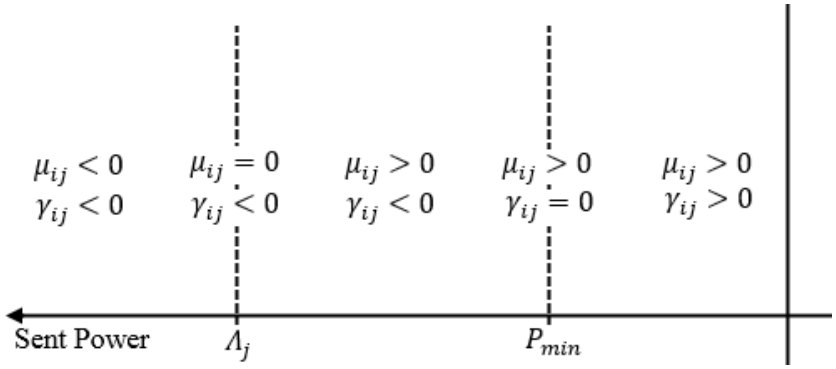


Figure 5.3 Possible combinations of μ_{ij} and γ_{ij} values.

A backwards message is given as:

$$\boldsymbol{y} = \boldsymbol{y} + \boldsymbol{B}^{-1}\left(-\boldsymbol{A} + \frac{G_j}{\Lambda_j}\mathbf{1} - (\mu_{ij} + G_j\gamma_{ij})\mathbf{1}\right). \tag{124}$$

5.3 Negative Quadratic Cost Curve

5.3.1 Method

In order to consider generation costs by a simple approximation of one of the unit commitment processes, we assume a negative cost curve objective function, which can be displayed in Figure 5.4.

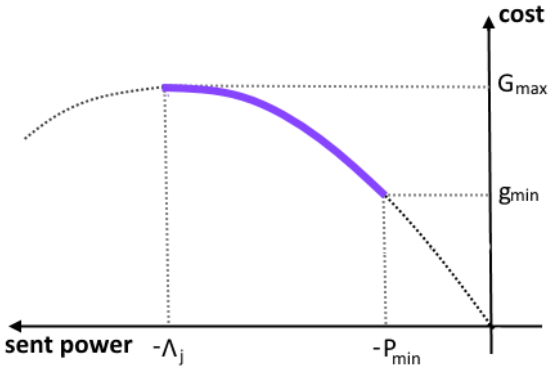


Figure 5.4 Experimental negative quadratic cost curve.

The curve considers that starting a generator incurs extra costs and therefore the difference between generating high to higher levels of power is much lower than the difference between generating none and a small amount, theoretically persuading the algorithm to choose higher power values at generators that are already producing instead of turning on other generators. The curve will be modelled by a negative quadratic function to maintain simplicity in the calculations. As the first and second derivatives of the VFE are required, functions such as the logarithm, the error function and the square root function become complicated quickly. The maximum cost is at the point where power flow is $-\Lambda_j$, and we mark this cost by the predetermined variable G_j. This gives us an objective function of:

$$-\frac{G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right)^2-\frac{2G_j}{\Lambda_j}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right). \tag{125}$$

The derivations are different in this case as the nonlinearity of the second derivative of function (125) is not zero when differentiated by a different power flow y_{jl} (this had been zero in earlier Equation (50)), and so we must now rewrite the derivation in matrix form:

$$-\frac{G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right)^T\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right)-\frac{2G_j}{\Lambda_j}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right). \tag{126}$$

As with the linear cost curve, it is possible to add the MW_{min} value, which automatically fixes the g_{min} value. Although this directly invalidates the main reason for trying negative quadratic curve costs, it still provides a non-linear model for the cost curve. We will consider this model primarily for comparison.

Calculating the optimal first and second derivative messages sent to node i using the Taylor expansion of the VFE with its new constraints:

$$\begin{aligned}F_{ij}=\boldsymbol{\varepsilon}^T\boldsymbol{A}+\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{B}\boldsymbol{\varepsilon}-\frac{G_j}{\Lambda_j^2}\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1}-y_{ij}\right)\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1}-y_{ij}\right)\\-\frac{2G_j}{\Lambda_j}\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1}-y_{ij}\right)+\mu_{ij}\left[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1}-y_{ij}+\Lambda_j\right]\\+G_j\gamma_{ij}\left[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1}-y_{ij}+P_j\right].\end{aligned} \tag{127}$$

The reason the generation cost is not considered within the Taylor expansion is that we only consider the cost of generation at node j for each VFE. Minimising with respect to $\boldsymbol{\varepsilon}$ gives:

$$\boldsymbol{\varepsilon}=\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\left(-\boldsymbol{A}+\frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}+\Lambda_j\right)^T\mathbf{1}-(\mu_{ij}+G_j\gamma_{ij})\mathbf{1}\right), \tag{128}$$

where $\boldsymbol{J}$ is the Jacobian matrix of size $c-1$. Putting the minimised $\boldsymbol{\varepsilon}$ back into F_{ij} gives the optimal VFE expression:

$$\begin{aligned}F_{ij}^*=-\frac{1}{2}\boldsymbol{A}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\boldsymbol{A}-\frac{G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right)^2-\frac{2G_j}{\Lambda_j}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}\right)\\+\frac{1}{2}\left(\mu_{ij}+G_j\gamma_{ij}\right)^2\mathbf{1}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}\\-\frac{2G_j^2}{\Lambda_j^4}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}+\Lambda_j\right)^2\mathbf{1}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}\\+\frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}+\Lambda_j\right)\boldsymbol{A}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}.\end{aligned} \tag{129}$$

Minimising the two Lagrange multipliers and recognising that both constraints cannot be unsatisfied at the same time, therefore one will always be zero, gives:

$$\mu_{ij} = \min\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j + \left(-\boldsymbol{A} + \frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right)^T\mathbf{1}\right)^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}\right], \quad (130)$$

$$\gamma_{ij} = \max\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j + \left(-\boldsymbol{A} + \frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right)^T\mathbf{1}\right)^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}{G_j\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}\right], \quad (131)$$

allowing us to obtain the messages:

$$A_{ij} = \begin{cases} \frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right) + \frac{4G_j^2}{\Lambda_j^4}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right)\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2}\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\boldsymbol{A} - \left(\mu_{ij} + G_j\gamma_{ij}\right) - \frac{2G_j}{\Lambda_j^2}\left(\mu_{ij} + G_j\gamma_{ij}\right)\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 1,} \\ \frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right) + \frac{4G_j^2}{\Lambda_j^4}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right)\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2}\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\boldsymbol{A} - 2\left(\mu_{ij} + G_j\gamma_{ij}\right) - \frac{4G_j}{\Lambda_j^2}\left(\mu_{ij} + G_j\gamma_{ij}\right)\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 2,} \\ \frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right) + \frac{4G_j^2}{\Lambda_j^4}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j\right)\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2}\mathbf{1}^T\left(\boldsymbol{B} - \frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\boldsymbol{A}, & \text{in case 3,} \end{cases} \quad (132)$$

where the cases are explained in Table 4, and:

$$B_{ij} = \begin{cases} \dfrac{1}{\mathbf{1}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}+\dfrac{2G_j}{\Lambda_j^2}, & \text{in case 1,} \\ \dfrac{14G_j}{\Lambda_j^2}+\dfrac{4}{\mathbf{1}^T\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}}+\dfrac{12G_j^2}{\Lambda_j^4}\mathbf{1}^T\left(\boldsymbol{B}-\dfrac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 2,} \\ -\dfrac{2G_j}{\Lambda_j^2}-\dfrac{4G_j^2}{\Lambda_j^4}\mathbf{1}^T\left(\boldsymbol{B}-\dfrac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 3,} \end{cases} \quad (133)$$

giving a backwards message of:

$$\boldsymbol{y} = \boldsymbol{y}+\left(\boldsymbol{B}-\frac{2G_j}{\Lambda_j^2}\boldsymbol{J}\right)^{-1}\left(-\boldsymbol{A}+\frac{2G_j}{\Lambda_j^2}\left(\boldsymbol{y}^T\mathbf{1}-y_{ij}+\Lambda_j\right)\mathbf{1} - \left(\mu_{ij}+G_j\gamma_{ij}\right)\mathbf{1}\right). \quad (134)$$

5.4 Positive Quadratic Cost Curve

5.4.1 Method

This curve will allow us to model more accurately the shape of the cost curve, but does not consider the case of the generator being off. The curve can be shown as:

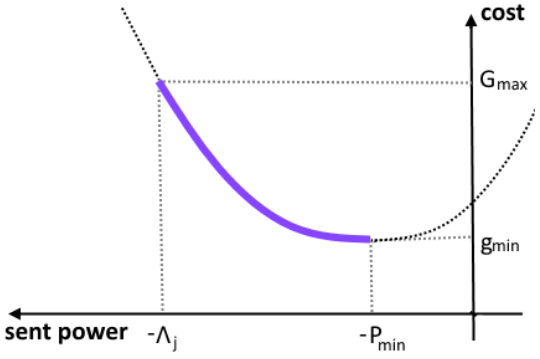


Figure 5.5 Experimental positive quadratic cost curve.

Here we can see that the power provided by the generator is required to be between the point MW_{max} and MW_{min}. As we are sending the first and second derivatives, any constant of the quadratic function would be lost. Therefore, we can only consider a function of the form $y = Ax^2 + Bx$ where other constants cannot be considered. If we chose the point -MW_{min} to be the minimum of the quadratic, at this point $0 = 2Ax + B$, and $x = -\frac{B}{2A}$. If we require at this point for the cost to be g_{min}, this gives us the function $y = \frac{g_{min}}{3P_j^2}x^2 + \frac{2g_{min}}{3P_j}x$. In this case, as we cannot determine the G_{max} value separately as its value is defined directly according to the function. Alternatively, G_{max} the cost required at MW_{max}, can be the coefficient we choose to fix, which gives $y = \frac{G_j}{\Lambda_j^2 - 2P_j\Lambda_j}x^2 + \frac{2P_jG_j}{\Lambda_j^2 - 2P_j\Lambda_j}x$, and g_{min} is automatically determined. To best compare the three methods, we will use the latter.

This allows us to write a new Taylor expansion of the VFE:

$$\begin{aligned} F_{ij} = \boldsymbol{\varepsilon}^T\boldsymbol{A} + \frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{B}\boldsymbol{\varepsilon} &+ \frac{G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij}\right)\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij}\right) \\ &+ \frac{2P_jG_j}{\Lambda_j^2 - 2P_j\Lambda_j}\left((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij}\right) \\ &+ \mu_{ij}\left[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij} + \Lambda_j\right] \\ &+ G_j\gamma_{ij}\left[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij} + P_j\right], \end{aligned} \tag{135}$$

giving the optimal Lagrange multipliers:

$$\mu_{ij} = \min\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j + \left(-\boldsymbol{A} - \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j\right)^T\mathbf{1}\right)^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}}\right], \tag{136}$$

$$\gamma_{ij} = \max\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j + \left(-\boldsymbol{A} - \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\left(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j\right)^T\mathbf{1}\right)^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}}{G_j\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}}\right], \tag{137}$$

allowing us to obtain the messages to ancestor i:

$$A_{ij} = \begin{cases} \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\boldsymbol{A} + \frac{4G_j^2}{\left(\Lambda_j^2 - 2P_j\Lambda_j\right)^2}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j)\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j) - (\mu_{ij} + G_j\gamma_{ij}) + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\mu_{ij} + G_j\gamma_{ij})\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 1,} \\ \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\boldsymbol{A} + \frac{4G_j^2}{\left(\Lambda_j^2 - 2P_j\Lambda_j\right)^2}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j)\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j) - 2(\mu_{ij} + G_j\gamma_{ij}) + \frac{4G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\mu_{ij} + G_j\gamma_{ij})\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 2,} \\ \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\boldsymbol{A} + \frac{4G_j^2}{\left(\Lambda_j^2 - 2P_j\Lambda_j\right)^2}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j)\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1} \\ -\frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j), & \text{in case 3,} \end{cases} \tag{138}$$

and,

$$B_{ij} = \begin{cases} \frac{1}{\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}} - \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}, & \text{in case 1,} \\ -\frac{14G_j}{\Lambda_j^2 - 2P_j\Lambda_j} + \frac{4}{\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}} + \frac{12G_j^2}{\left(\Lambda_j^2 - 2P_j\Lambda_j\right)^2}\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 2,} \\ \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j} - \frac{4G_j^2}{\left(\Lambda_j^2 - 2P_j\Lambda_j\right)^2}\mathbf{1}^T\left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\mathbf{1}, & \text{in case 3.} \end{cases} \tag{139}$$

Using the same cases as in Table 4. Backwards messages are given by:

$$\boldsymbol{y} = \boldsymbol{y} + \left(\boldsymbol{B} + \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}\boldsymbol{J}\right)^{-1}\left(-\boldsymbol{A} - \frac{2G_j}{\Lambda_j^2 - 2P_j\Lambda_j}(\boldsymbol{y}^T\mathbf{1} - y_{ij} + P_j)\mathbf{1} - (\mu_{ij} + G_j\gamma_{ij})\mathbf{1}\right). \tag{140}$$

5.5 Results

We will now show some properties of the minimising generation cost algorithms. Figure 5.6 to Figure 5.9 show pie charts of how different generation cost values change the proportion of power provided from each generator using a positive quadratic cost curve with P_{min} equals zero. When all generators have a low generation cost, the distribution cost is minimised with a higher priority allowing a variation of power sent per generator in Figure 5.6 due to topological

considerations. When all the generation costs are high, this leads to an even distribution of power provided among the generators as the cost of distribution is less important in comparison, as shown in Figure 5.7. If one node has a much higher generation cost than the others, we can see how this generator provides a negligible amount of power to the network in Figure 5.8 (this is because there is enough power in the network to satisfy the consumers, with one generator off; on a poorer network, the generator would still be required to provide some power). And finally, if one generator is much cheaper than the rest it provides more power than the other generators (but not exceeding its available capacity), Figure 5.9. It is to be recognized here that the power supplied is the maximum the station can provide, and assigning the node an even lower generation cost would not change the results.

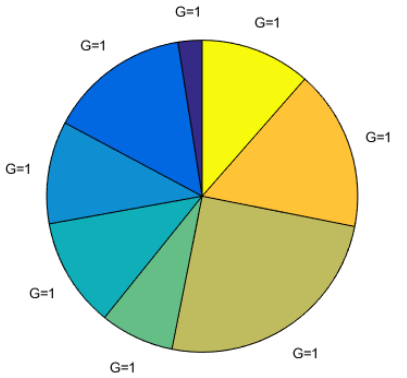


Figure 5.6 All generators have a generation cost value of 1, from an optimised synthetic network with generation larger than overall demand (20-nodes, 8 consumers with $\Lambda_j = \mathcal{F}(-2)$ And 8 generators with $\Lambda_j = \mathcal{F}(4)$). Pie chart represents the proportion of power provided to consumers from each generator, where G represents the generation cost of each generator.

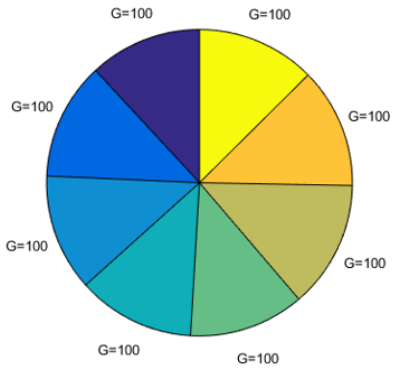


Figure 5.7 All generators have a generation cost value of 100. All other properties correspond with those of Figure 5.6.

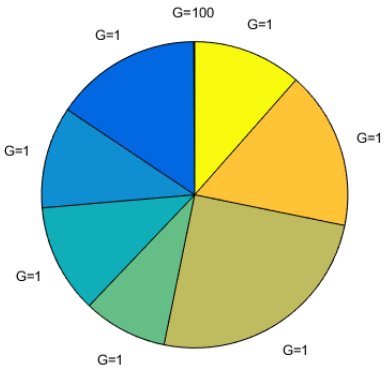


Figure 5.8 All generators have a generation cost value of 1, apart from one node with a generation cost of 100. All other properties correspond with those of Figure 5.6.

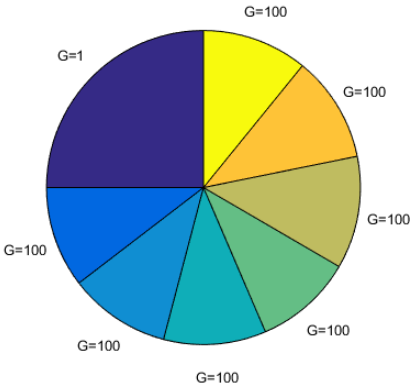


Figure 5.9 All generators have a generation cost value of 100, apart from one node with a generation cost of 1. All other properties correspond with those of Figure 5.6.

The following graphs indicate the power sent from each generator when one generators cost is increasing. Figure 5.7 and Figure 5.9 are demonstrated through Figure 5.10 where we see how when all generators costs are large, the demand is divided equally between nodes; but as one generators' cost decreases, making it cheaper to provide power, it sends more than the other nodes to reduce the amount the other generators must provide. The generator continues to send more power until it is fully utilised. Figure 5.6 and Figure 5.8 can be seen in Figure 5.11, where the generators all have low generation costs so distribution cost is considered more prominently, this results in each generator providing unequal amounts of power to reduce the distribution cost. As generator 1's generation cost increases, we see how the program aims to minimise the power this node must provide, as it is now more expensive, by sending the power from cheaper generators.

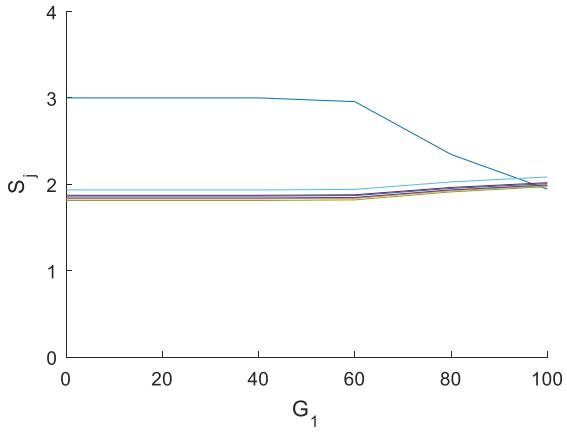


Figure 5.10 The power sent from all 8 generators from Figure 5.6, where generators 2 to 8 have generation cost $G_{2:8} = 100$ and G_1 is increasing. Capacities are fixed at $\mathcal{F}(3, -2)$.

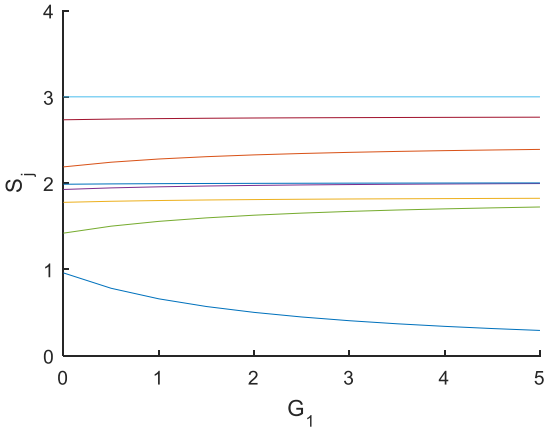


Figure 5.11 The power sent from all 8 generators from Figure 5.6, where generators 2 to 8 have generation cost $G_{2:8} = 1$ and G_1 is increasing. Capacities are fixed at $\mathcal{F}(3, -2)$.

Figure 5.12 represents the distribution of the power sent, S, from generator nodes to consumers. The graph shows how the connectivity of the network does not appear to largely affect the proportion of power provided by each of the nodes, however, this may be because the average generation cost is high; with a lower generation cost this may result in different effects because the distribution cost would have a higher priority, so the cost of each generator would be weighted more equally with the distance from consumer, in this case the curve would be expected to be less distinct especially for lower connectivities when distribution costs would be higher. Figure 5.13 shows how for any spread of the generation cost values, the nodes with lower costs send their maximum capacity first. For lower generation costs ($\mathcal{U}(0,1)$ and $\mathcal{U}(0,0.5)$) this correlation becomes slightly less pronounced as distribution cost has a higher priority than it does with larger average generator cost values. The graphs demonstrate how the approach works very well at prioritising cheaper generators to be used.

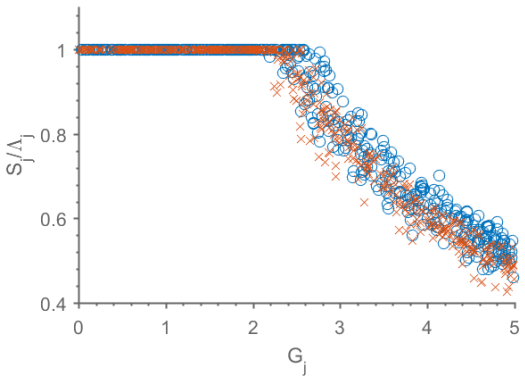


Figure 5.12 Proportion of power provided by each of the generators with respect to its own capacity. The degree connectivity of the different cases is $c = 5$ (blue), 7 (orange). The generation costs are marked on the x-axis and are randomly drawn $G_j = \mathcal{U}(0,5)$.

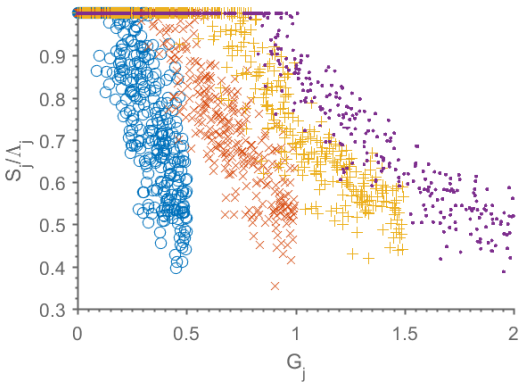


Figure 5.13 Proportion of power provided by each of the generators with respect to its own capacity. The generation cost is randomly chosen from a uniform distribution between $\mathcal{U}(0, G^+)$ where $G^+ = 0.5$ (blue) 1 (orange), 1.5 (yellow) and 2 (purple).

Figure 5.14 shows the very consistent and convex nature of the power supplied by each generator, as fraction of capacity, according to each nodes generation cost (x-axis values). One can see how as the network has less overall available power, the generators with the highest costs are prioritised as the cheaper generators provide more power. Figure 5.15 shows the difference in correlation between power sent and generation cost as the average capacity is now chosen from a normal distribution. We can see how the poorer networks (such as blue) result in most generators providing power, irrespective of generation cost, whereas more wealthy networks (such as purple) have less overall power sent (due to less demand), and the negative correlation between power sent and generation cost is very apparent. In blue it can still be seen that lower generation cost nodes have been fully utilised, but this graph demonstrates how these poorer networks are less affected by generation cost values, we assume this is because power flow is increased, and therefore distribution cost is higher and carries a higher weight against generation cost. A large difference can be seen between the Figure 5.14 and Figure 5.15, this is due to the distribution of the capacities; even if the mean is the same there will be a higher proportion of generators and a lower proportion of consumers with the normal distribution thus less need for sent power, in addition to less power being sent, the producers capacities are entirely independent of generation cost, meaning some low generation costs could have high capacity and satisfy consumers without sending all of their power and some high generator costs with low capacity, even when they send a small amount of power this will appear to be a bigger proportion, making the graph appear in general more random. This suggests that minimising generation costs may be aided by a more consistent set of capacities, although this may not be practical to implement.

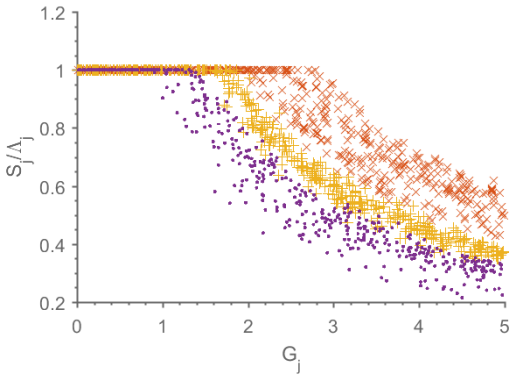


Figure 5.14 Proportion of power dispatched by each generator with respect to its own capacity and as a function of generation cost (x-axis). Where the capacities are given fixed values $\mathcal{F}(\Lambda^m, -1)$ where $\Lambda^m = 1.2$ (orange ×), 1.4 (yellow +) and 1.6 (purple •).

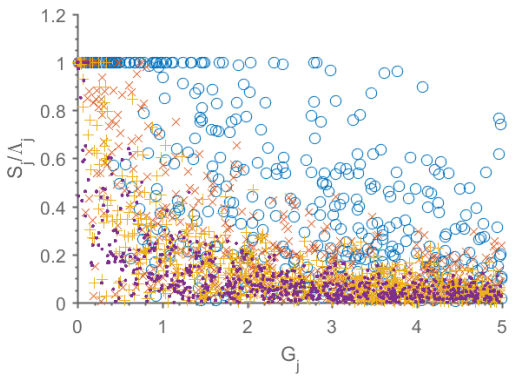


Figure 5.15 Proportion of power dispatched by each generator with respect to its own capacity and as a function of generation cost (x-axis). Where the capacities are chosen randomly from a Gaussian distribution $\mathcal{N}(\Lambda^m, 1)$ where $\Lambda^m =$ 0.2 (blue ○), 0.4 (orange ×), 0.6 (yellow +) and 0.8 (purple •).

Figure 5.16 and Figure 5.17 show two different cost curve algorithms tested on the same network, and it is obvious that the cost curve used has a high impact on the spread of power dispatched from the various generators. In Figure 5.16 we see a linear cost curve used within ten 100-node networks, there is a high proportion sending no power at all, a high proportion sending everything, and a low proportion sending some; with a very obvious transition for generators around some threshold cost value. This threshold values depends on the mean of the distribution of generation values. Figure 5.17 demonstrates the power dispatched from generators as a fraction of their capacities under the same conditions as in Figure 5.16 for a positive quadratic cost curve. We see a much higher proportion of nodes providing some power, a lower fraction sending everything and no generators sending no power at all. The curvature of each graph appears to reflect the shape of its cost curve. These diagrams show the importance of an accurate modelling of the cost curve; The efficacy of the cost model used in real scenarios depends on the preferred properties in the various cases. These results may vary with parameter changes of the cost curve, or different P_{min} values. Figure 5.16 and Figure 5.17 also shows how a higher average generation cost values results in approximately the same sent curves but at higher G values; as the generators with the highest costs are still prioritised in minimising the dispatch power.

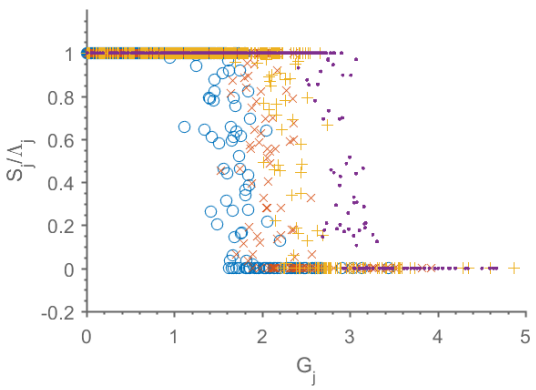


Figure 5.16 Proportion of available power supplied from generator nodes with respect to their individual capacities as a function of their specific cost values when the generation cost curve is linear. Plotted using ten 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the distribution of the generation cost values drawn from $\mathcal{N}(G^m, 1)$, where $G^m = 0.5$ (blue), 1 (orange), 1.5 (yellow), 2 (purple).

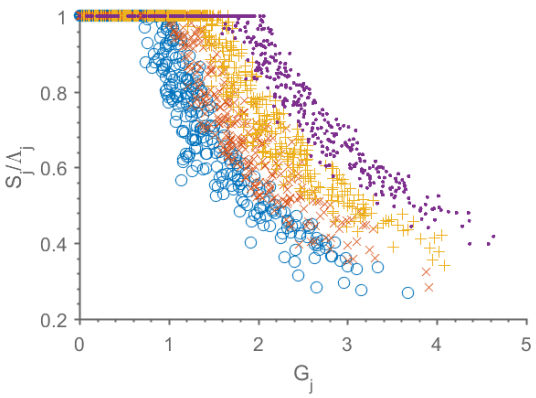


Figure 5.17 Proportion of available power supplied from generator nodes with respect to their individual capacities as a function of their specific cost values when the generation cost curve is positive quadratic. Plotted using ten 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the distribution of the generation cost values being $\mathcal{N}(G^m, 1)$, where $G^m = 0.5$ (blue), 1 (orange), 1.5 (yellow), 2 (purple).

A large difference in the proportion of power dispatched per generator can be seen in Figure 5.18, where the generation costs were assigned in correlation with the maximum capacity values of each generator. We can see that when there is a negative correlation between the two parameters, a very smooth negative cumulative type curve is formed, this is because generators of low capacity and a low cost will quickly be fully utilised and capacities will be gradually exhausted according to the increase in generation costs. A very different straight line between 0.2 and 0.4 can be seen when the parameters are positively correlated, since generators with higher capacities are used due to their low generation cost but the proportion of power dispatched with respect to their capacities remains fairly low, it is also more expensive to use the lower capacity generators and hence only a small proportion of these are used. This is a very useful graph when considering real networks, coal power stations usually have a smaller capacity than nuclear power stations, but higher generation cost; this coincides with a negative correlation (the blue curve). From Figure 5.19 we see how increasing the fixed value generation costs of all generators results in a decrease in the distribution of the power provided, converging towards the fixed power required by consumers. This demonstrates the importance of appropriate weighting between the generation cost and the distribution cost objective functions, we suggest a high average generation cost weight would be most appropriate to real objectives, to prioritise it over distribution costs. It can also be seen that the number of nodes in the network N does not appear to largely affect the power provided per generator.

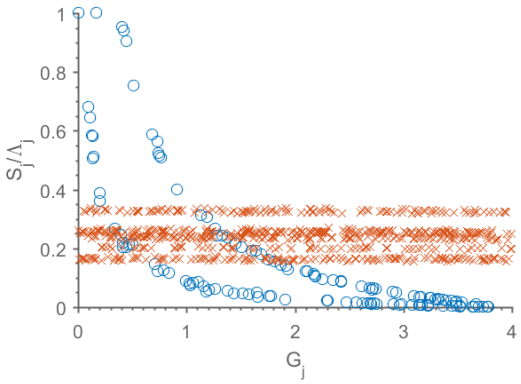


Figure 5.18 Proportion of available power supplied from generator nodes with respect to their individual capacities as a function of their specific cost values when the generation cost curve is positive quadratic. Plotted using ten 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the distribution of the generation cost values being $\mathcal{C}(-1$ (blue ○),$+1$ (orange ×), $\Lambda)$.

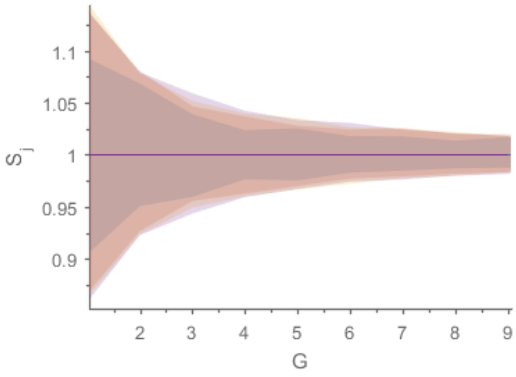


Figure 5.19 The power sent from each generator node as the fixed generation cost value increases when the generation cost curve is positive quadratic. Plotted using ten 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the size of the network $N = 40$ (blue), 80 (orange), 120 (yellow) and 160 (purple).

Figure 5.20 demonstrates the effect of increasing the fixed P_{min} value on each generator. We see how this decreases the distribution of the dispatched power. This is because as the minimum power sent from each generator increases, this additional power sent to consumers allows for the maximum power supplied from other generators to decrease. We see from Figure 5.21 how this increase in P_{min} value results in a quadratic increase in the overall networks generation cost because extra power is being sent from generators, that in the minimised generation cost algorithm were not optimal, to satisfy the constraint (the generation cost is calculated from the given cost function). We can also see how a larger fixed generation cost value increases the generation cost of the network, this is solely because of the fixed constant G_j value in the generation cost value $G^+ = \sum_j \frac{G_j}{\Lambda_j^2 - 2P_j\Lambda_j} S_j^2 + \frac{2P_jG_j}{\Lambda_j^2 - 2P_j\Lambda_j} S_j$; otherwise all curves follow the purple curve.

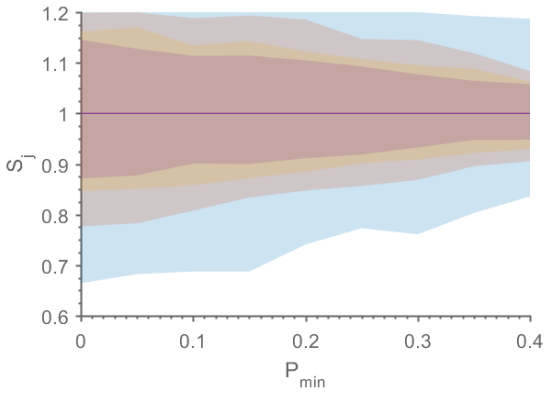


Figure 5.20 The power supplied from each generator node as the P_{min} value increases when the generation cost curve is positive quadratic. Plotted using 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the distribution of the generation cost values being $\mathcal{F}(0.25)$ (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

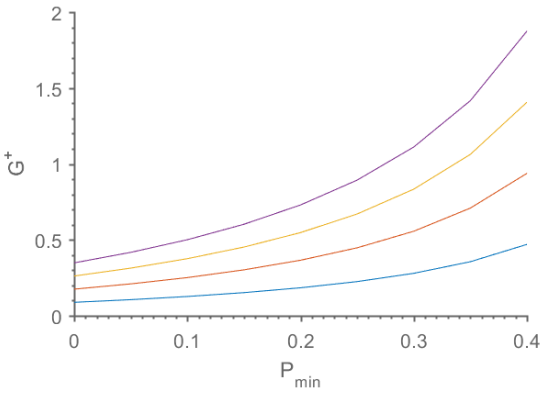


Figure 5.21 The generation cost value of the network as the P_{min} value increases when the generation cost curve is positive quadratic. Plotted using 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. Colours represent the distribution of the generation cost values being $\mathcal{F}(0.25)$ (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

Figure 5.22 shows an example of a network where the P_{min} values are fixed at a proportion of the nodes capacity. It can be seen how due to the generators having to dispatch a minimum amount according to their P_{min} value, excess power is given to consumers. Assuming no excess is necessary (ignoring fluctuations and $N-1$ contingencies) this scenario results in an unnecessary generation cost *and* distribution cost and a good power grid management would look to turn off some generators to avoid it.

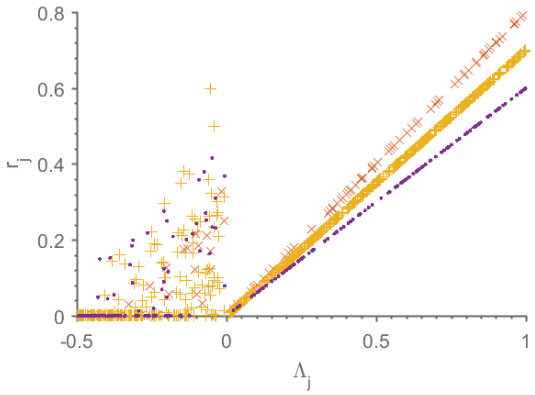


Figure 5.22 The remaining power at each node according to its original capacity from multiple 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{F}(-1,1.2)$. The P_{min} values on each generator are fixed at a percentage of the nodes capacity; $P_{min} = 0.2\Lambda_j$ (orange), $0.3\Lambda_j$ (yellow) and $0.4\Lambda_j$ (purple).

The main difference between the linear, the positive and the negative quadratic cost curve can be seen from Figure 5.23 - Figure 5.25. We see how for a positive quadratic, the higher capacity generators end up with higher reserves, this is because the cost curve looks to minimise the power supplied from each generator and favours a smaller value from each generator, whereas with a negative quadratic curve, the larger capacity generators appear to result in smaller remaining power as the cost curve favours generators maximising their capacities. The reserve given using the linear curve is entirely from larger capacities, and the curve appears to work similarly to the negative curve; utilising the higher capacities fully first before turning other generators on. The difference in the three graphs highlights the importance of the accurate modelling of the cost curve.

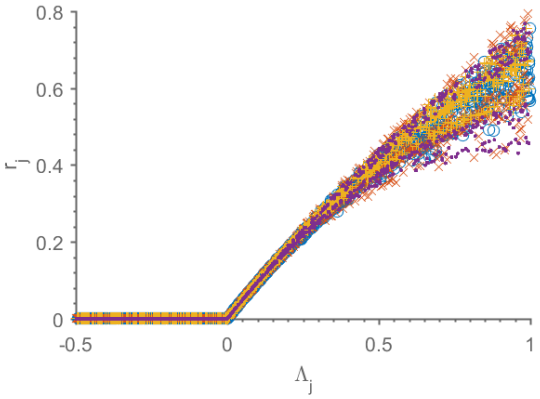


Figure 5.23 The reserve power remaining at each node as a function of its own Λ_j value when the generation cost curve is positive quadratic. Plotted using multiple 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{U}(-0.5,1)$. Colours represent the distribution of the generation cost values being $\mathcal{F}(0.25)$ (blue), 0.5 (orange), 0.75 (yellow) and 1 (purple).

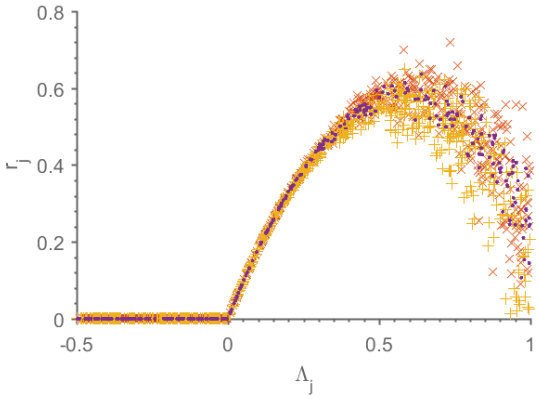


Figure 5.24 The reserve power remaining at each node as a function of its own Λ_j value when the generation cost curve is negative quadratic. Plotted using multiple 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{U}(-0.5,1)$. Colours represent the distribution of the generation cost values being $\mathcal{F}(0.5)$ (orange), 0.75 (yellow) and 1 (purple).

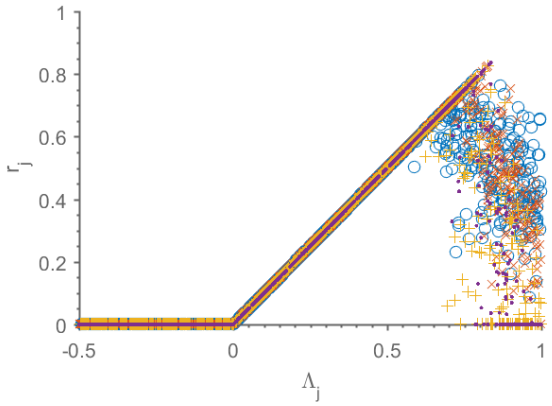


Figure 5.25 The reserve power remaining at each node as a function of its own Λ_j value when the generation cost curve is linear. Plotted using multiple 100-node networks with random regular connectivity 6 and $\Lambda = \mathcal{U}(-0.5,1)$. Colours represent the distribution of the generation cost values being $\mathcal{F}(0.5)$ (orange), 0.75 (yellow) and 1 (purple).

5.6 Discussion

All three models, linear, positive and negative quadratic, work well, but arguably the most representative model is the positive quadratic cost, which we have focused on. The results show that the model and the resulting message passing method can successfully redistribute the power flow to minimise the production cost. We see how increasing the cost of one generator effects the others in the network and how the value of the generation cost with respect to the distribution cost effects the power flow over lines. Results show that increasing generation costs decreases the distribution of power provided from generators, as does increasing the P_{min} values, but increases distribution cost. We see how P_{min} values are successful even when they result in power sent exceeding power demand, which results in reserve power at consumers. It is also shown how positive quadratic cost curves favour more generators being on and producing smaller amounts of power, whereas negative quadratic and linear cost curves prefer to maximise the generators capacities before turning on new ones. In addition, we see that a positive correlation of generation cost and capacity results in most generators sending equal proportions of their power, whereas a negative correlation favours maximising the utilisation of generators with low production costs.

Although considering the generation cost in this manner works relatively well, the objective function could be made more effective. For example, if the objective function was to allow a generator to produce nothing at zero cost this could represent the cost of turning off a generator at the unit commitment stage, this is possible by introducing more constraints with the help of more Lagrange multipliers.

The calculations also do not consider prohibited zones which are values of MW between MW_{min} and MW_{max} that are not accessible as explained in [169] and with the corresponding Figure 5.26. These zones may be due to turning on second and third reactors within the same power generation station for higher outputs. Time delays for the turning up or down of generators are not considered within these calculations as economic dispatch considers only instances at a time, not the dynamic state of the system. It is assumed that these can be taken care of before the start of the next time-frame or at the unit-commitment stage; further work suggested on minimising the deviation of variables in Section 8.2 may be able to consider this.

Figure 5.26 The cost curve considering prohibited zones [169].

6 Voltages

Work done so far has been quite abstract, because although we assume power is required by consumers and supplied by generators, we have talked until now about power being supplied over edges while ignoring the physical variables to be determined such as voltages, frequencies, currents and phases. We should therefore adapt the suggested framework to the physical variables used, for instance according to the simplified ACOPF (or DCOPF) in Equation (18); power is the difference between phase angles divided by reactance. In accordance with DCOPF, in this chapter we will use the variables, v_j for the phase angles at node j and R_{ij} as the reactance over the line (ij) (we will describe them as current, voltage and resistance for simplicity and in accordance with the comparison to DCOPF):

$$y_{ij} = \frac{v_j - v_i}{R_{ij}}. \tag{141}$$

6.1 Current Potential

6.1.1 Method

Adjusting voltages instead of assigning currents is more practical for power grids and as mentioned in the work done by Wong and Saad [76], the current y_{ij} can be expressed as current potential $(v_j - v_i)$, also mentioning that in the specific case where the objective function is quadratic, the chemical potential used in their framework and current potentials are interchangeable.

There are two obvious ways of considering the difference in voltages instead of the power flow (ignoring resistance for the moment). The model could either consider voltages as the control variables instead of the current, y_{ij}, or we could calculate the distribution of currents, and then take the difference to calculate the voltages; we will focus on the first model. The current message passing equations allows voltages to be at *any* value, as long as the difference between them is correct. Rewriting the equations in terms of voltages does have its complications; this is because adjusting one voltage affects all currents it is connected to. Therefore, if we were to adjust the voltage at a parent node j according to the information sent to the ancestor node (y_{ij}),

the current sent to the ancestor y_{ij}, would be changed by the adjustment, resulting in the backwards messages to nodes k, being unrepresentative of the present situation. In addition, the messages sent from node j to an ancestor, i, would be the derivatives of the VFE with respect to a single node v_i, these messages (A_i and B_i) would contain the information necessary for node i from node j, but the information would only depend on node i, disregarding the information sent from any of node i's other neighbours, which may result in inaccurate messages. One way to avoid these problems is to look at the situation of double bipartite factor graph.

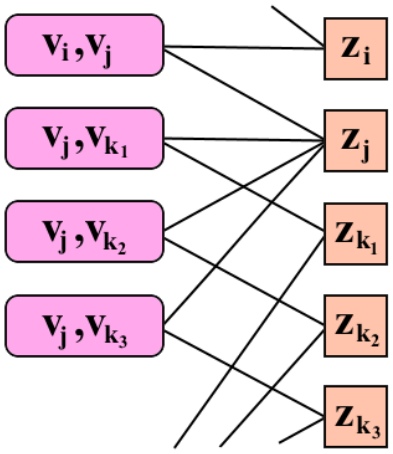


Figure 6.1 Example of a double variable bipartite factor graph.

For power distribution, we assign each double node (v_j, v_i) resulting in the equations of the power distribution, being almost exactly the same as before. The only real difference is that the adjustment, is formally made to $v_k - v_j$. However, we set the adjustment variables $\varepsilon_j = 0$ in order to avoid the extra degree of freedom and reduce the complexity of the expressions obtained; so, all adjustments are around v_k only. One variable adjustments may result in the equations not converging in cases where small loops are present.

The VFE Taylor Expansion now can be shown as:

$$F_{ij} = \sum_{k \neq i} \left[\left(A_{jk} + \phi'_{jk}\right)\varepsilon_k + \frac{1}{2}\left(B_{jk} + \phi''_{jk}\right)\varepsilon_k^2 \right] + \mu_{ij}\left[\sum_{k \neq i} \left(v_k + \varepsilon_k - v_j\right) - \left(v_j - v_i\right) + \Lambda_j \right], \tag{142}$$

where $A_{jk} = \frac{\partial F_{jk}}{\partial (v_k - v_j)}$, $B_{jk} = \frac{\partial^2 F_{jk}}{\partial (v_k - v_j)^2}$, $\phi'_{jk} = \frac{\partial \phi}{\partial (v_k - v_j)}$ and $\phi''_{jk} = \frac{\partial^2 \phi}{\partial (v_k - v_j)^2}$.

This process is very similar to previous calculations in Section 2.4 and minimising the expression with respect to ε_k gives:

$$\varepsilon_k = -\frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{\left(B_{jk} + \phi''_{jk}\right)}, \tag{143}$$

resulting in the optimal solution of F_{ij} given by:

$$F^*_{ij} = \frac{1}{2} \sum_{k \neq i} \mathcal{A}_{jk} \frac{\mu_{ij}^2 - \left(A_{jk} + \phi'_{jk}\right)^2}{B_{jk} + \phi''_{jk}}, \tag{144}$$

where:

$$\mu_{ij} = \min \left[0, \frac{\sum_{k \neq i} \mathcal{A}_{jk} \left(v_k - \frac{A_{jk} + \phi'_{jk}}{B_{jk} + \phi''_{jk}} - v_j \right) - v_j - v_i + \Lambda_j}{\sum_{k \neq i} \mathcal{A}_{jk} \frac{1}{B_{jk} + \phi''_{jk}}} \right]. \tag{145}$$

The derivatives of F^*_{ij} with respect to $v_j - v_i$ lead to the forward message A_{ij}, B_{ij} from node j to i:

$$A_{ij} \leftarrow -\mu_{ij}, \tag{146}$$

$$B_{ij} \leftarrow \frac{\Theta(-\mu_{ij} - \epsilon)}{\sum_{k \neq i} \mathcal{A}_{jk} \frac{1}{B_{jk} + \phi''_{jk}}}. \tag{147}$$

The backwards message is given by:

$$v_k \leftarrow v_k - \frac{A_{jk} + \phi'_{jk} + \mu_{ij}}{B_{jk} + \phi''_{jk}}; \tag{148}$$

this works well for the quadratic cost function.

6.2 Limiting Voltages

6.2.1 Method

One advantage of writing the equations in the form of voltages, is that it can allow us to limit the values of the voltages. In real-life the voltages need to be between 1.05 and 0.95 per unit of the prescribed values; these are our v_{min} and v_{max}. Previous work done on bandwidths [109] can be used to limit voltages.

A piecewise linear representation of the limits on the voltages can be described using Figure 6.2.

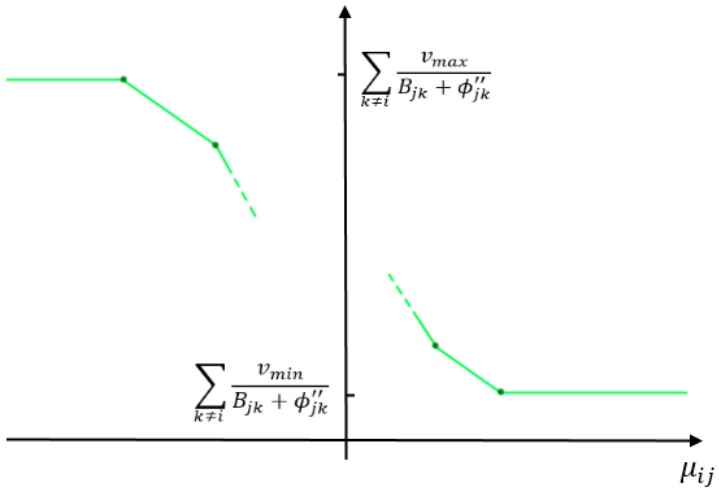


Figure 6.2 Piecewise linear function of μ_{ij} when voltages are limited by v_{min} and v_{max}.

This gives the equation:

$$\min\left[v_{max}, \max\left[v_{min}, v_k - \frac{A_{jk}+\phi'_{jk}}{B_{jk}+\phi''_{jk}}\right]\right] = cv_j - v_i - \Lambda_j, \tag{149}$$

where the values on the x-axis at each downward kink in Figure 7.2 are given by:

$$(v_k - v_{min})\left(B_{jk}+\phi''_{jk}\right) - A_{jk} - \phi'_{jk}, \tag{150}$$

and at the upwards kinks are:

$$(v_k - v_{max})\left(B_{jk}+\phi''_{jk}\right) - A_{jk} - \phi'_{jk}. \tag{151}$$

The corresponding y values are calculated as in Equations (68) and (69), with the first and last y-values at:

$$y_1 = \sum_{k \neq i} \frac{v_{max}}{B_{jk} + \phi''_{jk}}, \tag{152}$$

$$y_{2N} = -\sum_{k \neq i} \frac{v_{min}}{B_{jk} + \phi''_{jk}}. \tag{153}$$

6.3 Resistance and Reactance

6.3.1 Method

Every power line has some level of resistance and reactance; whether we are discussing ACOPF with assumptions or DC power flow. Reactance determines the amount of current passed according to phase angle differences in DCOPF and resistance influences the power lost from each edge as current passes through it, $P = I^2R$ in DC equations. This section will discuss both ways in which our message passing equations can incorporate resistance and reactance, starting with the latter.

Resistance

In DC power flow, to minimise power loss, similar to [109], we can use the power loss equation within the objective function on the edges:

$$\phi_{ij} = R_{ij}\left(v_j - v_i\right)^2. \tag{154}$$

When this is implemented into the message passing equations the algorithm can reduce the amount of current that passes through higher resistance edges.

Reactance

To bring the message passing derivations closer to that of a real AC network (with assumptions - DCOPF), we consider that the power at a node $P_j = \sum_j P_{ij}$ where P_{ij} is the power going into node j from all neighbours i and using the equation made from assumptions in Section 1.3:

$$P_{ij} = \frac{\theta_j - \theta_i}{x_{ij}}. \tag{155}$$

We can rewrite our F_{ij} as:

$$\begin{aligned} F_{ij} = \sum_{k \neq i} \mathcal{A}_{jk} \left[\left(A_{jk} + \phi'_{jk}\right)\varepsilon_k + \frac{1}{2}\left(B_{jk} + \phi''_{jk}\right)\varepsilon_k^2 \right] \\ + \mu_{ij} \left[\sum_{k \neq i} \left(\frac{\theta_k + \varepsilon_k - \theta_j}{x_{jk}} \right) - \left(\frac{\theta_j - \theta_i}{x_{ij}} \right) + \Lambda_j \right], \end{aligned} \tag{156}$$

where ε_k is now a small adjustment about the variable θ_k, A_{jk} and B_{jk} are the derivatives of the VFE with respect to the double node (θ_k, θ_j) and Λ_j is the required power.

This should more closely follow the equations of DCOPF and demonstrate some of the capacity of the message passing algorithms to be adapted to optimising an OPF problem.

6.4 Results

To test the efficacy of the new framework and the resulting algorithms, we have tested the message passing approach with control variables on the nodes against the same approach on edge variables from Section 2.4.

Figure 6.3 shows that the distribution cost is minimised to the same level when voltages or currents are used as the control variables, suggesting there is no compromise in using one or the other.

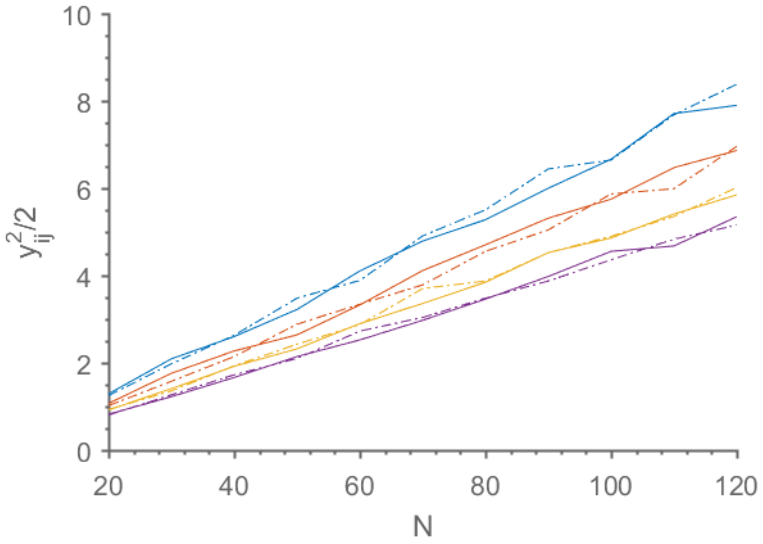


Figure 6.3 The distribution cost as the system size N increases when the topology is plotted using a Minimum Distance model (as described in Appendix B), in order to more accurately represent a power grid topology, with $\mathcal{M}(N, aN)$, where $a = 6$ (blue), 7 (orange), 8 (yellow) and 9 (purple) and where dashed lines indicate using an algorithm where control variables are on nodes (v_j) and full lines indicate control variables on edges (y_{ij}).

Although minimising voltage variables does appear to find a solution just as good as minimising current variables, the voltage algorithms do take longer to converge on MD simulated graphs, and are less likely to converge on RRGs. The reason for this is that although the calculations are very similar, the effect of adjusting the descendant voltage values, v_k, has a much bigger impact than adjusting the y_{jk} values; adjusting v_k changes $(c-1)+(c-1)^2$ currents, y_{jk} and y_{kl} for all neighbours of k, and for all descendant nodes, whereas adjusting y_{jk} changes $(c-1)$ current values. This results in a slow convergence as any small adjustment changes large sections of the network. We suggest that the algorithm works better on a MD topology due to its degree distribution.

Figure 6.5 gives a very simple example of a small network with loops. Although the algorithm is able to optimise the graph in this case, in a network this small where loops are obvious between all nodes, no assumption of independence between neighbours can be claimed. Therefore, the small graph is for demonstration purposes only. The example has been included to demonstrate how the difference in voltage magnitudes can result in the flow of current over an edge. The flow shown in the centre, is exactly the same as power flow calculated using the message passing algorithm with variables on y.

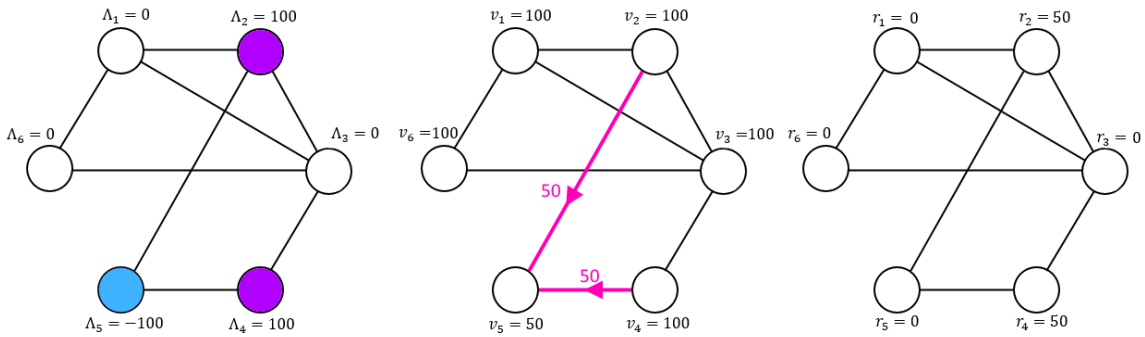


Figure 6.4 A very simple example of how current flows through a network according to the difference in voltages. Where capacities are given for each node on the left (generators are purple, consumer is blue), optimised voltages are shown (and their corresponding flow of current in pink) in the centre, and final reserve values according to optimised flow on the right.

One interesting behaviour is how adjusting the limits of v_{min} and v_{max} can result in unnecessary flow of current. The example here has limits of $v_{min} = 50$ for all nodes, $v_{max} = 75$ for node 6, and $v_{max} = 100$ for all other nodes. From Figure 6.4 we can see a network optimised without being restricted by voltage limits (in the case of no voltage limits, the current is identical but v values differ; here we limit voltages but this does not restrict the flow of current). The two neighbouring generators equally provide necessary current to node 5 and no extra current is sent. However, Figure 6.5 with a lower limit on node 6, results in its neighbours lowering their voltage to avoid having to send current they do not possess, but the generators who require their voltages at 100 to satisfy the consumer, cannot lower their voltages, and consequently send current to their already satisfied neighbours. This small example highlights the issue of having tight limits on variables and how this may increase distribution cost (assuming voltage is fixed per node and cannot be directional).

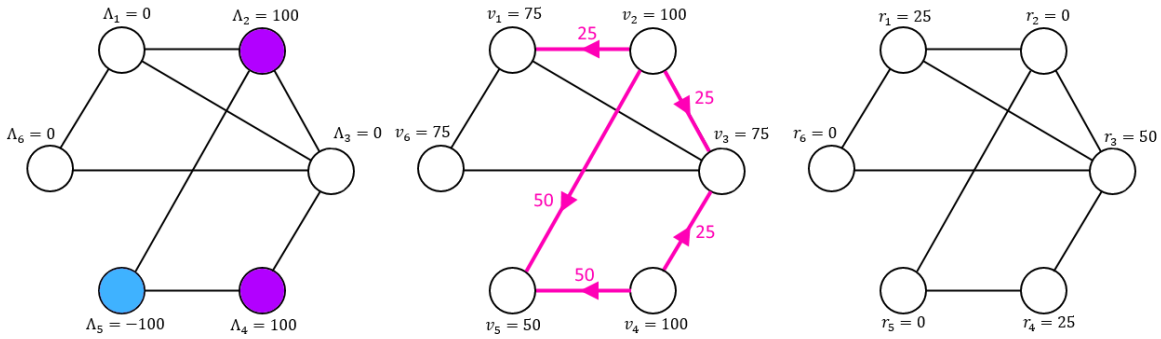


Figure 6.5 A very simple example of how a v_{max} limit can affect the power flow throughout a network. Where capacities are given for each node on the left (generators are purple, consumer is blue), optimised voltages are shown (and their corresponding flow of current in pink) in the centre when $v_{6max} = 75$, and the remaining v_{max} values are 100, all v_{min} values are 50, and final reserve values according to optimised flow on the right.

The effects of using the framework with voltages as control variables, when they are bounded between given limits can be seen in Figure 6.6; where the distribution of values v are plotted and it can be seen clearly how they are contained between the values v_{min} and v_{max} as the limits are increased. The distribution cost remains constant because the v values remain constant relative to each other.

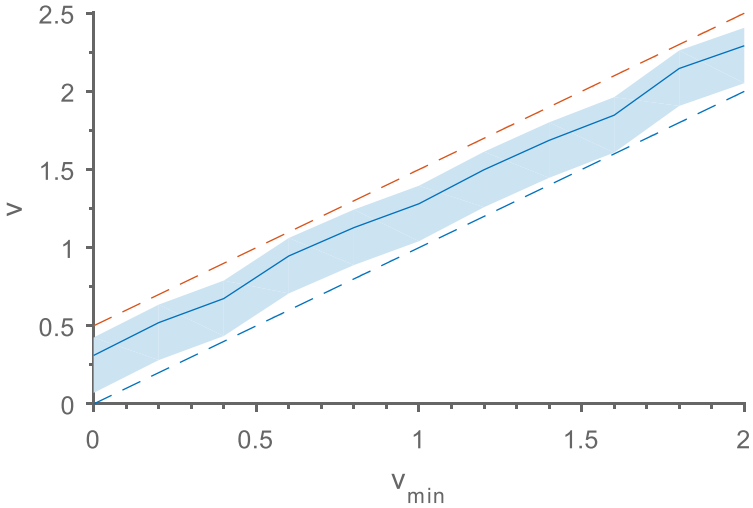


Figure 6.6 The distributed values of v over a network with limits of increasing v_{min} (blue dashed) and $v_{max} = v_{min} + 0.5$ (orange dashed).

Figure 6.7 gives a similar example to Figure 6.6, but where v_{max} is fixed. The graph shows how although the limits are large, the distribution of v remains small; this is because of the objective function which requires the difference between voltage values to be minimised. For every converged network, if the voltage is within the bounds, the v values can be anything, as long as they are correct relative to each other, this is why the mean value is very irregular for

increasing v_{min}; Figure 6.8 shows how for each convergence up to $v_{min} = 1.1$, even though the mean v values are varied, the distribution cost remains constant and minimised. At $v_{min} = 1.15$ the network is satisfied, but due to the limitations on v the distribution cost is increased. For any values higher than this, the network is unable to satisfy all consumers due to the tight constraints, until no power can be sent at $v_{min} = v_{max} = 1.5$ because there can be no difference in voltage.

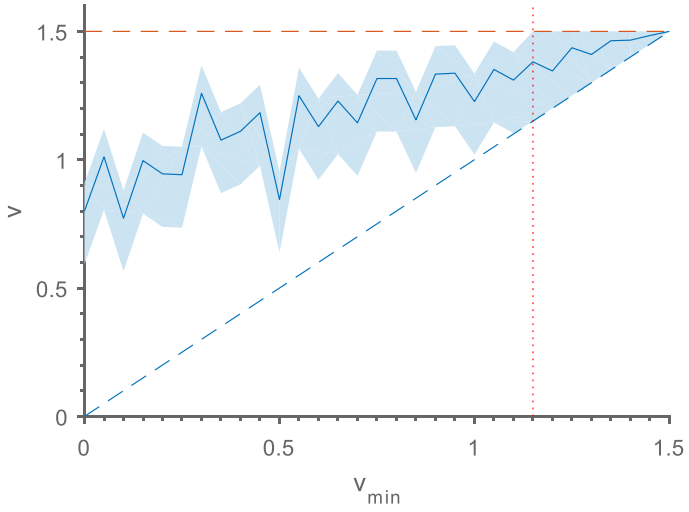


Figure 6.7 The distributed values of v over a network with limits of increasing v_{min} (blue dashed) and $v_{max} = 1.5$ (orange dashed). The red dotted line indicates when the algorithm was unable to satisfy all consumer nodes.

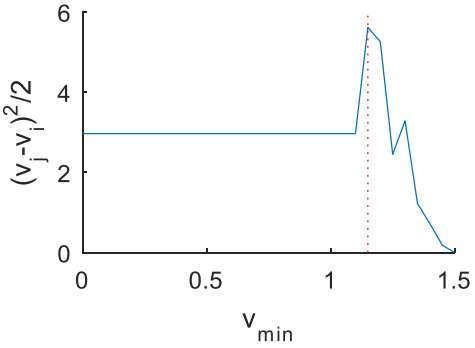


Figure 6.8 The distribution cost of the network as the limit of v_{min} increases and $v_{max} = 1.5$ as with Figure 6.7. The red dotted line indicates when the algorithm was unable to satisfy all consumer nodes (anything to the right has unsatisfied consumers).

6.5 Discussion

The ability for the message passing framework to consider voltages as control variables demonstrates the adaptability of the method and brings the algorithm closer to the criteria needed for an effective power distribution method. The work presented in this chapter is preliminary and serves as a proof of concept, to demonstrate the ability of the method to use different parameter sets.

We see from our results that the change in variables does not affect the minimisation of the objective function and we also see how the voltages can be limited between fixed values within this framework, but tight limitations result in the network being unable to find an optimal solution. The algorithm can minimise power loss when resistance is inhomogeneous, and we

include reactance within the constraints to highlight how the algorithm can be adapted to real-life electrical grids. Another point of interest is how strict limitations on voltages can in some cases result in unnecessary power sent to nodes with no demand.

Clearly, to make the algorithms more applicable one should add the many constraints and restrictions that exist in real systems. However, this is beyond the scope of this chapter which is aimed at demonstrating the potential of the approach to address more general settings.

7 Conclusion

The search for an effective ACOPF method is ongoing as no proposed method has been able to encompass the global optimisation of a non-convex, non-continuous, large network with multiple control variables and set parameters while adhering to the laws of physics, increasing reliability and considering robustness. The increasing penetration of fluctuating renewable sources such as wind and solar power have introduced a new component to the optimisation problem and although there have been a series of attempts to consider these uncertainties, they usually increase computational time, get stuck in local minima's, do not guarantee convergence or only deterministically consider the uncertainties through constraints. We therefore suggest a new distributed, computationally efficient and adaptive method which we believe will increase the reliability of the network.

The proposed method of message passing uses principled techniques to send conditional probabilities as messages locally in order to find the global optimal solution of the network. We recollect that the message passing method is able to pass continuous messages by sending the first and second derivatives of the VFE, and the VFE messages allow us to find the most probable, and optimal state for each node according to constraints and objective functions, within given approximations. Previous work has considered a distribution cost of $\phi = \frac{y^2}{2}$ for the objective function and bandwidths over edges. Comparing the message passing methods with the existing interior point method, we saw that message passing performed equally well in minimising the distribution cost, but found that the time taken was slightly more as the system size, and connectivity size increased. We suggest that with professional programming the message passing algorithm may be faster.

We have considered uncertainties and fluctuations at nodes by assuming capacities to be drawn from Gaussian distributions (any probability distribution can be used) and taking the average of the VFE. We find that the annealed approximation may only superficially consider the fluctuations, and that quenched averaging did not take into consideration a predetermined confidence level, so we suggested a quenched averaging method where the constraints included the predetermined confidence level (QAR). This method worked well and we found that by giving volatile nodes a higher reserve power the networks reliability in fluctuating time increased. We found that QAR gave a smoother change in power supplied as the uncertainties

increased but that the additional power increased the distribution cost and the number of active power lines. More work on incorporating volatility into message passing could include using a different probability distribution function from which capacities are drawn, and introducing bandwidths to the fluctuation calculations could be useful in preventing tripped power lines.

The thesis develops upon the message passing done by [76] by introducing the objective function of load shedding in the case where supply cannot meet demand; we focus on the brownout scenario, where any available power is shared amongst consumers and note that weighting of the the distribution cost should be negligible in this case, which can be achieved by introducing higher importance weights for consumers. Load shedding was done by minimising a variable, ζ_j, which denoted the amount of deficit at each node, according to its predetermined importance weight. We can limit the load shedding to consumers only by adding very large weights to generators, and suggest that weighting each consumer differently according to its importance allows us to prioritise load shedding. We point to the discontinuous jump in the deficit remaining at nodes between unsatisfiable and satisfiable networks, in terms of resources. We suggest that the advantage in this jump is an immediate reduction in the power supplied, which trades distribution costs with load shedding. Further work on this could include blackouts where some consumers are turned off altogether, keeping others fully satisfied, or incorporating rolling blackouts, which may require a step towards dynamic message passing.

A third objective function aimed to minimise the cost of generation. This is more important to power companies than reducing power losses because the monetary costs of generation are much higher. Three cost curves models were considered, linear for simplicity, negative quadratic to be inclusive of unit commitment decisions and positive quadratic to be most consistent with existing literature. We saw how linear and negative quadratic costs resulted in more generators being off or at full capacity, whereas a positive quadrative curve gave rise to more generators being on but not at full capacity. An additional minimal generator start-up value P_{min} was introduced, which required generators to produce a minimum capacity when turned on. We demonstrate how this can result in surplus reserves at consumer nodes when generators cannot be turned off. We suggest that further work could incorporate the unit commitment phase which makes decisions about whether generators should be turned on or off, this would be very useful for minimising overall costs for the entire network.

Finally, we demonstrated that message passing can optimise a voltage-based network rather than current/power-based flow. The algorithm is rearranged to make it more representative of

DCOPF by making voltage phase angles (described for simplicity in chapters as voltage) at each node the control variables and by showing how the algorithm can incorporate reactance (described as resistance). The algorithm can minimise distribution cost according to resistance over edges and we see how the difference in voltages returns similar results to the power flow used in the previous chapters, suggesting that message passing is able to accurately represent DCOPF in large networks. Limits on voltage values were also trialled and we saw that the algorithm effectively limits the voltage values. Further work on this would be to consider ACOPF equations without assumptions and to include a range of control variables. The complexity of changing from DCOPF to ACOPF may be significant as [35] suggests that "to switch to AC-based software would result in more precise dispatches and better marketing signals, but that the switch to AC-based software would increase the time to run a single scenario from minutes to over an hour". In terms of message passing, AC would require the use of more variables and more complex interactions.

In this thesis, we have found that message passing is a good method of distribution for power grid DCOPF as it is able to successfully consider multiple and varying objective functions, including constraints such as satisfaction levels, minimum and maximum dispatched power values and imposing variable limits over edges and nodes. The algorithm can be adapted to controlling variables on nodes instead of edges by using double variable bipartite factor graphs and resistance can be considered within the objective functions. Message passing has been shown to work as well as the interior point method in minimising distribution cost, the algorithm is able to minimise a 100-node network in an average of 35 seconds on a conventional laptop and the computational complexity is said to increase with $\mathcal{O}(N \ln N)$; although we do see that a higher connectivity increases the number of iterations needed to converge[13]. The method can inherently consider uncertainties within the calculations and we suggest that this can dramatically increase the reliability of networks with fluctuating variables.

Results show that the message passing method does appear to increase approximately linearly with the system size, which could result in much faster convergence than other methods such as Newton Raphson for large scale networks. The probabilistic nature has allowed us to incorporate fluctuations inherently within the calculations to accommodate for the stochastic nature of renewable sources making it arguably more reliable than other methods such as the

[13] This can be explained through the more specific order of $\mathcal{O}\left(N\left(\log\left(\frac{N}{k}\right)^2\right)\log(k)\right)$ from [192].

CCOPF method. The sparse topology of power grids allows us to find approximate global optimums, which for initial comparisons have been found to be exactly as optimal as the leading method of OPF, the interior point method, and the principled techniques allow us to guarantee a global optimal solution with few assumptions and more potential for further adaptation.

Disadvantages of the work done in this thesis include:

1. To consider volatility within the network, we model the node capacity as drawn from normal distributions of some mean. However, this allows to the fluctuations to flip the role of nodes from generator to consumer and vice versa, which in real-life may not be the case (unless microgenerators are considered to contribute to the network). Additionally, generators cannot produce infinite power, nor can consumers demand infinite levels of power, so the tails of a Gaussian are unrepresentative of real-life.
2. The message passing algorithm does not guarantee convergence to an optimal solution when the objective function is not quadratic and the number of iterations can be quite large on large scale networks.
3. The work done in this thesis does rely heavily on quadratic objective functions; this is because at higher orders or with exponential objective functions, the VFE would not be Gaussian, therefore the first and second derivatives would not be a precise representative of the VFE, and the solution would not be exact; although it may still provide a good approximate solution. Work done in [76] does consider linear and cubic objective functions, and other objective functions can also be used, but at the cost of accuracy.

However, the message passing appears to be very effective and easily adaptable. Here message passing methods have been used for power distribution but we would like to mention that they could also be used in other large network problems such as internet congestion, real traffic situations and other logistic scenarios.

8 Future Work

Finally, we would like to point to the potential of message passing equations by addressing additional challenges in the area of electricity grids. These could be future research directions using similar methodology to those described in this thesis.

8.1 Batteries

Electricity storage is becoming an important component in electricity grids that include volatile renewable components. Batteries and other storage devices are used to store power at times of plenty and release it when needed, thus mitigating the fluctuative nature of modern power generation. From a simple perspective, these could be described by a capacity that can be both negative and positive; when it is positive it acts as a generator but when the capacity is negative, the battery would be treated as a normal consumer that *must* be satisfied, contrary to the advantages of batteries, which allow them to be filled at off peak times or when generation exceeds predictions. Alternatively, the algorithm could make decisions on the state of the network and accordingly assign a certain amount of power to or from the battery according to its current reserves or demands. This could be done by the following constraint:

$$0 - \Lambda_j \leq \sum_{k \neq i} y_{jk} - y_{ij} \leq \beta_j - \Lambda_j, \tag{157}$$

where β_j is the maximum capacity of the battery, Λ_j indicates how much power the battery currently has stored (this should always be positive) and this constraint is only applied to the battery node. A corresponding production cost value can be added, similar to the generation cost G_j variable in minimising generation costs along with an objective function which encourages filling batteries at off-peak times. This constraint enforces that the power going in or out of the node stays within the battery limits, and when rearranged we can see that the left-hand side constraint is already enforced through the constraint in Equation (52); and the right-hand inequality is very similar to the constraint in Equation (115). This could be easily incorporated into the message passing algorithm. Power storage can be used as power sources when renewable sources are providing lower power levels than expected; however, this type of

inclusion of batteries requires consideration, mostly in the primary and secondary stage, once economic dispatch has optimised the distribution.

8.2 Deviation Costs

The provision of electrical power is highly regulated since it effects both industries and individuals, specifically machinery and appliances that are sensitive to voltage and current variation. An objective function could be designed so that deviation from the original values would be costly for the producer or provider. This would require the network to have y_{ij}^{old} values which indicate the power levels of each edge supplied in the previous 15 minutes. The algorithm could then create a positive quadratic where y_{ij}^{old} is the minima, discouraging large changes. To discourage small changes, a function such as $|\text{erf}(x)|$ could be used instead. A combination of minimising generations costs and deviation costs with weights allows he network to set generators who are expensive to deviate, as base levels, and others as more variable generators to be changed according to load; as seen in Figure 8.1 where deviation is more expensive at nuclear, coal and gas generators, and hydro is cheaper. This would be a large step towards making an effective, reliable method of power distribution for real-life networks.

Figure 8.1 An example of a typical days power demand in the US, and the proportion of power generated from different sources to satisfy the changes in demand; where natural gas has a high level of change, but nuclear remains smooth [170].

8.3 Correlations

With the increase in microgenerators and renewable sources in general, correlations are expected to be an important issue. For instance, wind farms, solar panels and combined heat and power generators are all conditioned on the weather. Although we know that wind and solar panels follow certain patterns according to the time of day, month or year, they all still depend on one joint set of variables – the weather. It is still under investigation whether there are correlations between forecasting errors of wind and solar generators; some work done by [171, 172] suggests that the error forecasts are weakly negatively correlated, and the correlations grow stronger with shorter time forecasts, larger geographical areas and in summer. Their study is done on such a large area in the US however, where the distance between nodes may be a lot larger than in the UK and so there may be differences in correlation properties. In addition, the proximity of neighbouring microgenerators may affect the network, along with increasing a negative correlation between consumer and generator values. It is expected that correlations between forecasting errors may be present, and therefore when considering fluctuations and uncertainty, to increase the reliability of economic dispatch it is inevitable that correlations should be considered. At the heart of the message passing method we assume a locally tree-like structure and this assumption, of all descendants being independent of one another if their common neighbour was removed, should remain for the derivation to be valid. However, correlations between neighbouring nodes could be accommodated in several ways, for instance, through joint covariance's when averages are carried out, or through the addition of non-power-bearing interactions between variables to represent existing correlations.

8.4 Preventing and Identifying Risk of Failure

Demand exceeding generation is not the only reason for power outages; faults on power lines are a regular cause of failures. Work done by Zdeborová et al [173] produce a stochastic local search algorithm called WalkGrid to develop a message passing scheme for identifying the transition areas between the network being satisfiable and unsatisfiable (the points where load shedding is required), and suggest mitigating against large scale network failures by switching within the network to ancillary lines (lines usually used only for backup or maintenance). This work uses the cavity method via population dynamics to obtain the macroscopic properties of network models without loops. Although WalkGrid performs slightly better than a BP-based

algorithm at finding an optimal configuration of switches on large networks, BP is able to count the number of satisfiable solutions, which can be used to accurately calculate how close the system is to failure; the two methods can be used together to calculate energy distribution and distance to failure predictions. Combining message passing and WalkGrid methods could be advantageous for economic dispatch, by highlighting risks for the controllers to mitigate. Alternatively, this could be used for network design, in order to find the most likely and worst case weaknesses of the network and find ways of strengthening them. Although work done by [174] suggested that some methods of reducing the likelihood of small outages can increase the risk of larger ones, highlighting these risks could be beneficial.

One way of minimising the risk of instability is that power flow equations are required to consider the $N-1$ contingency. This requires the network to continue to be satisfiable if any one of the nodes or edges are removed from the network, ensuring the system is able to continue to operate smoothly in the event of a sudden failure or attack. The $N-1$ contingency or "Security constrained optimal power flow" is not ideal because it assumes the probability of losing each node is the same, and also that the consequences of losing each node is the same. The method also, if continued to $N-2$ does not consider interdependencies between losses of nodes (e.g., deliberate attacks or local disasters). The message passing algorithm could incorporate this risk averse method by requiring as a condition that the network reserve must be equal to or higher than the highest capacity generator and that flow should be secured under realistic constraints. Alternatively, the network could assign risks to each node and minimise the satisfiability constraint according to them; the probabilistic nature of the algorithm may allow for correlations between risks to be considered.

8.5 Minimising Environmental Costs

One way in which linear minimisation costs (such as in Section 5.2) could be used is in minimising environmental costs. We have done work on a similar equation to the linear cost curve which suggests the introduction of a predetermined value E_j which correspond to the environmental damage caused by producing energy by a given generator; minimising the environmental cost can be done in addition to minimising other costs and does not increase the computational complexity of the algorithms (See Appendix E).

References

[1] EnerData, “Global energy statistical yearbook 2015,” 2015. [Online]. Available: https://yearbook.enerdata.net/. [Accessed 15 March 2016].

[2] Energy UK, “Electricity Generation,” 2010. [Online]. Available: http://www.energy-uk.org.uk/energy-industry/electricity-generation.html. [Accessed 15 March 2016].

[3] Department of Energy & Climate Change, “Renewable Sources of Energy,” National Statistics, London, 2016.

[4] Fossil Fuel, “How Natural Gas is Formed,” 2016. [Online]. Available: www.fossil-fuel.co.uk. [Accessed 15 March 2016].

[5] S. Shafiee and E. Topal, “When will fossil fuel reserves be diminished?,” *Energy Policy,* vol. 37, pp. 181-189, 2009.

[6] S. Fetter, “How long will the world's uranium supplies last?,” 2009. [Online]. Available: http://www.scientificamerican.com/article/how-long-will-global-uranium-deposits-last/. [Accessed 15 March 2016].

[7] World Nuclear Association, “Chernobyl Accident 1986,” 2015. [Online]. Available: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/chernobyl-accident.aspx. [Accessed 15 March 2016].

[8] World Nuclear Association, “Fukushima Accident,” 2016. [Online]. Available: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/fukushima-accident.aspx. [Accessed 15 March 2016].

[9] R. Black, “Nuclear power 'gets little public support worldwide',” BBC News, 25 November 2011. [Online]. Available: http://www.bbc.co.uk/news/science-environment-15864806. [Accessed 3 August 2016].

[10] W. Krewitt, F. Hurley, A. Trukenmüller and R. Friedrich, “Health Risks of Energy Systems,” *Risk Analysis,* vol. 18, no. 4, pp. 377-383, 1998.

[11] B. K. Sovacool, “A Critical Evalution of Nuclear Power and Renewable Electricity in Asia,” *Journal of Contemporary Asia,* no. 0047-2336, 2010.

[12] Department of Energy & Climate Change, “UK Solar PV Strategy Part 1: Roadmap to a Brighter Future,” 2013.

[13] M. Grimston, “Memorandum,” 2006. [Online]. Available: http://www.publications.parliament.uk/pa/cm200506/cmselect/cmenvaud/584/584we34.htm. [Accessed 15 March 2016].

[14] Department of Energy & Climate Change; The Rt Hon Edward Davey, “Record investments of £40 billion in renewable electricity to bring green jobs and growth to the UK,” Low Carbon Technologies, 2013.

[15] ReEnergy Holdings Lcc., “What is Biomass?,” [Online]. Available: www.reenergyholdings.com/renewable-energy/what-is-biomass/. [Accessed 20 November 2015].

[16] P. J. Crutzen, A. R. Mosier, K. A. Smith and W. Winiwarter, “N20 release from agro-biolfuel production negates global warming reduction by replacing fossil fuels,” *Atmospheric Chemistry and Physics,* vol. 8, 2008.

[17] Department of Energy & Climate Change; Charles Hendry, “UK and Iceland sign energy agreement,” Low Carbon Technologies, 2012.

[18] A. Withnall, “David Cameron to announce plan to power UK by harnessing Iceland's volcanoes,” October 2015. [Online]. [Accessed 15 March 2016].

[19] S. Shead, “Iceland's volcanoes could power the UK: but at what cost?,” April 2014. [Online]. Available: www.theengineer.co.uk/icelands-volcanoes-could-power-the-uk-but-at-what-cost/. [Accessed 15 March 2016].

[20] British Hydropower Association, “Hydro Facts,” 2011. [Online]. Available: www.british-hydro.org/hydro_facts.html. [Accessed 15 March 2016].

[21] Renewable UK, “Wind Energy,” 2016. [Online]. Available: www.renewableuk.com/en/renewable-energy/wind-energy/index.cfm. [Accessed 15 March 2016].

[22] D. Levitan, “Why Wave Power Has Lagged Far Behind as Energy Source,” 28 April 2014. [Online]. Available: http://e360.yale.edu/feature/why_wave_power_has_lagged_far_behind_as_energy_source/2760/. [Accessed 17 October 2016].

[23] Department of Energy & Climate Change, “About Us,” 2016.

[24] GHD, “Renewable energy – political will versus economic viability,” 18 August 2015. [Online]. Available: http://www.ghd.com/united_kingdom/about_us/blog/news_blog_2015/renewable_energy-political/. [Accessed 1 August 2016].

[25] Department for Business Innovation & Skills, “Microgeneration,” 2009. [Online]. Available: webarchive.nationalarchives.gov.uk/+/http:/www.berr.gov.uk/energy/sources/sustainable/microgeneration/index.html. [Accessed 15 March 2016].

[26] Department of Communities and Local Government, “Building a Greener Future: policy statement,” Department of Communities and Local Government, London, 2007.

[27] HM Government, “Government Response to the Committee on Climate Change,” Williams Lea Group, London, 2015.

[28] Department for Business Enterprise & Regulatory Reform, “Low Carbon Buildings Programme - Phase 2,” Department for Business Enterprise & Regulatory Reform, Watford, 2008.

[29] Electronics Tutorials, “DC Circuit Theory,” ApensCore, 1 January 2017. [Online]. Available: http://www.electronics-tutorials.ws/dccircuits/dcp_1.html. [Accessed 15 June 2017].

[30] National Grid, “What we do in the Electricity Industry,” 3 March 2016. [Online]. Available: http://www2.nationalgrid.com/uk/our-company/electricity/. [Accessed 9 May 2017].

[31] National Grid, “What we do,” 3 August 2016. [Online]. Available: http://www2.nationalgrid.com/about-us/what-we-do/. [Accessed 3 August 2016].

[32] National Grid, “Frequency response services,” 12 October 2016. [Online]. Available: http://www2.nationalgrid.com/uk/services/balancing-services/frequency-response/. [Accessed 17 October 2016].

[33] S. Blumsack, “Economic Dispatch and Operations of Electric Utilities,” Department of Energy and Mineral Engineering, Pennsylvania State University, 4 March 2016. [Online]. Available: https://www.e-education.psu.edu/eme801/node/532. [Accessed 9 May 2017].

[34] J. Carpentier, “Contribution á l'étude du dispatching économique,” *Bulletin de la Société Française des Électriciens,* vol. 8, no. 3, pp. 431-447, 1962.

[35] M. B. Cain, R. P. O'Neill and A. Castillo, “History of Optimal Power Flow and Formulations: Optimal Power Flow Paper 1,” 2012.

[36] D. K. Molzahn, B. C. Lesieutre and C. L. DeMarco, “A Sufficient Condition for Power Flow Insolvability with Applications to Voltage Stability Margins,” *IEEE Transactions on Power Systems,* vol. 28, no. 3, pp. 2592-2601, 2013.

[37] Power Optimisation, “Unit Commitment and Economic Dispatch Software to Optimise the Short-term Scheduling and Longer-term Planning of Electrical Power Generation,” [Online]. Available: http://www.powerop.co.uk/. [Accessed 3 August 2016].

[38] R. D. Zimmerman, C. E. Murillo-Sánchez and R. J. Thomas, “MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education,” *IEEE Transactions on Power Systems,* vol. 26, no. 1, pp. 12-19, 2011.

[39] A. Zhu and H. Pi, “A Method for Improving the Accuracy of Weather Forecasts Based on a Comprehensive Statistical Analysis of Historical Data for the Contiguous United States,” *Journal of Climatology & Weather Forecasting,* vol. 2, no. 1, pp. 1-9, 2014.

[40] A. Fabbri, T. G. S. Román, J. R. Abbad and V. H. M. Quezada, “Assessment of the Cost Associated With Wind Generation Prediction Errors in a Liberalized Electricity Market,” *IEEE Transactions on Power Systems,* vol. 20, no. 3, pp. 1440-1446, 2005.

[41] E. Gosden, “Blackouts report: death, disorder and other key consequences,” 2015. [Online]. Available: www.telegraph.co.uk/news/earth/energy/11314480/Blackouts-report-death-disorder-and-other-key-consequences.html. [Accessed 15 March 2016].

[42] J. Moylan, “Electricity blackouts would cause 'severe economic consequences',” 2014. [Online]. Available: www.bbc.co.uk/news/business-30221520. [Accessed 15 March 2016].

[43] European Commission, “Renewable energy progress report,” Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2015.

[44] M. T. Arif, A. M. T. Oo and A. B. M. S. Ali, “Estimation of Energy Storage and Its Feasibility Analysis,” *Energy Storage: Technologies and Applications,* 2013.

[45] The Learning Network, “Northeast Is Hit by Blackout,” 9 November 2011. [Online]. Available: http://learning.blogs.nytimes.com/2011/11/09/nov-9-1965-northeast-is-hit-by-blackout/?_r=0. [Accessed 9 November 2015].

[46] R. Cormier, “The 12 Biggest Electrical Blackouts In History,” 9 November 2015. [Online]. Available: http://mentalfloss.com/article/57769/12-biggest-electrical-blackouts-history. [Accessed 10 November 2015].

[47] J. Hooper, “Italy's blackout raises questions,” Guardian, 30 September 2003. [Online]. Available: http://www.theguardian.com/world/2003/sep/30/italy.johnhooper. [Accessed 10 November 2015].

[48] S. Z. Al-Mahmood, “Bangladesh Power Restored After Nationwide Blackout,” 2 November 2014. [Online]. Available: http://www.wsj.com/articles/bangladesh-power-restored-after-nationwide-blackout-1414915894. [Accessed 10 November 2015].

[49] RT, “Turkey struck by biggest power cut in 15 years, investigation underway (VIDEO),” 31 March 2015. [Online]. Available: https://www.rt.com/news/245529-massive-power-outage-turkey/. [Accessed 10 November 2015].

[50] Sky News, “Militant Attack Plunges Pakistan Into Darkness,” 26 January 2015. [Online]. Available: http://news.sky.com/story/1414477/militant-attack-plunges-pakistan-into-darkness. [Accessed 10 November 2015].

[51] Global Wind Energy Council, “Global Cumulative Installed Wind Capacity 1997-2014,” 2014. [Online]. Available: http://www.gwec.net/wp-content/uploads/2012/06/Global-Cumulative-Installed-Wind-Capacity-1997-2014.jpg. [Accessed 10 November 2015].

[52] *Newton-Raphson Power Flow [PowerPoint]*, GUC Faculty of Information Engineering and Technology.

[53] T. Tingting, L. Yang, L. Ying and J. Tong, “The Analysis of the Convergence of Newton-Raphson Method Based on the Current Injection in Distribution Network Case,” *International Journal of Computer and Electrical Engineering,* vol. 5, no. 3, pp. 288-290, 2013.

[54] H. Yang, F. Wen and L. Wang, “Newton-Raphson on Power Flow Algorithm and Broyden Method in the Distribution System,” in *International Conference on Power and Energy* , Johor Baharu, 2008.

[55] G. Chang, S. Chu and H. Wang, “An Improved Backward/Forward Sweep Load Flow Algorithm for Radial Distribution Systems,” *IEEE Transactions on Power Systems,* vol. 22, no. 2, pp. 882-884, 2007.

[56] B. Widrow and J. M. McCool, “A Comparison of Adaptive Algorithms Based on the Methods of Steepest Descent and Random Search,” *Transactions on Antennas and Propagation ,* vol. 24, no. 5, pp. 615-637, 1976.

[57] S. Haykin, “Least-Mean-Square Algorithm,” in *Neural Networks - a compresensive foundation,* Michigan, Macmillan, 1994, pp. 121-135.

[58] W.-M. Lin, C.-H. Huang and T.-S. Zhan, “A Hybrid Current-Power Optimal Power Flow Technique,” *IEEE Transactions on Power Systems,* vol. 23, no. 1, pp. 1-9, 2008.

[59] H. Wei, H. Saski, J. Kubokawa and R. Yokoyama, “An Interior Point Nonlinear Programming for Optimal Power Flow Problems with a Novel Data Structure,” *IEEE,* vol. 97, no. 0-7803-3717-1, pp. 134-141, 1997.

[60] F. Capitanescu, M. Glavic and L. Wehenkel, “An interior-point method based optimal power flow,” in *ACOMEN,* Belgium, 2005.

[61] F. M. Tuaimah and M. F. Meteb, “A Linear Programming Method Based Optimal Power Flow Problem for Iraqi Extra High Voltage Grid (EHV),” *Journal of Engineering,* vol. 20, no. 4, pp. 23-35, 2014.

[62] C. Coffrin and P. Van Hentenryck, “A Linear-Programming Approximation of AC Power Flows,” *INFORMS Journal on Computing,* vol. 26, no. 4, pp. 718-734, 2013.

[63] J. Lavei, A. Rantzer and S. Low, “Power flow optimization using positive quadratic programming,” in *IFAC Proceedings,* 2011.

[64] T. Bouktir, L. Slimani and M. Belkacemi, “A Genetic Algorithm for Solving the Optimal Power Flow Problem,” *Leonardo Journal of Science,* vol. 4, pp. 44-58, 2004.

[65] M. Younes, M. Rahli and L. Abdelhakem-Koridak, “Optimal Power Flow Based on Hybrid Genetic Algorithm,” *Journal of Information Science and Engineering,* vol. 23, pp. 1801-1816, 2007.

[66] D. B. Fogel, Evolutionary computation toward a new philosophy of machine intelligence., New York: IEEE Press, 1995.

[67] M. A. Abido, “Optimal power flow using particle swarm optimization,” *Electrical Power and Energy Systems,* vol. 24, no. 563-571, 2002.

[68] M. Li, W. Du and F. Nian, “An Adaptive Particle Swarm Optimization Algorithm Based on Directed Weighted Complex Network,” *Mathematical Problems in Engineering,* vol. 2014, 2014.

[69] T. Bouktir and L. Slimani, “Optimal Power Flow of the Algerian Electrical Network using an Ant Colony Optimization Method,” *Leonardo Journal of Sciences,* pp. 42-57, 2005.

[70] A. Rahiminejad, A. Alimardani and S. H. Hosseinian, “Shuffled frog leaping algorithm optimization for AC-DC optimal power flow dispatch,” *Turkish Journal of Electrical Engineering & Comupter Sciences,* vol. 22, pp. 874-892, 2014.

[71] T. Govindaraj and S. Udayakumar, “Optimal Reactive Power Planning And Real Power Loss Minimization Using Cuckoo Search Algorithm,” *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering,* vol. 2, no. 2, pp. 1-5, 2014.

[72] S. Orike and D. W. Corne, “Improved evolutionary algorithms for economic load dispatch optimization problems,” in *Workshop on Computational Intelligence* , UK, 2012.

[73] V. L. Paucar and M. J. Rider, “Artificial neural networks for solving the power flow problem in electrical power systems,” *Electric Power Systems Research,* vol. 62, no. 2, pp. 139-144, 2002.

[74] Beal, “Variational algorithms for approximate Bayesian inference,” University of London, London, 2003.

[75] K. Y. M. Wong and D. Saad, “Equilibration through local information exchange in networks,” *Physical Review E,* vol. 74, no. 010104, pp. 1-4, 2005.

[76] K. Y. M. Wong and D. Saad, “Inference and optimization of real edges on sparse graphs: A statistical pysics perspective,” *Physical Review E,* vol. 76, no. 011115, pp. 1-22, 2007.

[77] M. Mézard and A. Montanari, Information, Physics and Computation, New York: Oxford University Press Inc., 2009.

[78] Y. Kabashima, T. Murayama and D. Saad, “ Typical Performance of Gallager-type Error Correcting Codes,” *Physical Review Letters,* vol. 84, no. 6, pp. 1355-1358, 2000.

[79] M. Mézard and R. Zecchina, “Random k-satisfiability problem: From an analytic solution to an efficient algorithm,” *Physical Review E,* vol. 66, no. 056126, pp. 1-27, 2002.

[80] L. L. Peterson and B. S. Davie, Computer Networks: a systems approach, San Francisco: Morgan Kaufmann Publishers, 2003.

[81] Y. C. Ho, L. Servi and R. Suri, “A class of center-free resource allocation algorithms,” *Large Scale Systems,* vol. 1, no. 1, pp. 51-62, 1980.

[82] S. Shenker, D. Clark, D. Estrin and S. Herzog, “Pricing in computer networks: Reshaping the research agenda,” *ACM SIGCOMM Computer Communication Review,* vol. 26, no. 2, pp. 19-43, 1996.

[83] J. F. Dopazo and H. M. Merrill, “Optimal Generator Maintenance Scheduling Using Integer Programming,” *IEEE Transactions on Power Apparatus and Systems,* Vols. PAS-94, no. 5, pp. 1537-1545, 1975.

[84] J. Pearl, “Reverend Bayes on Inference Engines: A Distributed Hierachal Approach,” in *Proceedings of the Second National Conference on Artificial Intelligence,* California, 1982.

[85] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing, New York: Oxford University Press Inc., 2001.

[86] M. Mézard and A. Montanari, “Belief Propagation,” in *Information, Physics and Computation,* New York, Oxford University Press, 2009, pp. 292-326.

[87] K. Murphy, Y. Weiss and M. Jordan, “Loopy Belief Propagation for Approximate Inference: An Empirical Study,” in *Uncertainty in Artificial Intelligence,* Stockholm, 1999.

[88] D. MacKay and N. RM, “Near Shannon limit performance of low density parity check codes,” *Electronics Letters,* vol. 32, no. 18, pp. 1645-1646, 1996.

[89] C. Berrou, A. Glavieux and P. Thitimajshima, “Near Shannon Limit Error-correcting coding and decoding: Turbo-codes (1),” in *IEEE International Conference on Communications,* Geneva, 1993.

[90] J. Yedidia, W. Freeman and Weiss, “Understanding Belief Propagation and its Generalizations,” Mitsubishi Electric Research Laboratories, Massachusettes, 2001.

[91] J. Yedidia, W. Freeman and Y. Weiss, “Bethe free energy, Kikuchi approximations, and belief propagation algorithms,” Mitsubishi Electric Research Laboratories, Massachusettes, 2001.

[92] W. Wiegerinck and T. Heskes, “Fractional Belief Propagation,” in *Advances in Neural Information Processing Systems,* 2003.

[93] A. Yuille and A. Rangarajan, “The concave-convex procedure (CCCP),” in *Advances in Neural Information Processing Systems 2,* 2002.

[94] M. Mézard, P. G and R. Zecchina, “Analytic and Algorithmic Solution of Random Satisfiability Problems,” *Science,* vol. 297, no. 5582, pp. 812-815, 2002.

[95] A. Braunstein, M. Mézard and R. Zecchina, “Survey Propagation: An Algorithm for Satisfiability,” *Wiley Periodicals Inc.,* pp. 201-226, 4 March 2005.

[96] B. Frey and D. Dueck, “Clustering by Passing Messages Between Data Points,” *Science,* vol. 315, pp. 972-976, 2007.

[97] S. Mertens, M. Mézard and R. Zecchina, “Threshold Values of Random K-SAT from the Cavity Method,” *Random Structures and Algorithms,* vol. 28, no. 3, pp. 340-373, 2005.

[98] M. Opper and D. Saad, Advanced Mean Field Methods, MIT Press, 2001.

[99] D. J. C. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge: Cambridge University Press, 2003.

[100] L. Saul and M. Jordan, “Learning in Boltzmann Trees,” *Neural Computation,* vol. 6, no. 6, pp. 1174-1184, 1994.

[101] S. L. Lauritzen, “Propagation of Probabilities, Means, and Variances in Mixed Graphical Association Models,” *Journal of the American Statistical Association,* vol. 87, p. 1098, 1992.

[102] N. Skantzos, I. P. Castillo and J. P. L. Hatchett, “Cavity approach for real variables on diluted graphs and application to synchronization in small-world lattices,” *Physical Review E,* vol. 72, no. 6, p. 066127, 2005.

[103] T. Minka, “Expectation Propagation for Approximate Bayesian Inference,” in *Uncertinty in Artificial Intelligence,* Washington, 2001.

[104] Y. Weiss and W. Freeman, “Correctness of belief propagation in Gaussian graphical models of arbitrary topology,” Mitsubishi Electric Research Laboratories, Inc., Massachusettes, 1999.

[105] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin and J. Hellerstein, “GraphLab: A New Framework for Parallel Machine Learning,” in *Uncertainty and Artificial Intelligence,* California, 2010.

[106] J. Moharil, P. May and D. B. H. Gaile, “Belief propagation in genotype-phenotype networks,” *Statistical Applications in Genetics and Molecular Biology,* vol. 15, no. 1, pp. 39-53, 2016.

[107] K. Y. M. Wong, D. Saad and Z. Gao, “Message passing for task redistribution on sparse graphs,” *Advances in Neural Information Processing Systems*, vol. 18, p. 1529, 2005.

[108] K. Y. M. Wong, D. Saad and Z. Gao, “Resource allocation in sparse graphs,” in *European Conference on Complex Systems*, Paris, 2005.

[109] C. H. Yeung and K. Y. M. Wong, “Optimal Resource Allocation in Random Networks with Transportation Bandwidths,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 1, p. P03029, 2009.

[110] M. Collas, Resource Allocation on Sparse Graphs, Aston University, 2005.

[111] E. Harrison, D. Saad and K. Y. M. Wong, “Optimal Load Shedding in Electricity Grids with Renewable Sources via Message Passing,” *Energy Procedia*, vol. 107, pp. 101-108, 2016.

[112] Department of Energy & Climate Change, “Supply and consumption of electricity,” Energy Trends: electricity, 2015.

[113] B.-M. Hodge, A. Florita, K. Orwig, D. Lew and M. Milligan, “A comparison of wind power and load forecasting error distributions,” *National Renewable Energy Laboratory*, pp. 1-8, 2012.

[114] Department of Energy and Climate Change, Energy Consumption in the UK, London: Crown Copyright, 2015, pp. 8, 14-16.

[115] Department of Energy and Climate Change, “Energy consumption in the UK 2016,” Department of Business, Energy and Industrial Strategy, 2016.

[116] J. Miller, “Why expanded alternative energy increases the need for natural gases,” The Energy Collective, 1 January 2013. [Online]. Available: http://www.theenergycollective.com/jemillerep/178096/expanded-wind-and-solar-power-increase-need-natural-gas. [Accessed 2017 April 3].

[117] Committee on Climate Change, “The Renewable Energy Review,” 2011.

[118] Ofgem, “About RO,” Ofgem, 5 Novemer 2016. [Online]. Available: https://www.ofgem.gov.uk/environmental-programmes/ro/about-ro. [Accessed 7 December 2016].

[119] The Renewable Energy Website, “Wind Speed Distribution Weibull,” 2014. [Online]. Available: http://www.reuk.co.uk/Wind-Speed-Distribution-Weibull.htm. [Accessed 20 November 2015].

[120] B.-M. Hodge and M. Milligan, “Wind Power Forecasting Error Distributions over Multiple Timescales,” Michigan, 2011.

[121] M. Lange, “On the Uncertainty of Wind Power Predictions - Analysis of the Forecast Accuracy and Statistical Distribution of Errors,” *Journal of Solar Energy Engineering*, pp. 1-8, 2005.

[122] Y. V. Makarov, Z. Huang, P. V. Etingov and J. Ma, “Incorporating Wind Generation and Load Forecast Uncertainties into Power Grid Operations,” *U.S. Department of Energy*, 2010.

[123] F. Massen, A. Kies, N. Harpes and Students of the LCD, “Seasonal and Diurnal CO2 Patterns at Diekirch LU, 2003-2005,” MeteoLCD, 2007.

[124] RenSmart Wind Archive, “The average wind speed for each month over the period 2000 to 2010 - Birmingham/Airport,” 2010. [Online]. Available: http://www.rensmart.com/Weather/WindArchive\#monthlyLayer. [Accessed 20 November 2015].

[125] RenSmart Wind Archive, “The distribution of wind speed over the period 2000 to 2010 - Birmingham/Airport,” 2010. [Online]. Available: http://www.rensmart.com/Weather/WindArchive\#monthlyLayer. [Accessed 20 November 2015].

[126] G. Goodstal, Electrical Theory for Renewable Energy, New York: Delmar Cengage Learning, 2013, p. 183.

[127] D. Lew and M. Milligan, “The value of wind power forecasting,” National Renewable Energy Laboratory, 2011.

[128] Wind Power Program, “Wind statistics and the Weibull distribution,” [Online]. Available: http://www.wind-power-program.com/wind_statistics.htm. [Accessed 20 November 2015].

[129] K. Methaprayoon, W. J. Lee, C. Yingvivatanapong and J. Liao, “An Integration of ANN Wind Power Estimation into UC Considering the Forecasting Uncertainty,” *IEEE*, pp. 116-119, 2005.

[130] V. S. Pappala, I. Erlich, K. Rohrig and J. Dobschinski, “A Stochastic Model for the Optimal Operation of a Wind-Thermal Power System,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 940-950, 2009.

[131] E. D. Castronuovo and J. A. Peças Lopes, "On the Optimization of the Daily Operation of a Wind-Hydro Power Plant," *IEEE Transactions on Power Systems,* vol. 19, no. 3, pp. 1599-1606, 2004.

[132] R. Doherty and M. O'Malley, "A New Approach to Quantify Reserve Demand in Systems with Significant Installed Wind Capacity," *IEEE Transactions on Power Systems,* vol. 20, no. 2, pp. 587-595, 2005.

[133] S. Bofinger, A. Luig and G. Beyer, "Qualification of wind power forecasts," pp. 1-6, 2002.

[134] H. Bludszuweit, J. A. Domínguez-Navarro and A. Llombart, "Statistical Analysis of Wind Power Forecast Error," *IEEE Transactions on Power Systems,* vol. 23, no. 3, pp. 983-991, 2008.

[135] S. Tewari, C. J. Geyer and N. Mohan, "A Statistical Model for Wind Power Forecast Error and its Application to the Estimation of Penalties in Liberalized Markets," *IEEE Transactions on Power Systems,* vol. 26, no. 4, pp. 2031-2039, 2011.

[136] Slide Share, "Offshore wind farm design - Offshore wind climate," [Online]. Available: http://www.slideshare.net/DelftOpenEr/offshore-wind-farm-design-offshore-wind-climate-13773950. [Accessed 20 November 2015].

[137] L. E. Jones, Renewable Energy Integration, 1 ed., London: Elsevier Inc., 2014.

[138] Y. D. Arthur and K. B. Gyamfi, "Probability Distributional Analysis of Hourly Solar Irradiation in Kumasi-Ghana," *The Journal of Business,* vol. 3, no. 3, pp. 63-75, 2013.

[139] J. Harrison, "Solar Power," 2015. [Online]. Available: http://www.jaharrison.me.uk/Misc/Solar/index.html. [Accessed 20 November 2015].

[140] Met Office, "Met Office in the Media," 2011. [Online]. Available: http://blog.metoffice.gov.uk/2011/06/23/met-office-in-the-media-23-june-2011/. [Accessed 20 November 2015].

[141] B.-M. Hodge, M. Hummon and K. Orwig, "Solar Ramping Distributions over Multiple Timescales and Weather Patterns," *National Renewable Energy Laboratory,* 2011.

[142] S. Mathew, G. S. Philip and C. M. Lim, "Analysis of Wind Regimes and Performance of Wind Turbines," *Advances in Wind Energy Conversion Technology,* pp. 71-83, 2011.

[143] D. Beinstock, M. Chertkov and S. Harnett, “Chance Constrained Optimal Power Flow: Risk-Aware Network Control Under Uncertainty,” 2013.

[144] A. Nemirovski and A. Shapiro, “Convex Approximations of Change Constrained Programs,” *Society for Industrial and Applied Mathematics Journal Optimisation,* vol. 17, no. 4, pp. 969-996, 2006.

[145] R. Bent, D. Beinstock and M. Chertkov, “Synchronization-Aware and Algorithm-Efficient Chance Constrained Optimal Power Flow,” *IREP Symposium-Bulk Power System Dynamics and Control -IX,* 2013.

[146] G.-C. Liao, “A novel evolutionary algorithm for dynamic economic dispatch with energy saving and emission reduction in power system integrated wind power,” *Energy,* vol. 36, pp. 1018-1029, 2011.

[147] S.-Y. Lin and J.-F. Chen, “Distributed optimal power flow for smart grid transmission system with renewable energy sources,” *Energy,* vol. 56, pp. 184-192, 2013.

[148] L. Armijo, “Minimization of functions having Lipschitz continuous first partial derivatives,” *Pacific Journal of Mathematics,* vol. 16, no. 1, pp. 1-3, 1966.

[149] S. Brini, H. H. Abdallah and A. Ouali, “Economic Dispatch for Power System included Wind and Solar Thermal energy,” *Leonardo Journal of Sciences,* vol. 14, pp. 204-220, 2009.

[150] G. J. Osório, J. M. Lujano-Rojas, J. C. O. Matias and J. P. S. Catalao, “A Probabilistic Approach to Solve the Economic Dispatch Problem with Intermittent Renewable Energy Sources,” *Energy,* vol. 82, pp. 949-959, 2015.

[151] B.-F. B, E. Farjah and R. Azizipanah-Abarghooee, “An efficient scenario-based and fuzzy self-adaptive learning particle swarm optimization approach for dynamic economic dispatch consideing load and wind power uncertainties,” *Energy,* vol. 50, no. 1, pp. 232-244, 2013.

[152] M. Schwartz, “Relation between quenched and annealed random ferromagnetic systems,” *Physics Letters,* vol. 72A, no. 1, pp. 37-40, 1979.

[153] The Illinois Center for a Smarter Electric Grid (ICSEG), “WSCC 9-Bus System,” 2015.

[154] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, B. R, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour and C. Singh, “The IEEE Reliability Test System-1996.

A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee," *IEEE Transactions on Power Systems,* vol. 14, no. 3, pp. 1010-1020, 1999.

[155] K. Wong and D. Saad, "Minimizing unsatisfaction in colourful neighbourhoods," *Journal of Physics A,* vol. 41, no. 32, p. 324023, 2008.

[156] R. Christie, "IEEE 118 Bus Test Case," University of Washington, 2009.

[157] E. Harrison, D. Saad and K. Wong, "Optimal distribution in smart grids with volatile renewable sources using a message passing algorithm," *International Journal of Smart Grid and Clean Energy,* vol. 5, no. 4, pp. 221-228, 2016.

[158] E. Harrison, D. Saad and K. Wong, "Power Grids with Volatile Sources - Message Passing for Optimising Electricity Distribution and Load Shedding," in *2nd International Conference on Offshore Renewable Energy*, Glasgow, 2016.

[159] E. Harrison, D. Saad and K. Wong, "Message Passing for Distributed Optimisation of Power Allocation with Renewable Resources," in *2nd International Conference on Intelligent Green Building and Smart Grid*, Prague, 2016.

[160] M. Rohden, A. Sorge and D. Witthaut, "Self-Organized Synchronization in Decentralized Power Grids," *Physical Review Letters,* vol. 109, p. 064101, 2012.

[161] C. Wu, G. Hug and S. Kar, "Risk-limiting Economic Dispatch for Electricity Markets with Flexible Ramping Products," *IEEE Transactions on Power Systems,* vol. 31, no. 3, pp. 1990-2003, 2015.

[162] White House, "Incorporating renewables into the electricity grid: Expanding opportunites for smart markets and energy storage," Executive Office of the President of the United States, Washington, 2016.

[163] Eskom, "What is load shedding?," Eskom, 8 November 2016. [Online]. Available: http://loadshedding.eskom.co.za/loadshedding/description. [Accessed 8 December 2016].

[164] N. Tatbul, U. Çetintemel and S. Zdonik, "Staying FIT: Efficient Load Shedding Techniques for Distributed Stream Processing," in *International Conference on Very Large Data Bases*, Vienna, 2007.

[165] K. Y. M. Wong, C. H. Yeung and D. Saad, "Distributed algorithms for global optimization on sparse networks of arbitrary bandwidths," Aston University, Birmingham, 2006.

[166] S. Kaplan, “CRS Report for Congress,” 13 November 2008. [Online]. Available: https://www.fas.org/sgp/crs/misc/RL34746.pdf. [Accessed 10 November 2016].

[167] L. V. N. Rao, “PSO Technique for Solving the Economic Dispatch Problem Considering Generator Constraints,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* vol. 3, no. 7, pp. 10439-10454, 2014.

[168] Z.-L. Giang, “Particle Swarm Optimization to Solving the Economic Dispatch Considering the Generator Constraints,” *IEEE Transactions on Power Systems,* vol. 18, no. 3, pp. 1187-1195, 2003.

[169] P. Jain, K. K. Swarnkar, D. Wadhwani and A. K. Wadhwani, “Prohibited Operating Zones Constraint with Economic Load Dispatch using Genetic Algorithm,” *International Journal of Engineering and Innovative Technology ,* vol. 1, no. 3, pp. 179-183, 2012.

[170] C. Arcus, “For Maximum Renewable Integration, Load Following is King,” Clean Technica, 5 July 2016. [Online]. Available: https://cleantechnica.com/2016/07/05/maximum-renewable-integration-load-following-king/. [Accessed 2017 May 22].

[171] J. Zhang, B.-M. Hodge and A. Florita, “Investigating the Correlation Between Wind and Solar Power Forecast Errors in the Western Interconnection,” in *International Conference on Energy Sustainability ,* Minnesota, 2013.

[172] J. Zhang, B.-M. Hodge and A. Florita, “Joint Probability Distribution of Correlation Analysis of Wind and Solar Power Forecast Errors in the Western Interconnection,” in *American Society of Civil Engineers,* 2014.

[173] L. Zdeborová, A. Decelle and M. Chertkov, “Message passing for optimization and control of a power grid: Model of a distribution system with redundancy,” *Physical Review,* vol. 80, no. 046112, 2009.

[174] I. Dobson, B. A. Carreras, V. E. Lynch and D. E. Newman, “Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization,” *Chaos,* vol. 17, no. 026103, pp. 1-13, 2007.

[175] A. Pagani and M. Aiello, “The Power Grid as a Complex Network: a Survey,” *Physics and Society,* vol. 1105.3338, no. 2, p. 34, 2012.

[176] P. Erdos and A. Renyi, “On Random Graphs I,” *Publicationes Mathematicae,* vol. 6, no. 1, pp. 290-297, 1959.

[177] E. Gilbert, “Random Graphs,” *The Annals of Mathematical Statistics,* vol. 30, no. 4, pp. 1141-1144, 1959.

[178] D. Watts and S. Strogatz, “Collective Dynamics of 'small-world' networks,” *Nature,* vol. 393, no. 6684, pp. 440-442, 1998.

[179] B. Cloteaux, “Limits in modelling Power Grids Topology,” 2013.

[180] A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” *Science,* vol. 286, pp. 509-512, 1999.

[181] P. Hines, S. Blumsack, E. Cotilla Sanchez and C. Barrows, “The Topological and Electrical Structure of Power Grids,” 2010.

[182] B. Bollobás, W. Fulton, A. Katok, F. Kirwan and P. Sarnak, Random Graphs, 1 ed., vol. 2, Cambridge: Cambridge University Press, 2001, pp. 50-55.

[183] D. Chassin and C. Posse, “Evaluating North American Electrica Grid Reliability Using the Barabási-Albert Network Model,” *Physica A: A Statistical Mechanics and its Applications,* vol. 355, no. 2-4, pp. 667-677, 2005.

[184] L. Amaral, A. Scala, M. Barthémy and H. Stanley, “Classes of small-world networks,” *Proceedings of the National Academy of Sciences,* vol. 97, no. 21, pp. 11149-11152, 2000.

[185] R. Albert, I. Albert and G. Nakarado, “Structural vulnerability of the North American power grid,” *Physical Review E,* vol. 69, no. 1, p. 025103, 2004.

[186] R. Albert and A. Barábasi, “Statistical Mechanics of Complex Networks,” *Reviews on Modern Phsyics,* vol. 74, no. 1, pp. 47-97, 2002.

[187] B. Bollobás and W. Fernandez de la Vega, “The Diameter of Random Regular Graphs,” *Combinatorica,* vol. 2, no. 2, pp. 125-134, 1982.

[188] S. Pahwa, A. Hodges, C. Scoglio and S. Wood, “Topological Analysis of the Power Grid and Mitigation Strategies Against Cascading,” 2010.

[189] M. Newman, “Assortative mixing in networks,” *Physical Review Letters,* vol. 89, no. 208701, 2002.

[190] PSCAD, “IEEE 39 Bus System,” IEEE, 6 December 2016. [Online]. Available: https://hvdc.ca/knowledge-base/read,article/28/ieee-39-bus-system/v:. [Accessed 7 December 2016].

[191] C. Lo, “Sun storage: the quest for 24-hour solar power,” 2014. [Online]. Available: www.power-technology.com/features/featuresun-storage-the-quest-for-24-hour-solar-power-4168692/. [Accessed 15 March 2016].

[192] M. McAsey, L. Mou and W. Han, “Convergence of the forward-backward sweep method in optimal control,” *Computational Optimization Applications,* vol. 53, no. 1, pp. 207-226, 2012.

[193] V. Chandar, D. Shah and G. Wornell, “A Simple Message-Passing Algorithm for Compressed Sensing,” in *International Symposium on Information Theory*, Austin, Texas, 2010.

Appendices

A.Fundamental Concepts in Statistical Physics

Since the message passing approach utilised in this thesis has been inspired by concepts adopted from statistical physics, this appendix looks to introduce some of the basic concepts and terminology that appears throughout the thesis.

Statistical physics typically comprises of many variables that interact with one another through predefined or dynamically changing interactions. This set of variable configurations or states of the system appear with some probability. The sum of all probabilities of being in a certain state, s_i, is equal to one:

$$\sum_i P_i = 1, \tag{158}$$

where P_i is the probability of the system being in state, s_i, and the sum runs over all possible states. Each state has an energy, E_i, and the expected energy, E, can be written as:

$$\sum_i P_i E_i = E. \tag{159}$$

The entropy of the system is defined as:

$$S = -\sum_i P_i \log P_i. \tag{160}$$

The entropy is a measure of uncertainty of the state s_i. The larger the entropy, the less prior information is known about the node being in state s_i.
To maximise the entropy subject to the constraints of probability, we can write the function using two Lagrange multipliers α and β:

$$S + \alpha\left(\sum_i P_i - 1\right) + \beta\left(\sum_i P_i E_i - E\right), \tag{161}$$

and optimising this with respect to P_i gives:

$$\log P_i + 1 + \alpha + \beta E_i = 0, \tag{162}$$

or:

$$P_i = e^{-(1+\alpha)-\beta E_i}. \tag{163}$$

If we then rename the function $e^{1+\alpha} = Z$, the partition function, which is just a normalisation constant, this can be rewritten as:

$$P_i = \frac{1}{Z} e^{-\beta E_i}. \tag{164}$$

This is called the Boltzmann distribution, it is the equilibrium distribution, or the maximum entropy subject to the constraints.

Considering that the sum of all probabilities equals one, from Equation (158), and (164) we find the equation for the partition function, and its derivative:

$$Z(\beta) = \sum_i e^{-\beta E_i}, \tag{165}$$

$$\frac{\partial Z}{\partial \beta} = -\sum_i E_i e^{-\beta E_i}. \tag{166}$$

Considering the entropy in terms of probabilities, in Equations (160), and (164), one can rearrange the equations to get:

$$S = \beta E + \log Z\,(\beta)\,. \tag{167}$$

Using Equations (164) and the energy in terms of the energy of each state multiplied by the probability of each state, (159), we get:

$$E = \frac{1}{Z} \sum_i E_i e^{-E_i}, \tag{168}$$

$$E = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \log Z}{\partial \beta}. \tag{169}$$

The energy is made up of two main components; the free energy, F, and the entropy multiplied by some temperature, T:

$$E = F + TS, \tag{170}$$

where the free energy is the freedom of the state to change; minimising this will find the most likely, or optimal set of states of system. Differentiating with respect to entropy gives:

$$\frac{dE}{dS} = T. \tag{171}$$

We also write the free energy as:

$$F = -T \log Z\,; \tag{172}$$

from which most macroscopic properties of the system at equilibrium can be derived. By fully differentiating Equation (167):

$$dS = \beta dE + Ed\beta + \frac{d \log Z}{d\beta} d\beta. \tag{173}$$

From Equation (169) and then (171) this gives:

$$dS = \beta dE \quad \text{or} \quad \beta = \frac{dS}{dE} = \frac{1}{T}. \tag{174}$$

We can now use these equations of the energy, free energy, partition function and Boltzmann distribution to find a recursive relation in terms of a function called the vertex free energy, which can then be passed iteratively as messages through a system to obtain an optimal solution according to some objective functions.

B. Electricity Grid Topology

To test and analyse message passing algorithms we need to generate multiple networks of different sizes which are representative of the Power Grid (PG) topology and its main characteristics. These are required for generating statistical properties of networks and measure the performance of the devised algorithms. Although there are IEEE models[14] available, these do not give us a large enough variety for analysis, controlled evaluation and adequate statistics. There are many different types of graph and methods for creating them. I will be looking at the main and most relevant methods according to literature on modelling the PG topology [175].

Methods of Network Generation

Some of the models suggested in literature are regular lattice, random graph [176, 177], random regular graph, small-world [178], Cloteaux's graph [179], Barabási-Albert's preferential attachment [180] and Hines' Minimum-Distance method [181] generated graphs.

Regular lattices (RL) are very organised structures, each node is connected to c other nodes (the neighbourhood or degree), c is the same for every node[15] and nodes are connected in a sequence. Some examples of a random lattice can be seen in Figure B.0.1 and Figure B.0.2. They are created by positioning the nodes and then linking them according to some pattern. These graphs are easy to produce and fast to compute, but they are unrealistic compared to the complex structure of a PG.

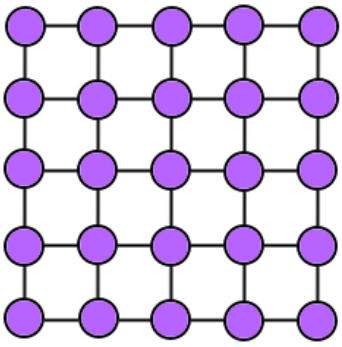


Figure B.0.1 Example of a square lattice.

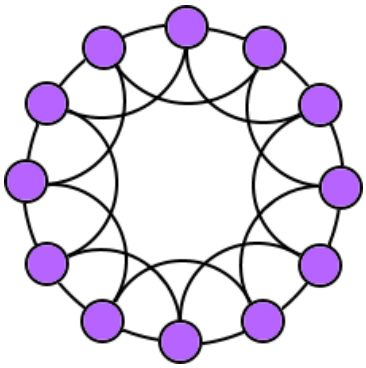


Figure B.0.2 Example of a ring lattice.

[14] IEEE provides synthetic grids which accurately represent real PG networks.

[15] In the case of looped, infinite or ring lattices c is constant. In Figure B.0.1 we see that all the inner nodes have fixed connectivity but outer nodes can have a lower connectivity

Random Graphs (RG) are unstructured and are generated by connecting randomly selected nodes. One of the most celebrated graphs is the Erdös-Renyi [176] graph $G(N, v)$; this is created by choosing, at random, one of all the possible graphs with N nodes and v edges. In the $G(N, p)$ model suggested by Gilbert [177], a graph is generated by selecting, at random, edges between nodes and creating them with probability $p = \frac{v}{N}$, independently. As described by Erdös and Renyi, with random graphs there is a probability of the graph not being completely connected. This may look similar to the topology of a PG, but lacks control and the structure that emerges from landscape and demand characteristics. PG networks are developed over time and placed according to geographical data, population density, overloaded edges, connectivity of nodes, etc. Although we do not need this level of accuracy, we do not want to ignore the control aspect completely.

Random Regular Graphs (RRG) [182] are similar to random graphs, but each node has a fixed connectivity c, where $c \geq 3$. This is less trivial to generate. There are many ways to generate a random regular graph; one method is the pairing model which takes Nc nodes and places them into N sections. Random matching between the Nc nodes is done such that each node is connected to a node from another section and edges are formed; then each section is made into a single vertex, with c connections. If there are repeated edges present, the process is repeated.

This is a very easy model to produce and quick to compute. It can be argued that the fixed connectivity makes it less representative of PGs than random graphs, but it can be systematically controlled. It is a good model for analysis due to its limiting variables.

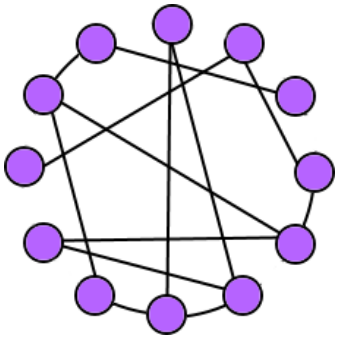


Figure B.0.3 Example of a Random Graph.

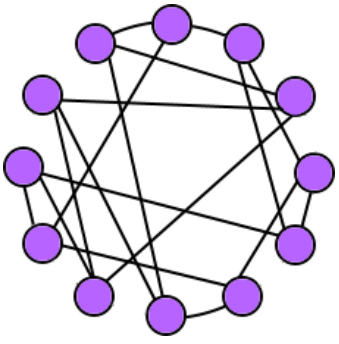


Figure B.0.4 Example of a Regular Random Graph where $c = 3$.

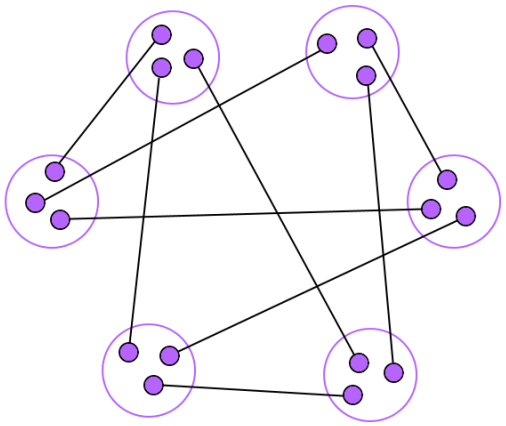


Figure B.0.5 Generating a RRG using the pairing method.

Small-World (SW) graphs [178] can be explained as the middle ground between random graphs and regular lattices. For instance, they can be created by first considering a ring lattice; then each edge is rewired to any node with probability p. If $p = 0$ then the graph is a regular lattice and if $p = 1$ the graph is an Erdös-Renyi random graph. ($N \gg c \gg \ln(N) \gg 1$ guarantees that the graph will be completely connected). According to [175] the transmission grid of a PG has been found in many studies to satisfy the SW conditions, while it is less representative of the distribution stage topology (this work will focus specifically on the transmission stage).

Like RRGs we can have more control on the resulting topology, but with an element of randomness. This topology is amenable for analysis as we can plot results against the increasing p.

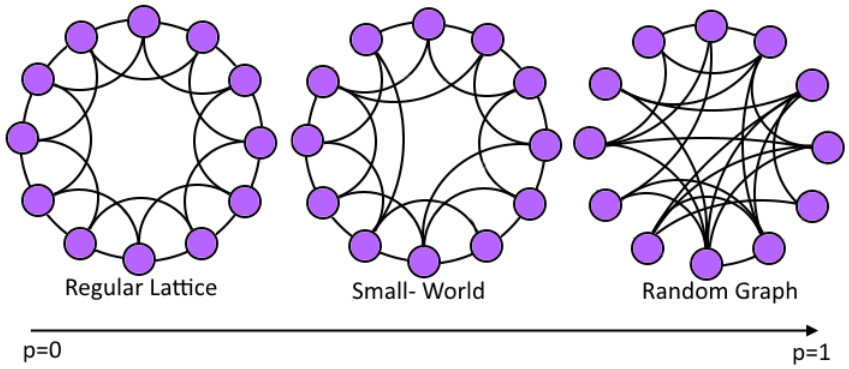


Figure B.0.6 Small-World example in relation to regular and Erdös-Renyi random graphs.

Preferential Attachment is another common method of generating graphs. We consider the method of Barabási-Albert (BA) [180] for generating them. This is done by starting with an

initially connected graph[16], denoted as m_0. New nodes are added to the network one at a time. Each new node is connected to each existing node, i, with a probability, p_i, that is proportional to the number of links that the existing node already has. This can be formally written as:

$$p_i = \frac{c_i}{\sum_j c_j}, \tag{175}$$

where c_i is the number of links that node i already has and the sum at the denominator is over all the pre-existing nodes j. This means that heavily linked nodes will accumulate even more links while nodes with few links are unlikely to be chosen.

This is a good model for growth of political or social networks, for instance the more someone is known the more likely you are to know them. However, for PGs this may not be an optimal representation because network designers may be more likely to add edges to less connected nodes for obtaining a more stable, robust and secure network (if a network contains only a few nodes which are very highly connected this makes it more vulnerable to directed attack). An adaptation to this model to describe a PG network would be to use a minimal spanning tree[17] over a geographical area as the initially connected graph, m_0.

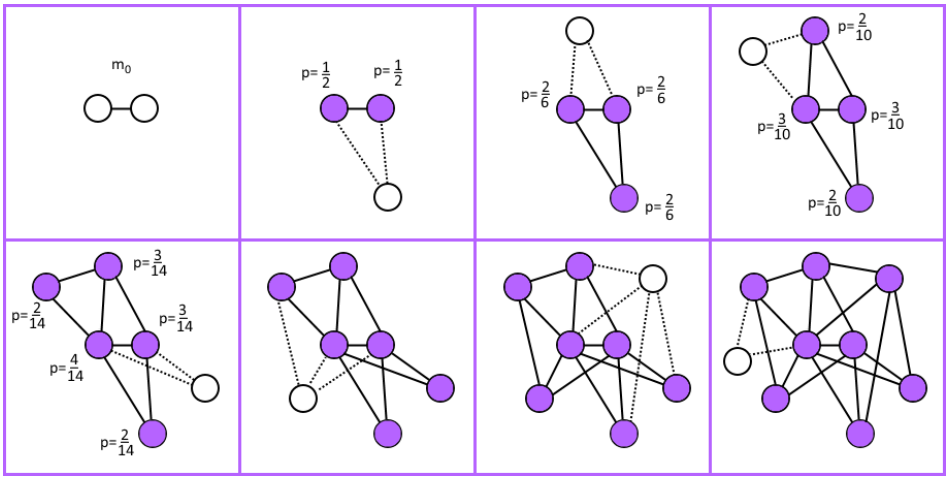


Figure B.0.7 Generating a Barabási-Albert graph.

Cloteaux (CG) [179] suggests a model which is designed specifically to represent PGs. It is computed by creating a minimal spanning tree over a geographical area. Next, links are added

[16] The starting network m_0 can be any network as long as each node has at least one edge, it is usually described as two nodes connected by an edge.

[17] A spanning tree is a tree which connects all nodes in a network, a minimum spanning tree is the tree whose total edge weights are the smallest.

between these nodes provided that they do not cross any other links. The added edges are connected between nodes that are vital to the graph (if they were removed the graph would be unconnected), this helps to enhance robustness and the theory corresponds to the way the PG is actually built. This model was not extensively analysed and Cloteaux explains that PGs throughout the world are very different and therefore have different properties.
It is a very well thought out model as it depends on geographical area and introduces edges to increase robustness. This is one of the more complicated models to produce and without further analysis we cannot easily understand how accurately it describes PGs.

Minimum Distance (MD), a method by Hines [181], is generated by the following algorithm:

Randomly generate planar coordinates (x_a, y_a) from a uniform distribution within a fixed area; these are to represent the position on a geographical plane of node 1.
for $a = 2{:}\,N$

1. Randomly generate planar coordinates (x_a, y_a) from a uniform distribution within a fixed area; these are to represent the position on a geographical plane of node a.
2. Generate $\left\lfloor \frac{v}{N} \right\rfloor$ links between a and already existing nodes, b, by iteratively selecting nodes b to minimise the Euclidean distance between a and b, where $\mathcal{A}_a$ is the set of neighbouring nodes of a:

$$\min_b [(x_a - x_b)^2 + (y_a - y_b)^2] \quad s.t.\, b \notin \mathcal{A}_a. \tag{176}$$

3. Generate one additional link between a and another node b with probability

$$p = \frac{v}{N} - \left\lfloor \frac{v}{N} \right\rfloor \tag{177}$$

(presumably to consider the decimal ignored in (176) by using the *floor* function).

There is also a second method of "Minimum Distance with Bisection" which includes creating a new node which either becomes a new branch to the tree or bisects an existing edge:

$$\begin{gathered} \min(\mathcal{C}_1, \mathcal{C}_2) \\ \text{s.t.}\ \ \mathcal{C}_1 = \min_{b \notin \mathcal{A}_a} [(x_a - x_b)^2 + (y_a - y_b)^2] \\ \mathcal{C}_2 = \min_{i \in \{1,\dots,v\}} d(a \to e) + f\ , \end{gathered} \tag{178}$$

where $d(a \rightarrow e)$ is the Euclidean distance between point a and the nearest point along the line segment e and f is an externally selected bisection cost. If $\mathcal{C}_1$ is smaller than $\mathcal{C}_2$, a new link is created. If $\mathcal{C}_2$ is smaller, the new node bisects an existing link e. Although this does not consider geography or population density, the method does consider the positioning and the evolution of a PG over time.

All of these methods of generating graphs create networks which all have very different properties. Next we will look at the different properties a network can have, and how accurately these models relate to PGs.

Properties of a Network

There is a range of ways that the structure of a network can be described. The reason we need to consider these is that flow dynamics within a network depends on its topology, therefore we should model it as closely as possible.

Degree Distribution:

A neighbourhood or degree, c, describes the number of nodes a given node n is connected to. Average degree is the average numbers of connections per node, $\langle c \rangle$. Degrees can also be drawn from a distribution.

The degree distribution of a PG is widely debated. Some papers [183] say that PGs have a power-law distribution, $P(c) = \alpha c^{-\gamma}$ where typically $1.5 < \gamma < 3$. A network whose degree distribution follows a power law is said to be scale-free because the distribution is independent of scale (constant on a logarithmic scale). Cloteaux [179] explains that this could be because population distributions follow a power law with $1.74 < \gamma < 1.91$ and PGs may follow suit as they are built to accommodate populations.

The opposing idea [184, 185] is that degree distributions follow an exponential distribution, $P(c) = \alpha e^{\beta c}$. This may be due to the number of power lines going into a power station being proportional to the logarithm of the population of the surrounding areas. The survey done by Pagani and Aiello [175] found that most PGs follow an exponential degree distribution with $\beta \approx 0.5$ and Hines [181] shows this with the IEEE 300 model [154]. It is important to consider that both suggestions may be correct for different PGs.

The reason degree distribution is an important property of a network when it comes to analysis of a grid is, for example, that when it comes to robustness of a network, a scale-free network is resilient to random attack, but it is very vulnerable to direct attacks, whereas networks which follow an exponential degree distribution are equally vulnerable to random and direct attacks.

Some networks, such as regular lattice and RRG, can have fixed degrees per node and therefore have a uniform distribution. Preferential attachment models are scale-free, with the BA model at $P(c) \approx c^{-3}$, whereas MD graphs follow an exponential distribution along with some types of random graph. Erdös-Renyi graphs follow a Poisson degree distribution.

Characteristic Path Length:

A path is a route through edges from node i to another node j. The diameter is the minimum number of links to get from node i to j, $d(i,j)$, and the characteristic path length (CPL) is the mean of the path lengths from a node to any other:

$$d_i = \frac{1}{N-1}\sum_{i \neq j} d(i,j). \tag{179}$$

A PGs average path length, $\mathcal{L}$, is proportional to the logarithm of the number of nodes:

$$\mathcal{L} \sim \log N. \tag{180}$$

This is important because the path length can be the minimum distance between a generator and a consumer. A high CPL may result in a higher distribution cost and possibly computing time.

Random lattices and MD networks have a CPL which increases linearly with N. BA networks have an interesting CPL of:

$$\mathcal{L} \approx \frac{\log N}{\log \log N}. \tag{181}$$

SW networks where $p > 0$, random graphs and RRG networks are most representative with a CPL which, like PG, increases by $\ln N$ [186, 187].

Clustering Coefficient:

A clustering coefficient, C, is the measure of the degree to which nodes in the graph tend to cluster together, it is defined by:

$$C_n(k) = \frac{\langle c \rangle}{\binom{c_n}{2}}. \tag{182}$$

PGs have a high clustering coefficient, it is much higher than that of random graphs but not as high as that of SW networks, whose clustering coefficient is at $C(c) = c^{-1}$. BA's networks' clustering coefficient is a power law of $C \sim N^{-0.75}$ [182]. Cloteaux's networks also have a high clustering coefficient.

Some work [188] looks at isolating clusters in the event of a cascading failure to minimise and contain the damage and suggest that one could rely on positioning renewable sources in these areas to sustain the clusters.

High clustering coefficients could, depending on the positions of the generators and consumers within the clusters, be useful in justifying the use of local optimisation methods.

[178] illustrates the change in path length and clustering coefficient in a network for a changing small world parameter p, shown in Figure B.0.8.

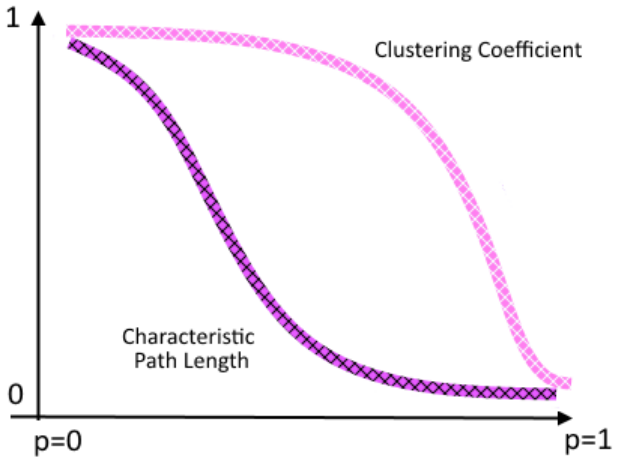


Figure B.0.8 The approximate curves of path length and cluster coefficient for SW topologies of changing p.

Betweenness Distribution:

A shortest path is the shortest distance from node i to node j. When all the shortest paths between node pairs in a network have been calculated, the number of shortest paths which pass through each node is called 'node betweenness'. The betweenness distribution is the distribution of the node betweenness throughout the network of the transmission stages in PGs is scale-free, around $B(c) \sim c^{-1.5}$.

Pagani and Aiello explain that it is not always true that the most central nodes are the most important, in terms of robustness, or that the nodes with the highest betweenness see the biggest electricity flow. This property is more important to the design and robustness of the PG.

Assortativity:

Assortativity is the extent to which nodes connect to others with a similar degree [189]. This may affect the length of distribution paths, or robustness of a graph. A more positively correlated assortativity would lead to higher distribution costs as nodes with lower connectivity may require longer paths to reach necessary nodes. The case of disassortative networks may lead to shorter paths due to better utilisation of the topology.

A PG it typically disassortative, whereas SW and MD networks are assortative. RG, RL and RRG networks are neither assortative nor disassortative.

Choosing a Model

With all these properties in mind, it is easy to forget to consider that PGs differ world-wide. It is not possible to find a model which will perfectly match all existing PGs because they are all different; depending on the landscape of the area, the reliance on electricity in the country, the types of power sources used, the specific design and more.

Table 5 shows the properties each model has and their usefulness for modelling a PG. Throughout my work I will use RRG for analysis of the work done because of its simplicity and the ability to minimise changing variables such as c, and MD will be used for its more realistic application to PGs, to gather information such as computational time.

Table 5. List of properties associated with different network models.

		Properties							
		Simple	Realistic	Ease of Production	Degree Distribution	CPL	Clustering Coefficient	Betweeness Distribution	Assortatitvity
Model	Regular Lattice	✓		✓					
	Random Graph	✓		✓	✓$_e$				
	Random Regular Graph	✓		✓		✓			
	Small- World	✓	✓	✓		✓			
	Barabási-Albert	✓		✓	✓$_{pl}$				
	Cloteaux's Graph		✓				✓		
	Minimum Distance		✓		✓$_e$				

C. Leaf Nodes

From most previous work on message passing, the optimisation has been tested on sparse random regular graphs. This thesis hopes to display how these algorithms could be adapted to work in a real-life power optimisation problem, and so we need to consider what electrical power networks actually look like. Looking at a benchmark network from New England (39-Node Grid) [190] (Figure C.0.9) we see that the network is relatively sparse, but small loops are present. The work of [87] shows us that in most cases small loops are not a problem in MP due to the law of large numbers and normalization. Although MP algorithms are strictly only guaranteed an exact optimal solution for tree-like networks (no loops), networks with loops have been found to give very good approximate optimal solutions if the algorithm converges.

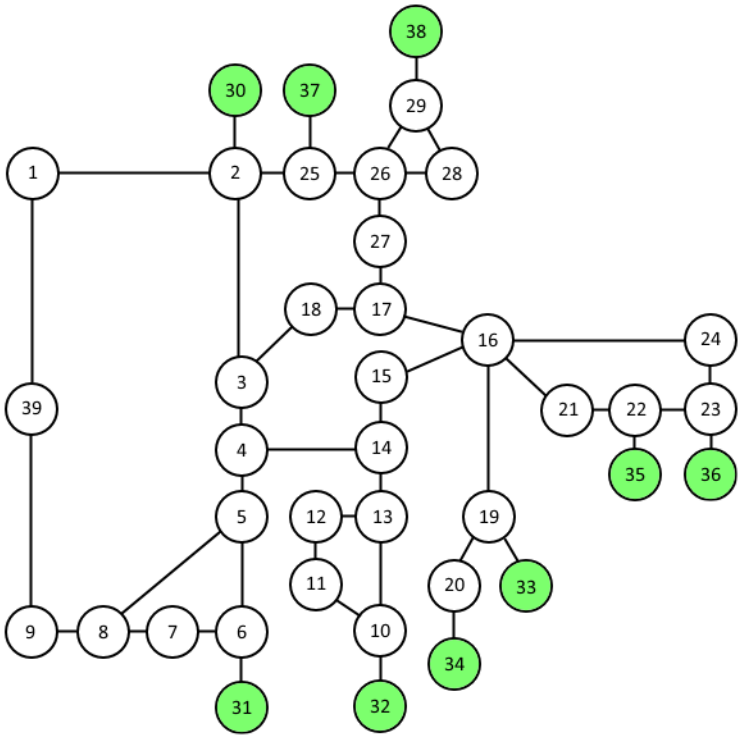


Figure C.0.9 New England 39-node network [190]. Demonstrating leaf nodes in green.

Another problem encountered is that there are many leaf nodes in a power grid topology. A leaf node is a node that is only connected to one other node (Figure C.0.10), this is a problem as the message passing techniques focus on taking information from one node and passing in onto others, which a leaf node cannot do. One way to avoid this would be to modify the network topology by transferring all the capacity of the leaf node into its neighbour and delete the leaf.

Although this may find an optimal distribution for the edited network, due to the change in the topology this may not be optimal for the original topology, as the distribution costs of edges deleted would not be taken into consideration.

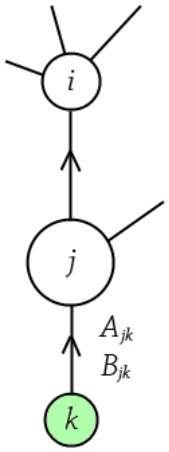


Figure C.0.10 An example of a leaf node (green) locally.

Alternatively, one can use the original calculations and rewrite them explicitly for the case of one neighbour. The neighbour can be considered separately in different iterations as the ancestor, and then as the descendant. The main problem with the existing formulation is that, if node j is a leaf node and node i is the ancestor, there is an absence of descendants; $\sum_{k\neq i}\mathcal{A}_{jk}\left(y_{jk}-\frac{A_{jk}+\phi'_{jk}}{B_{jk}+\phi''_{jk}}\right)=0$. To derive messages from/to leaf nodes one can expand this vanishing term about zero and exploit their independencies to obtain the messages. As the ancestor, the following messages to be passed to node i are:

$$A_{ij}=\begin{cases}0, & \Lambda_j \geq y_{ij},\\ \infty, & \Lambda_j < y_{ij},\end{cases} \tag{183}$$

$$B_{ij}=\begin{cases}0, & \Lambda_j > y_{ij},\\ \infty, & \Lambda_j \leq y_{ij}.\end{cases} \tag{184}$$

Once the neighbour i is the parent subject and node j is the descendant, the infinity values will be incorporated within the messages from descendants. The values will divide one by the other, cancelling out both infinity values (along with objective functions) and a function of $y_{ij}-\Lambda_j$ emerges for the backwards current back to the leaf node.

The neighbour is then assessed as the descendant. Now μ_{ij} will be zero or negative and we send a back message of:

$$y_{jk} = \begin{cases} y_{jk} - \dfrac{A_{jk} + \phi'_{jk}}{B_{jk} + \phi''_{jk}}, & \text{if} \quad \mu_{ij} = 0, \\ -\Lambda_j, & \text{if} \quad \mu_{ij} < 0. \end{cases} \tag{185}$$

This method works effectively on networks such as the 39-node example grid. The ability to pass messages in topologies with leaf nodes allows the algorithm to be more applicable to real-life power networks. However, this method of considering leaf nodes may mean that the node becomes satisfied unconditionally, which becomes more complicated when considering minimising load shedding or fluctuations. Also, because of infinite values of A_{ij} and B_{ij}, the paired objective functions will be neglected and the true minimal distribution may not be found.

Other methods of considering leaf nodes would be to consider a price iteration scenario at these nodes, this would give:

$$\mu_i = \min\left[\mu_j + \Lambda_j, 0\right], \tag{186}$$

which is similar to the function that emerges from (183).

D. Load Shedding Calculations

From the function:

$$\langle \mu_{ij}^2 \rangle_{\Lambda_j} = \begin{cases} 0, & \text{if } \Lambda_j - x \leq 0, \\ \dfrac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2} \displaystyle\int \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j, & \text{if } -\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b - \Lambda_j < 0, \\ \dfrac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2 + \dfrac{1}{\alpha_j}} \displaystyle\int \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j, & \text{if } -\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b - \Lambda_j \geq 0, \end{cases} \tag{187}$$

we can write this as an integral, however it does depend on whether x or $-\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b$ is larger because this determines the limits of each integral.

If $-\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b < x$:

$$\begin{aligned} \langle \mu_{ij}^2 \rangle_{\Lambda_j} = \int_x^\infty 0 \, \mathrm{d}\Lambda_j + \int_{-\infty}^{-\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b} & \frac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2} \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j \\ + \int_{-\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b}^{x} & \frac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2 + \dfrac{1}{\alpha_j^2}} \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j, \end{aligned} \tag{188}$$

or if $-\sum_{\substack{b \neq j \\ b \in N}} \Lambda_b > x$:

$$\langle \mu_{ij}^2 \rangle_{\Lambda_j} = \int_x^\infty 0 \, \mathrm{d}\Lambda_j + \int_{-\infty}^{x} \frac{1}{\left(\sum_{k \neq i} \mathcal{A}_{jk} \dfrac{1}{B_{jk} + \phi''_{jk}} \right)^2} \left[\Lambda_j^2 - 2x\Lambda_j + x^2 \right] \mathrm{d}\Lambda_j. \tag{189}$$

E. Minimising Environmental Costs

The cost of environmental impact should be an increasing concern in power grid distribution. To consider this is as an objective function and exploiting the generation cost minimisation framework to reduce the environmental impact one could use a similar framework to (117):

$$F_{ij} = \boldsymbol{\varepsilon}^T \boldsymbol{A} + \frac{1}{2}\boldsymbol{\varepsilon}^T \boldsymbol{B}\boldsymbol{\varepsilon} - E_j\big((\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij}\big) + \mu_{ij}\big[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij} + \Lambda_j\big] + E_j\gamma_{ij}\big[(\boldsymbol{y}+\boldsymbol{\varepsilon})^T\mathbf{1} - y_{ij}\big], \tag{190}$$

where E_j is zero for all consumers, and a predetermined, individual environmental cost value for each generator. There is also a second Lagrange multiplier, γ_{ij}, which indicates that generators must not receive power; since all consumers have a $E_j = 0$ cost, the constraint will only be enforced on generators. We are assuming the environmental impact to be linear to power provided.

Minimising the VFE with respect to the Lagrange multipliers gives:

$$\mu_{ij} = \min\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + \Lambda_j + \big(-\boldsymbol{A} + E_j\mathbf{1}\big)^T\boldsymbol{B}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{B}^{-1}\mathbf{1}}\right], \tag{191}$$

$$\gamma_{ij} = \max\left[0, \frac{\boldsymbol{y}^T\mathbf{1} - y_{ij} + \big(-\boldsymbol{A} + E_j\mathbf{1}\big)^T\boldsymbol{B}^{-1}\mathbf{1}}{G_j\mathbf{1}^T\boldsymbol{B}^{-1}\mathbf{1}}\right], \tag{192}$$

which gives $\mu_{ij} = f(\gamma_{ij})$, and $\gamma_{ij} = g(\mu_{ij})$ but as both functions are non-zero only if they are not obeyed, and as they are non-overlapping constraints, we can assume that if either function is non-zero the other must be zero, allowing us to omit each one from the others equation, but not their derivatives.

Minimising the VFE again to find the optimal messages to pass to the ancestor node gives the first and second derivative messages:

$$A_{ij} = \begin{cases} E_j - (\mu_{ij} + E_j\gamma_{ij}), & \text{in case 1,} \\ E_j - 2(\mu_{ij} + E_j\gamma_{ij}), & \text{in case 2,} \\ E_j, & \text{in case 3,} \end{cases} \tag{193}$$

$$B_{ij} = \begin{cases} \dfrac{1}{\mathbf{1}^T \boldsymbol{B}^{-1} \mathbf{1}}, & \text{in case 1,} \\ \dfrac{4}{\mathbf{1}^T \boldsymbol{B}^{-1} \mathbf{1}}, & \text{in case 2,} \\ 0, & \text{in case 3,} \end{cases} \tag{194}$$

where the cases are as in Table 4, and the backwards message is:

$$\boldsymbol{y} = \boldsymbol{y} + \boldsymbol{B}^{-1}\left(-\boldsymbol{A} + E_j \mathbf{1} - \left(\mu_{ij} + E_j \gamma_{ij}\right)\mathbf{1}\right). \tag{195}$$