Markov chain methods for rare event sampling and applications to energy systems

Maldon Patrice Goodridge, ID 170751203

Supervisor: Professor John Moriarty

Submitted in partial fulfilment of the requirements of the
Degree of Doctor of Philosophy

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on April 23, 2022.

I, Maldon Goodridge, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Maldon Goodridge
Date: December 20, 2021

Details of collaboration and publications:

1. *Distribution of cascade sizes in power system emergency responses*;
   Authors: Maldon Goodridge, John Moriarty, Andrea Pizzoferrato;
   Conference proceeding: Probabilistic Methods Applied to Power Systems (PMAPS);
   Publication date: July 2020.

2. *A rare-event study of frequency regulation and contingency services from grid-scale batteries*;
   Authors: Maldon Goodridge, John Moriarty, Andrea Pizzoferrato;
   Journal: Philosophical Transactions of the Royal Society, A;
   Publication date: June 2021.

3. *Skipping between distant basins*;
   Authors: Maldon Goodridge, John Moriarty, Jure Vogrinc, Alessandro Zocca;
   Journal: Journal of Global Optimization;
   Accepted. Publication date: May 2022.

Abstract

This thesis attempts to address the problems of sampling rare events in power system operations, global optimisation studies and in higher dimensions. Our primary algorithmic tool is the *skipping sampler*, an existing Metropolis-class algorithm designed to efficiently draw samples from a distribution π, whose support C, consists of connected components. First, we apply the skipping sampler to a cyber-physical-statistical power system simulation model to sample power injections from renewable energy sources, conditioned on the activation of frequency-related emergency responses. Such emergency responses, designed to protect sensitive equipment from deviations in system frequency, occur infrequently, and can be considered a rare event. We also explore how the application of large battery energy storage systems can mitigate this risk. Methodologically, we apply the skipping sampler to the field of global optimisation, where we present the *basin hopping with skipping* algorithm, which replaces the perturbation step of the well-known basin hopping routine with the proposal function of the skipping sampler. Results indicate that, for energy landscapes with well-separated basins, the basin hopping with skipping algorithm is both more effective and efficient at locating the global minima than the basin hopping routine. Finally, to address the problem of drawing samples of rare events in higher dimensions, we propose the *Sequential Monte Carlo with skipping* (SMC-S) algorithm, which use the skipping sampler as the transition kernel of a sequential Monte Carlo framework. To address the challenge of sampling particle paths which intersect with regions of interest in high dimensions, the skipping sampler kernel samples the direction particle paths from a data-driven, empirical distribution, based on the relative positions of particles. Experiments suggest that the SMC-S, using this approach, outperforms both MCMC and other SMC routines in drawing samples of rare events in high dimensions.

To my wife, Kamesha- thank you for your patience, support and encouragement.

Per ardua, ad astra.

Acknowledgements

This thesis represents the efforts of many, whose direct or indirect actions contributed to its formulation and completion. I take this opportunity to acknowledge the individuals, without whom, this work would not be possible.

Firstly, I express my heartfelt thanks and gratitude to Professor John Moriarty who supervised my research activities during my Ph.D tenure. I acknowledge your patient guidance, timely advice and mentoring over the past four years. Please know your tireless efforts are appreciated- thank you.

I also extend gracious thanks to the Principal of Queen Mary, University of London, the Research Degrees office and the School of Mathematics, who supported my research financially through the provision of the Principal's studentship for postgraduate research.

I also thank Dr. Jure Vongrinc, who worked with me to develop the Markov chain and sequential Monte Carlo theorems and algorithms discussed in Chapter 6. Special thanks goes to Dr. Andrea Pizzoferrato for his assistance in the mathematical formulation of the protection systems of the power system model. Similarly, thanks goes Professor Alessandro Zocca for his assistance in the design of the global optimisation algorithm. Finally, I thank Professor Pierluigi Mancarella and Dr. Mehdi Dozein for their guidance in the area of power systems engineering and the formulation of the power system model used in this thesis.

Special thanks also goes to the Ph.D cohort during my tenure, who never declined my request for assistance or support. Of note are Ali Raad, Evangelos Mitsokapas, Francesco Coghi, Asma Farid, Marco Grammatica and Katerina Zahradova. I thank you all for your time, your efforts and especially your friendship.

To those who supported me personally, my wife- Kamesha, my mother- Marcia, my father- McArthur; without your continued support and encouragement, I could not accomplish this. Finally, all thanks goes to my God, whose faithfulness and mercy is the source of my success.

Thank you.

Contents

List of Figures

List of Tables

0.1 Introduction

This thesis is motivated by the critical importance of rare events, such as blackouts, in electrical power systems. Rare event analysis is pertinent to this context because the power system of the UK, like those of many other countries, is designed and operated to very high standards of reliability. Thus the events which have the most impact on the power system, such as blackouts, occur so rarely that there is very little relevant empirical data available for study. This fact motivates the use of other approaches such as simulations.

For essentially the same reasons, the simulation of rare events in power systems is computationally challenging. This creates an exciting interdisciplinary research opportunity which requires careful study in both computational statistics and power systems engineering, and it is this interface which is addressed here. At the level of methodology the main tool used in this thesis is the *skipping sampler*, a Metropolis-class rare-event sampling algorithm recently developed for this application by Moriarty, Vogrinc and Zocca [75]. This sampler proceeds by jumping or 'skipping' across the complement of the rare event of interest. It is particularly suitable for use with complex, 'black box'-type simulation models because it requires only pointwise knowledge of the rare event.

The first interdisciplinary contribution of this thesis is the precise mathematical description of a cyber-physical-statistical power system simulation model. The physical part is sufficiently detailed to capture both the necessary continuous electrical and mechanical dynamics. It is a so-called *third order* model including both voltage and power dynamics, which are necessary to model the large power disturbances which are required to cause power systems to fail [98, 56]. The statistical part reflects the fact that randomness plays an increasingly important role in our decarbonised power systems, due mainly to the increased penetration of stochastic renewable generation. It models *power disturbances*, which represent random changes (either positive or negative) to the electrical power injected at the different nodes of the system. The 'cyber' part accounts for both the discrete-time actions of power system emergency protection equipment and the physical discontinuities caused by its operation. This enables the modelling of cascading failures, as the emergency responses of protection equipment lead to changes in the topology and other physical characteristics of the power system, potentially causing additional system instability triggering further emergency response, and so on.

The thesis's second interdisciplinary contribution is to couple the skipping sampler, for the first time, with a detailed power system simulator. Inspired by the ongoing adoption of grid-scale battery storage, this model combines the above cyber-physical-statistical model with a model of automatic voltage regulation and a battery storage model recently developed in the engineering literature. In a detailed case study, this model is then applied to a benchmark power system model, namely the

Kundur two-area system [57].

In common with many Metropolis-Hastings algorithms, the skipping sampler's performance degrades as the dimension of the sample space increases. The third contribution of this thesis is an empirical study of promoting the skipping sampler to a sequential Monte Carlo (SMC) algorithm [21] and, hence, evolving a population of samples rather than just one. This attempts to use the fact that at each step of the algorithm, the population contains information about the geometry of the rare event of interest. The idea is to use this information to guide the sampling process during the next step of the algorithm, in order to help mitigate the effect of high dimension.

In order to make the best possible system design and capital planning decisions, *optimisation* also plays an important role in power systems engineering. It was briefly noted in [75] that the skipping sampler could be combined with *basin hopping* (BH), a stochastic algorithm for global non-convex optimization [111]. The connection is made by first noting that BH applies random walk sampling to search for the global minimum. Then, in challenging non-convex optimisation problems, this search may be considered a problem of rare event sampling. This motivates substituting the skipping sampler for the random walk sampler, giving the so-called *basin hopping with skipping* (BH-S) algorithm. A fourth contribution of this thesis is then to global non-convex optimization, providing the first systematic exploration of the BH-S algorithm on a variety of benchmark optimisation surfaces.

The thesis is organised as follows. Chapter 1 provides background on rare-event sampling before focusing on the skipping sampler. Two novel proofs are provided, concerning the reversibility both of the sampler and of a more efficient implementation (the 'doubling trick'). Moving next to the perspective of power systems engineering, Chapter 2 presents the cyber-physical part of the power system model. To validate the use of this model for a statistical study of cascading failures, Chapter 3 carries out a case study on 'toy' power systems with differing network topologies. The skipping sampler is first applied in Chapter 4. Here the use of the Kundur two-area system, together with careful calibration of the model parameters, results in a cyber-physical-statistical model which is robust in the sense that failures are rare events, which we analyse in detail. Chapter 5 presents the application of skipping to global non-convex optimisation while Chapter 6 details the empirical investigation of the skipping sampler's application to sequential Monte Carlo.

Chapter 1

Markov Chain Monte Carlo methods for rare event sampling

Specialised rare-event sampling algorithms use numerical methods to encourage a simulation model to generate more realisations of a particular rare event of interest than would be generated by brute force simulation [11]. In the literature, Markov chain algorithms such as Markov chain Monte Carlo (MCMC) and non-Markov algorithms such as *importance sampling* have been discussed as methods for rare event sampling. This thesis will focus on developing Markov chain approaches by presenting a series of MCMC algorithms specially designed for sampling rare events. These methods will be derived from the *skipping sampler*, an MCMC algorithm for intended to draw samples from a distribution π with connected support introduced in [76] (see Section 1.4).

This chapter proceeds as follows: in Section 1.1 we provide relevant definitions from Markov chain literature while Section 1.2 will present the *Metropolis-Hastings* algorithm. These are intended to familiarise the reader with the terminology used throughout the thesis. Section 1.4 presents a review of the skipping sampler, with alternatives proofs of Proposition 5 from [75] provided in Section 1.5. Finally Section 1.6 presents a discussion, proof and case study of the *doubling sampler*, an extension of the skipping sampler proposed by the authors of [76].

In this chapter, the state space is taken to be $\Omega = \mathbb{R}^d$, with capitalised Latin letters representing random variables and lower-case Latin letters representing realised outcomes. Density functions and their corresponding distributions will be denoted by the same Greek symbols.

1.1 Markov chain background

This section presents relevant theorems and definitions from Markov chain literature used in subsequent chapters of the thesis. Unless referenced otherwise, this section presents the theorems and definitions from [107].

Definition 1.1.1 (Markov chain). A stochastic process $X(t) = X_0, X_1, \ldots$ is a discrete time Markov chain once it satisfies:

$$\mathbb{P}(X_{n+1} = y | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} | X_n = x_n).$$

where the quantities $\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$ are called *transition probabilities*.

If the transition probabilities do no depend on the time step n, then the Markov chain is said to be *time homogeneous*:

Definition 1.1.2 (Time homogeneous Markov chain). A Markov chain $\{X_i\}_{i \geq 0}$ is *time homogeneous* if $\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$.

For general state spaces, transitions from state $X_{n-1} = x$ at time $n - 1$ to state $X_n = y$ at time n are given by a *Markov kernel*, defined as:

Definition 1.1.3 (Markov Kernel, [81]). For measurable spaces $(\mathbb{R}^d, \mathcal{A})$ and $(\mathbb{R}^d, \mathcal{A}_1)$, a *Markov kernel* M_1, is a map $M_1 : \mathbb{R}^d \times \mathcal{A}_1 \to [0, 1]$ such that:

i $\forall x \in \mathbb{R}^d, M_1(x, \cdot)$ is a probability measure on $\mathcal{A}_1$;

ii $\forall A_1 \in \mathcal{A}_1, M_1(\cdot, A_1)$ is $\mathcal{A}$-measurable.

As previously stated, we aim to construct a Markov chain sample from a density π. To achieve this, we can design a Markov such that π is the unique *invariant/stationary distribution*:

A time-homogeneous Markov chain with *invariant distribution* π is a sequence of random variables $\{X_n\}_{n \geq 0}$ such that the transition kernel P given by:

$$P(X_n, A) = \mathbb{P}\{X_{n+1} \in A | X_0, \ldots, X_n\}$$

satisfies:

$$\pi(A) = \int \pi(x) P(x, A) \tag{1.1}$$

for all measurable sets A.

We desire to construct a Markov chain such that π is the only stationary distribution, regardless of X_0. The conditions required for convergence to a unique stationary distribution are established in Theorem 1.1.7. We first introduce the following key concepts:

Definition 1.1.4 (π-irreducibility). A transition kernel P on $(\mathbb{R}^d, \mathcal{A})$ is π-*irreducible* if for each initial state $x \in \mathbb{R}^d$, and for each $A \in \mathcal{A}$ with $\pi(A) > 0$, there exists an integer $n = n(x, A) \geq 1$ such that $P^n(x, A) > 0$.

Irreducibility of the transition kernel implies there should be a non-zero probability of transitioning to any state x in the support of π. Additionally, transitions between disjoint sets in the sample space should not occur cyclically, formally, P must be *aperiodic:*

Definition 1.1.5 (Periodicity:). A π-irreducible transition kernel P is *periodic* if there exists an integer $d \geq 2$ and a sequence $\{E_0, E_1, \ldots, E_{d-1}\}$ of d non-empty disjoint sets in $\mathbb{R}^d$ such that, for all $i = 0, \ldots, d-1$ and all $x \in E_i$,

$$P(x, E_j) = 1 \quad \text{for } j = i + 1 \pmod{d}$$

Otherwise, the kernel is *aperiodic.*

A key concept for the convergence of theory of generate state space chains is *recurrence:*

Definition 1.1.6 (Recurrence). A π-irreducible chain $\{X_n\}_{n \geq 1}$ with invariant distribution π is *recurrent* if, for each A with $\pi(A) > 0$,

$$\mathbb{P}\{X_n \in A \text{ i.o.}\} > 0 \text{ for all } x, \tag{1.2}$$

$$\mathbb{P}\{X_n \in A \text{ i.o.}\} = 1 \text{ for } \pi\text{-almost all } x. \tag{1.3}$$

The chain is *Harris recurrent* if $\mathbb{P}\{X_n \in A \text{ i.o. }\} = 1$ [1] for all x.

The requirements for a Markov chain to converge to its unique, stationary distribution are given in Theorem 1.1.7 (Theorem 1 in [107]):

Theorem 1.1.7. *Suppose P is π-irreducible and $\pi P = \pi$. Then P is positive recurrent and π is the unique invariant distribution of P. If P is also aperiodic, then for π-almost all x,*

$$||P^n(x, \cdot) - \pi|| \to 0,$$

with $|| \cdot ||$ denoting the total variation distance. If P is Harris recurrent, then the convergence occurs for all x.

Finally, we present the concept of *reversibility:*

Definition 1.1.8 (Reversibility). A Markov chain is *reversible* with respect to a probability distribution $\pi(\cdot)$ if:

$$\pi(x)p(x, y) = \pi(y)p(y, x) \qquad \forall (x, y) \tag{1.4}$$

---

[1]i.o.: Suppose that $\{A_n : n \geq 1\}$ is a sequence of events in a probability space. Then the event A(i.o.) = $\{A_n$ occurs for infinitely many $n\}$ is given by $A(i.o.) := \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n$.

This relationship is also known as the *detailed balance equation*, and is a sufficient condition for π to be the unique stationary density of the Markov chain [31](page 194).

The reversibility of the chain is used in the construction of the Metropolis-Hastings algorithm, a class of MCMC algorithm which will be discussed later in Section 1.2.

1.2 The Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is a class of Markov Chain Monte Carlo methods used when a sample must be taken from a distribution π from which sampling is difficult, complicated or expensive due to the intractability of the calculation [31, 107], or because the dimensionality of the problem is large [19]. The basic aim of the MH algorithm is to construct a Markov chain with state space $\mathbb{R}^d$ and stationary distribution π, then simulate a long run of the chain. Following an initial, transient phase, realisations of the chain can be used as samples from the density of interest π [94].

A Metropolis-Hastings algorithm is presented in Algorithm 1:

Algorithm 1: Metropolis-Hastings algorithm; adapted from [31]
---
1 Initialize iteration counter $i = 0$. Set initial value for investigation: X_0;
2 **for** *i = 1 to N* **do**
3 $\quad$ Propose new value Y using $q(X_i, Y)$;
4 $\quad$ Calculate the acceptance probability α:
5

$$\alpha(X_i, Y) = \min\left\{1, \frac{\pi(Y)q(Y, X_i)}{\pi(X_i)q(X_i, Y)}\right\}$$

$\quad$ Generate a uniform random number $U \in [0, 1]$;
6 $\quad$ **if** $\alpha \geq U$ **then**
7 $\quad\quad$ Accept the proposed state Y and set $X_{i+1} = Y$;
8 $\quad$ **else**
9 $\quad\quad$ Set $X_{i+1} = X_i$;
10 $\quad$ **end**
11 $\quad$ Increase the iteration counter by 1: $i = i + 1$;
12 **end**

1.2.1 The Metropolis-Hastings transition kernel

The MH algorithm presented by Hastings in [40] constructs a Markov transition kernel $P(x, y)$ such that π is the equilibrium distribution of the resulting Markov chain by ensuring the transition kernel P satisfies the detailed balance equation (1.4). As mentioned in Section 1.1, this is a sufficient condition for π to be the unique stationary distribution of the resulting Markov chain.

Suppose the transition kernel, for some function $p(x, y)$, can be expressed as

$$P(x, dy) = p(x, y)dy + \delta_x(dy)\left[1 - \int_{\mathbb{R}^d} p(x, y)dy\right], \tag{1.5}$$

where $p(x, x) = 0$ and δ_x denotes point mass at x.

Theorem 1.2.1. *If the function $p(x, y)$ satisfies the detailed balance condition (Equation (1.4)), then π is the invariant density of P. [107]*

Proof. We now verify this claim: Starting with the right-hand side of Equation (1.1) and letting $r(x) = 1 - \int_{\mathbb{R}^d} p(x,y)dy$:

$$
\begin{aligned}
\int P(x,A)\pi(x)dx &= \int \left[\int_A p(x,y)dy \right]\pi(x)dx + \int r(x)\delta_x(A)\pi(x)dx \\
&= \int_A \left[\int p(x,y)\pi(x)dx \right]dy + \int_A r(x)\pi(x)dx \\
&= \int_A \left[\int p(y,x)\pi(y)dx \right]dy + \int_A r(x)\pi(x)dx \\
&= \int_A (1 - r(y))\pi(y)dy + \int_A r(x)\pi(x)dx \\
&= \int_A \pi(y)dy
\end{aligned}
$$

$\square$

For the MH algorithm, the function $p(x,y)$ is decomposed into a *proposal density* $q(x,y)$ and an acceptance probability $\alpha(x,y)$ as in Equation (1.6):

$$
p(x,y) = q(x,y)\alpha(x,y) \qquad x \neq y. \tag{1.6}
$$

The proposal density $q(x,y)$, with $\int q(x,y)dy = 1$ [19] is used to generate a potential move of the chain- i.e. q is the conditional density of proposing state y given that the chain is currently at $X_n = x$. Positivity of the proposal density ($q(x,y) > 0$ for all $x,y \in \mathbb{R}^d$) ensures the irreducibility of the resulting MH chain ([109], Equation 2.1).

The *acceptance probability* $\alpha(x,y)$ is the probability of accepting the proposal $Y = y$ into the Markov chain, so $X_{n+1} = y$ [31]. The choice and structure of the acceptance probability is important for the form of the algorithm. Hastings, in the formulation of the generalised *Metropolis-Hastings* algorithm, proposed that $\alpha(x,y)$ be constructed as in Equation (1.7):

$$
\alpha(x,y) = \min \left\{ 1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} \right\}, \tag{1.7}
$$

such that, when combined with the proposal density q, the resulting MH transition kernel P satisfies the detailed balance Equation (1.4) [40]. Note Equation (1.6) is defined for $y \neq x$. The probability for the chain to remain at x is given by Equation (1.8):

$$
p(x,x) = 1 - \int q(x,y)\alpha(x,y)dy. \tag{1.8}
$$

Combining Equations (1.6) and (1.8) yields the expression of the MH transition kernel P:

$$P(x, A) = \int_A q(x, y)\alpha(x, y)dy + \delta_x(dy)\left[1 - \int q(x, y)\alpha(x, y)dy\right], \qquad (1.9)$$

for any set $A \subset \mathbb{R}^d$.

1.2.2 Metropolis-Hastings implementation

In the implementation of the MH algorithm, one common class of proposal density is the *symmetric proposal*, where $q(x, y) = q(y, x)$. When the proposal density is symmetric, the acceptance probability of the move simplifies to

$$\alpha(x, y) = \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}. \qquad (1.10)$$

which does not depend on the proposal density q, and is usually simpler to evaluate than the Metropolis-Hastings acceptance probability in Equation (1.7). MH algorithms with this simplified acceptance probability are referred to as *Metropolis* algorithms.

A second class of proposal density is the *random-walk proposal* where $q(x, y) = q_1(y - x)$. The proposed state y is drawn following the process $y = x_i + W_i$ where $W_i \sim q_1$ is a random variable, usually with distribution independent of the chain. This leads to the exploration of the state space through a random walk process, with the resulting chain referred to as a *random-walk chain*. When $q_1(c) = q_1(-c)$, that is, q_1 is symmetric around zero, the acceptance probability again simplifies to Equation (1.10). Common choices for q_1 are the Gaussian, Student's t and the uniform distributions centred at the origin [31]. The seminal work for MCMC algorithms employs a random-walk chain (see [71]), with such algorithms referred to as *random-walk Metropolis* algorithms.

A third class of proposal density is the *independent proposal*, where $q(x, y) = q_1(y)$ leading to *independent chains* [19]. For example, if the form of π is known and $\pi \propto f(x)h(x)$ where $h(x)$ is a density that can be sampled and $f(x)$ is uniformly bounded, if $q(x, y) = h(y)$, then the acceptance ratio can be simplified to

$$\alpha(x, y) = \frac{f(y)}{f(x)}. \qquad (1.11)$$

For example, in a Bayesian context, one implementation of the independent chain is for h to be a prior density and f the likelihood function [31].

The above classes of proposal functions are not exhaustive. In fact, subject to the conditions stated above, the researcher can design proposal functions for MH algorithms to address problem-specific challenges which may arise in sampling. In this thesis, we exploit the flexibility in the construction of q to develop specialised Metropolis-Hastings algorithms for rare event sampling. Following the notation used in [76], we will denote by $MH(\pi, q)$ a Metropolis-Hastings algorithm with target

density π and proposal function q.

The dispersion of the proposal density has implications for the performance of the algorithm [19] and its choice is a consideration in the "tuning" of the algorithm, i.e- choosing parameters to minimise the iterations required for the chain to reach its stationary density. The dispersion of the proposal density affects both the *acceptance rate* (the ratio of the number of accepted moves to the total number of proposals) and the region of the sample space explored by the chain. For example, the dispersion of a Gaussian random-walk proposal $q(x, y) \sim \mathbb{N}(X_n, \sigma^2)$ can be controlled by the choice of the variance σ^2. As $\sigma^2 \to 0$, q generates minute displacements relative to the current state $X_n = x$, requiring a large number of iterations to explore the state space, and for the chain to converge to the desired stationary distribution [19], a phenomena known as 'slow mixing'. Conversely, for $\sigma^2 \gg 0$, if X_n is in the high density region of π and a proposal is selected from a lower density region, then acceptance rates may be low as such proposed states will more likely be rejected [14, 31]. Tuning of proposal dispersion should have the aim to achieve an optimal acceptance rate. However, no single optimal rate has been prescribed in the literature, with rates between 15% to 48% quoted as desirable to avoid slow mixing. [31].

A key characteristic of Metropolis-Hastings algorithms is 'burn-in' or 'warm-up', referring to the number of trials which should be discarded from the start of the chain, when the chain is yet to achieve its desired stationary distribution. While the burn-in quantity can be identified by measuring the auto-covariances, in practice however it is usually sufficient for users to build long chains using the MH algorithm, and discard the first 1 - 2% of states as burn-in [34].

1.3 Metrics

1.3.1 The batch means test

The *Batch Means* statistic is one method used to calculate Monte Carlo Standard Errors for MCMC algorithms. The batch means method attempts to reduce auto-correlation by batching observations.

In general, if it is intended to estimate the expectation of some function $\mu = E(g(X))$ where X is drawn from a distribution π, then a non-overlapping *batch means* method requires the output Markov chain $\{X_N\}$ to be divided into blocks of equal size, that is, $N = ab$ with $a, b \in \mathbb{N}$. This divides the sample generated from the Markov process into a batches each of size b. For each batch a, a sample mean Y_k of the kth batch is given by:

$$Y_k = \frac{1}{b} \sum_{i=(k-1)b+1}^{kb} g(X_i) \quad \text{for } j = 1, ..., a. \tag{1.12}$$

The batch means estimate of the sample variance given by:

$$\hat{\sigma}_g^2 = \frac{b}{a-1} \sum_{k=1}^{a} (Y_k - \hat{\mu})^2, \tag{1.13}$$

where $\hat{\mu} = \frac{\sum_{i=1}^{a} Y_i}{a}$.

The batch mean estimate of the Monte Carlo Standard Error (MCSE) is given by:

$$MCSE = \frac{\hat{\sigma}}{\sqrt{N}}. \tag{1.14}$$

The batch size b should be such that each Y_k is independent. Furthermore, the authors of [50] revealed that if the batch size and the number of batches are allowed to increase as the overall length of the simulation increases, then $\hat{\sigma}_g^2 \to \sigma_g^2$ almost surely as $n \to \infty$. As such, a reasonable heuristic for b is $b = \sqrt{N}$, and $a = N/b$ to ensure that the estimator $\hat{\sigma}_g^2$ converges to σ_g^2 asymptotically [17].

The MCSE can be used as to determine the stopping time of an MCMC simulation. This is done by first determining an acceptable MCSE for the experiment's parameters. As the algorithm is running, periodically compute the standard errors for the parameters using the batch means methodology. The chain can be halted when the standard errors reach a desired level. In general, lower values for the MCSE are preferred, as they indicate a chain with consistent mean between arbitrary batches
.

1.3.2   Mean squared jump distance

The mean squared jump distance (MSJD) statistic is used to calculate the average of the square of the Euclidean distance between elements in the Markov chain. This is given by:

$$MSJD := \frac{1}{n-1} \sum_{i=1}^{n-1} \|x_{i+1} - x_i\|_2^2. \tag{1.15}$$

where x_i are elements of the Markov chain. This statistic can be used to evaluate the mixing of the chain, which is directly related to how quickly the algorithm converges to its stationary distribution. Broadly stated, higher MSJD values are associated

with high acceptance rates, while lower values are associated with low acceptance rates and high rejection rates [9].

1.4 The skipping sampler: Efficient sampling of targets with non-convex support

1.4.1 Background

Much of the novel work developed in subsequent sections of this thesis are based upon the results of [76]. Given its relevance, this section summarises its findings, which introduces the *skipping sampler*, an easy to implement Metropolis-class algorithm intended to improve exploration of target densities π with non-trivial support C. The target density π is assumed to possess the following characteristics:

Assumption 1: π *is a probability density function on* $\mathbb{R}^d$ *whose support*

$$C = supp(\pi) := \{x \in \mathbb{R}^d : \pi(x) > 0\}$$

satisfies $Leb(C^c) > 0$ *where* C^c *is the complement of* C *and* Leb *denotes Lebesgue measure on* $\mathbb{R}^d$. (page 1, [75])

Relevant to this thesis, targets of this form were said to arise in rare event sampling problems, and could pose a challenge for random-walk Metropolis algorithms which conduct local exploration moves for the next state of the chain, as any proposals in C^c would be rejected. Instead, the *skipping sampler* attempts to cross C^c in a sequence of linear steps in order to generate a relevant proposal, with the authors noting this method can be advantageous when C is composed of connected components. In addition, the *skipping sampler* is a Metropolis-class algorithm due to the symmetry in the construction of its proposal density q (discussed below). Thus, q need not have a convenient closed-form expression as it is not required for the Metropolis acceptance probability (Equation (1.10)). The implementation of the skipping sampler requires no *a priori* knowledge of the target density π or its support, only a means to determine if a state is in C.

1.4.2 The skipping sampler

Assumption 2. *Let* $q : \mathbb{R}^d \mapsto \mathbb{R}$ *be a symmetric* $(q(x) = q(-x))$ *continuous probability density function with* $q(\mathbf{0}) > 0$. *We refer to* q *as the underlying proposal density.* (Page 3, [75])

The symmetry of the proposal density given by Assumption 2 allows the use of the simplified Metropolis acceptance probability (1.10). We relax the symmetric assumption for q in Sections 1.4.6 and 6.3, requiring an acceptance probability with form similar to (1.7). The skipping sampler enables a random-walk Metropolis algorithm to traverse regions where the target density π is zero. The skipping sampler initially generates proposal $Y_{n+1} = y$ using a random walk proposal. If $Y_{n+1} \in C^c$, then this state would be rejected in a random walk Metropolis-Hastings algorithm; instead, in the skipping sampler, states $Y_{n+1} \in C^c$ are updated by adding jumps of random size in a linear trajectory until C is entered, or the updating procedure (referred to as 'skipping') is halted. This is formalised in Algorithm 2.

Algorithm 2: Skipping sampler algorithm (adapted from [76])

1 Set $X := X_n$ and $Z_0 = X$;

2 Generate the initial proposal Y of the skipping chain distributed with $u \mapsto q(u - X)$;

3 Calculate the direction $\Phi = \frac{(Y-X)}{|Y-X|}$;

4 Generate an independent halting index $K \sim \mathcal{K}$;

5 Set $k = 1$ and $Z_1 := Y$;

6 **while** $Z_k \in C^c$ **and** $k < K$: **do**

7 $\quad$ Generate an independent ditance increment R distributed as $|Y - X|$ given Φ;

8 $\quad$ Set $Z_{k+1} := Z_k + \Phi R$;

9 $\quad$ Set $k = k + 1$;

10 **end**

11 Set $Z := Z_k$;

12 Evaluate acceptance probability:

$$
\alpha(X, Z) = \begin{cases} \min\left(1, \frac{\pi(Z)}{\pi(X)}\right) & \text{if } \pi(X) \neq 0; \\ 1, & \text{otherwise} \end{cases}
$$

Generate $A \sim U(0, 1)$;

13 **if** $A \leq \alpha(X, Y)$;

14 **then**

15 $\quad$ $X_{n+1} = Y$;

16 **else**

17 $\quad$ $X_{n+1} = X$;

18 **end**

19 **return** X_{n+1}.

1.4.3 The skipping chain

With reference to Algorithm 2, the *halting index* K is an independent random variable distributed $K \sim \mathcal{K}$ on $\mathbb{Z}_{>0} \cup \{\infty\}$. When $K = 1$, Y is a random-walk Metropolis proposal. Alternatively, if $K > 1$, the initial proposal Y is updated as described in

the previous section using the *skipping chain* $\{Z_k\}_{k\geq 0}$ on $\mathbb{R}^d$, constructed so that $Z_0 := X$, with $X = X_n$ as the current state of the chain and following the update rule:

$$Z_{k+1} := Z_k + \Phi R_{k+1}, \quad k \geq 0, \tag{1.16}$$

where

$$\Phi := \frac{Y - X}{||Y - X||} \tag{1.17}$$

$$R_1 := ||Y - X||, \tag{1.18}$$

where the *distance increments* $\{R_k\}_{k\geq 2}$ are independent realisations from the distribution of the radial component $||Y - X||$ conditioned on the angular component Φ. Let T_C be the first entry time of the skipping chain $\{Z_k\}_{k\geq 0}$ into the support C. Specifically,

$$T_C := \min\{k \geq 1 : Z_k \in C\} \tag{1.19}$$

where $\min\{\emptyset\} := \infty$.

Assumption 3. *The support $C = supp(\pi)$ and distribution $\mathcal{K}$ are such that $\mathbb{E}[T_A \wedge K] < \infty$ where $T_C \wedge K$ denotes the smaller of the two indices T_A and K (page 3, [76]).*

The skipping proposal $Z := Z_{T_C \wedge K}$ output by Algorithm 2 is well defined as $Y \neq X$ since q is a density and $T_C \wedge K < \infty$ by Assumption 3 in [76].

Proposition 2.1:

i *Algorithm 2 is a symmetric Metropolis-class algorithm on the domain C. That is, there exists a transition density $q_\mathcal{K}$ which depends on the halting index distribution $\mathcal{K}$ satisfying $q_\mathcal{K}(x, z) = q_\mathcal{K}(z, x)$ for all $x, z, \in C$ such that Algorithm 2 is $MH(\pi, q_\mathcal{K})$;*

ii *The inequality $q_\mathcal{K}(x, z) \geq q(z - x)$ holds for every $x, z \in C$. (page 4, [75]).*

1.4.4 Choice of underlying proposal q

In addition to assumptions 2 and 3, it is sufficient that a Metropolis-Hastings algorithm with target π and proposal density q (denoted here and in [75] as $MH(\pi, q)$) be π-irreducible to ensure the Strong Law of Large Numbers holds. This is true when π is continuous and bounded, and $q > 0$ everywhere. For the skipping sampler, the authors suggested q can be chosen as if it was to be used in a random walk

Metropolis algorithm, with the aim of achieving an acceptance rate of approximately 25% of proposals, although the authors note that empirical results indicate that a lower acceptance rate of 15% may stimulate skipping.

1.4.5 Proposal density q and the jump increment R_i

Since radial jump increments $R_1, R_2, \ldots$ are sampled conditioned on the event $\{\Phi = \varphi\}$, q should be chosen such that samples can be efficiently drawn from

$$||Y - X|| \text{ conditional on } \{\Phi = \varphi\} \tag{1.20}$$

for all φ in the unit sphere $\mathbb{S}^{d-1}$ with $\{z \in \mathbb{R}^d : |z| = 1\}$ [75]. Radially symmetric densities are a natural choice for q, so that conditioning on Φ is not necessary. A second option for q is $q \sim \mathcal{N}(0, \Sigma)$ for some $d \times d$ covariance matrix Σ, where, given direction $\Phi = \varphi$, R_i follows a generalised Gamma distribution with density given:

$$R_i|\Phi = \varphi \sim \frac{(\varphi^T \Sigma^{-1} \varphi)^{\frac{d}{2}}}{2^{\frac{d}{2}-1}\Gamma(\frac{d}{2})} r^{d-1} e^{-(\varphi^T \Sigma^{-1}\varphi)\frac{R^2}{2}}. \tag{1.21}$$

A third option provided by the authors was that q could be chosen indirectly by instead directly choosing the unconditional distribution of Φ and the conditional distribution $R|\Phi$.

A final suggestion by the authors is highlighted when sampling the conditional distribution $R|\Phi$ is computationally expensive. In such cases, $R_k = R$ for all k, requiring only a single realisation of R to be generated. This will lead the skipping chain to skip in the direction Φ with jump increments of equal size.

1.4.6 Anisotropy in C

In the implementation of the skipping sampler, if the support C has known anisotropy, the angular component of the proposal density can be selected to favour certain directions φ over others. This could be achieved, for example, by adjusting the covariance matrix of a Gaussian proposal function. It was noted this technique may be particularly beneficial in higher dimension problems, where the skipping chain may fail to enter C with high probability. This "biasing" of the selection of Φ requires the Metropolis acceptance ratio used by the skipping sampler to additionally depend on the ratio of the angular densities. Letting $q_\varphi(x, \phi)$ be the density of direction ϕ at the location x, for $\Phi = \frac{Y_{n+1} - X_n}{||Y_{n+1} - X_n||}$ then the acceptance probability is given by Equation (1.22).

$$\alpha(X_n, Y_{n+1}) = \min\left(\frac{\pi(Y_{n+1})q_\varphi(Y_{n+1}, -\Phi)}{\pi(X_n)q_\varphi(X_n, \Phi)}\right). \tag{1.22}$$

This result will be revisited in Section 6.3 which discusses using the skipping proposal with non-uniform sampling of Φ as the transition kernel of a sequential Monte Carlo method. This is intended to improve sampling from π in high-dimension problems.

1.4.7   Choice of $\mathcal{K}$

The authors of [75] suggest a number of choices for the distribution of the halting index. The simplest option is to employ a non-random halting index $K \equiv k_s \in \mathbb{Z}_{>1}$, applying a vector of the k_s skips to be applied to states Z_i for $i = 1, \ldots, k_s$ of the skipping chain respectively. The advantage of this choice is it could permit the user to evaluate whether $Z_i \in C$ in parallel, allowing potential running speed of the skipping sampler to approach that of the random walk Metropolis algorithm.

In cases where some knowledge of the geometry of C^c is known, the choice of q can influence the choice of K. The authors provide an example where the upper bound D of the diameter of C^c is known. In this case, $k_s = \frac{D}{sup_\varphi \sigma_\varphi}$ where σ_φ is the standard deviation of the conditional jump density in the direction φ. Alternatively, if C has a known anisotropy, the halting index can constructed to be direction-dependent while ensuring $\mathcal{K}_\varphi = \mathcal{K}_{-\varphi}$ for each $\varphi \in \mathbb{S}^{d-1}$ to maintain the symmetry of the proposal. The authors do caution that an unbounded distribution $\mathcal{K}$ should only be considered if C^c is known to be bounded due to the potentially high computational costs in C is not re-entered.

1.5   Alternate proof of Proposition 5.1 from [75]

In this section, we provide alternative proofs for Proposition 5.1 and Lemma 5.2 from [75]. We first present the transition kernel of the skipping chain for states $x \neq y$ as well as definitions for the *killed skipping chain* and the *stopped skipping chain* from [75].

Definition 1.5.1. For every $z \in \mathbb{R}^d \backslash \{x\}$, set $r_z = ||z - x||$ and $\varphi_z = \frac{z-x}{||z-x||}$.

$$\xi_1(x, z) := q_{r|\varphi}(r_z|\varphi_z)$$
$$\xi_k(x, z) := \int \xi_{k-1}(x, x + t\varphi_z) \mathbb{1}_{C^c}(x + t\varphi_z) q_{r|\varphi}(r_z - t|\varphi_z) dt,$$

where $\xi_1(x, z)$ is the density of proposing z in one step conditional on the chain is currently at x, conditioned on direction φ_z. Similarly, $\xi_k(x, z)$ is the density of proposing state z in k steps, conditional on the chain's current state being x and the sampled direction φ_z, and follows from the Chapman-Kolmogorov equation.

With reference to the skipping chain discussed in section 1.4, the authors of [75] also present the following related chains:

1.

Definition 1.5.2 (Killed skipping chain). Letting $\Delta \notin \mathbb{R}^d$ be a cemetery state, the *killed skipping chain* $\{\tilde{Z}_k\}_{k\geq 1}$ on $\mathbb{R}^d \cup \Delta$ is defined as

$$\tilde{Z}_k = \begin{cases} Z_k, & k \leq T_C \\ \Delta, & k > T_C \end{cases}$$

and

2.

Definition 1.5.3 (Stopped skipping chain). The *stopped skipping chain* $\{S_k\}_{k\geq 1}$ on $\mathbb{R}^d$ is given by

$$S_k := Z_{T_C \wedge k},$$

where $T_C \wedge k := \min T_C, k$.

Proposition 5.1 (page 13 [75]):

Conditional on $\{\Phi = \varphi\}$, the state $\tilde{Z}$ of the killed skipping chain has a density with respect to Lebesgue measure on the line $L_\varphi := \{x + t\varphi : t \geq 0\}$. This conditional density is given by the function $t \mapsto \xi_k(x, x + t\varphi)$. That is, for all $A_r \in \mathcal{B}([0,\infty])$ and $k \geq 1$:

$$\mathbb{P}[\tilde{Z}_k \in x + A_r\varphi | \Phi = \varphi] = \int_{A_r} \xi_k(x, x + t\varphi)dt$$

An alternate, simple proof of Proposition 5.1, not found in [75], is provided below:

Proof. For $0 \leq s \leq t$, let $P_k(s,t)$ be the k-step transition density of the killed skipping chain on the line L_φ with the given parametrization. Then, by the Chapman-Kolmogorov equation:

$$P_k(s,t) = \int_s^t P_{k-1}(s,u)P_1(u,t)du$$

Let:

$$C_u = \{u \in (s,t) : x + u\varphi \in C\}$$

Then by construction:

$$P_1(u,t) = \begin{cases} 0 & if\, u \in C_u \\ q_{r|\varphi}(t-u|\varphi) & if\, u \notin C_u \end{cases}$$

Hence,

$$P_k(s,t) = \int_s^t P_{k-1}(s,u)\mathbb{1}_{C_u^c}(u)P_1(u,t)du$$

.

Suppose that the proposition is true, for $k = n-1$. Then setting $s = 0, t = r_z$ and $\varphi = \varphi_z$ we have:

$$\begin{aligned} P_n(0,r_z) &= \int_0^t P_{n-1}(0,u)\mathbb{1}_{C_u^c}(u)P_1(u,r_z)du \\ &= \int_0^t \xi_{n-1}(x,x+u\varphi_z)\mathbb{1}_{C^c}(x+u\varphi_z)q_{r|\varphi}(r_z-u|\varphi_z)du \\ &= \xi_n(x,z) \end{aligned}$$

$\square$

The characteristics of the stopped skipping chain $\{S_k\}$ are established in Lemma 5.2 of [75]:

Lemma 5.2:

For $x, z \in \mathbb{R}^d$ with $x \prec z$ and $x \neq z$ we have:

i *For $k \geq 2, \xi(x,z)=$*

$$\int \cdots \int_a \left(\prod_{i=1}^{k-1} q_{r|\varphi}(r_i|\varphi)\mathbb{1}_{C^c}\left(x + (\sum_{j=1}^{i} r_j)\varphi\right) \right) q_{r|\varphi}(r_z - \sum_{i=1}^{k-1} r_i|\varphi)dr_{k-1}...dr_1 \quad (1.23)$$

where:

$$a = r_1 \prec \sum_{i=1}^{2} r_i \prec ... \prec \sum_{i=1}^{k-1} r_i \leq r_z$$

ii *The identity $\xi_k(x,z) = \xi_k(z,x)$ holds for every $k \in \mathbb{N}$,*

A simple proof of lemma 5.2, not presented in [75], is presented below:

i *Proof.* Suppose the statement holds for $k = n - 1$. Then by definition:

$$\xi_n(x, n) = \int \xi_{n-1}(x, x + t\varphi)\mathbb{1}_{C^c}(x + t\varphi)q_{r|\varphi}(r_z - t|\varphi)dt$$

$$= \int_t \int \cdots \int_{0 \leq t_i, \sum_{i=1}^{n-2} t_i \leq r_z} \prod_{i=1}^{n-2} q_{r|\varphi}(t_i|\varphi)\mathbb{1}_{C^c}\left(x + \left(\sum_{j=1}^{i} t_j\right)\varphi\right)q_{r|\varphi}(t - \sum_{i=1}^{n-2} t_i|\varphi)dt_{n-2}...dt_1$$

$$\times \mathbb{1}_{C^c}(x + t\varphi)q_{r|\varphi}(r_z - t|\varphi)$$

If we let $t_{n-1} := t - \sum_{i=1}^{n-2} t_i$. The we obtain by substitution:

$$\xi_n(x, z) = \int \cdots \int_{0 \leq t_i, \sum_{i=1}^{n-1} t_i \leq r_z} \prod_{i=1}^{n-1} q_{r|\varphi}(t_i|\varphi)\mathbb{1}_{C^c} \times$$

$$\left(x + \left(\sum_{j=1}^{i} t_j\right)\varphi\right)q_{r|\varphi}(r_z - \sum_{i=1}^{n-1} t_i|\varphi)dt_{n-1}...dt_1$$

as required.

$\square$

ii *Proof.* Exchange the roles of x and z in (i), noting that:

$r_x := |x - z| = |z - x| = r_z$ and $\varphi_x = \frac{x-z}{|x-z|} = -\frac{z-x}{|z-x|} = -\varphi_z,$

by assumption 2:

$$q_{r,\varphi}(r, -\varphi) = q(x - r\varphi, x)r^{d-1}$$

$$= q(x, x + r\varphi)r^{d-1} = q_{r,\varphi}(r, \varphi)$$

.

Thus $q_\varphi(\varphi) = q_\varphi(-\varphi)$ and $q_{r|\varphi}(r|\varphi) = q_{r|\varphi}(r| - \varphi)$.

Then:

$$\xi_n(z, x) = \int \cdots \int_{0 \le t_i, \sum_{i=1}^{n-1} t_i \le r_z} \prod_{i=1}^{n-1} q_{r|\varphi}(t_i| - \varphi_z) \mathbb{1}_{C^c}\left(z - \left(\sum_{j=1}^{i} t_j\right)\varphi_z\right) \times$$

$$q_{r|\varphi}(r_z - \sum_{i=1}^{n-1} t_i| - \varphi)dt_{n-1}...dt_1$$

Setting $s_1 = r_z - \sum_{i=1}^{n-1} t_i, s_2 = t_{n-1}, s_3 = t_{n-2}, ...s_{n-1} = t_2,$

we have $r_z - \sum_{i=1}^{n-1} s_i = t_1$ and:

$$z - t_1\varphi_z = z - (r_z - \sum_{i=1}^{n-1} s_i)\varphi_z = (z - r_z\varphi_z) + (\sum_{i=1}^{n-1} s_i)\varphi_z$$

$$= x + (\sum_{i=1}^{n-1} s_i)\varphi_z$$

also:

$$z - (\sum_{i=1}^{2} t_i)\varphi_z = x + (\sum_{i=1}^{n-2} s_i)\varphi_z$$

and so on...

Therefore, $\xi_n(x, y) = \xi_n(y, x)$. $\qquad \square$

1.6 The doubling sampler

The authors of [76] present a methodology to accelerate the skipping sampler under the restriction that C^c is convex referred to in this section as the *doubling sampler*. Summarising [76], since the states of the skipping chain are the partial sum $Z_k = x + \Phi \sum_{i=1}^{k} R_i$ where R_i are iid distance increments and $R_1 = |Y - x|$, the convexity of C^c induces an ordering on the skipping chain:

$$\begin{cases} Z_k \in C^c, & \text{if } k < T_C, \\ Z_k \in C, & \text{if } k \geq T_C, \end{cases} \tag{1.24}$$

where T_c is the first entry time of the skipping chain into the support C. Additionally, the convexity of C^c implies the halting index can such that $P[K_\phi = \infty] = 1$.

With reference to Algorithm 2, T_C (1.19) is determined by sampling the partial sums $\{R_k\}_{k \geq 1}$ sequentially. Alternatively, the skipping sampler algorithm can be accelerated if, for any k, $\sum_{i=1}^{k} R_i$ can be sampled directly, both unconditionally and conditioned on the value of $\sum_{i=1}^{2k} R_i$ at a comparable cost to sampling R_1. For example, this is possible when the distance increments R_i are exponentially distributed.

One strategy to develop a Metropolis-class algorithm which uses this accelerated skipping proposal (henceforth called the doubling proposal $q_d(x, y)$) is to design the doubling proposal such that it it has the same density as the skipping proposal q_κ. Defining

$$S_C := \min\{k \geq 1 : Z_{2^k - 1} \in C\}, \tag{1.25}$$

the doubling proposal would search forward through the exponential subsequence $Z_1, Z_3, Z_7, \ldots, Z_{2^k - 1}, \ldots$ until $k = \tilde{k} = S_C$ such that $Z_{2^{\tilde{k}} - 1} \in C$, then perform a logarithmic search of the sequence $Z_{2^{\tilde{k}-1} - 1}, \ldots, Z_{2^{\tilde{k}} - 1}$ to identify T_C. We then sample Z_m for $m = 2^{\tilde{k}-1} - 1 + 2^{\tilde{k}-2}$; if $Z_m \in C$, the sequence is reduced to $Z_{2^{\tilde{k}-1} - 1}, \ldots, Z_m$. Alternatively, if $Z_m \in C^c$, the subsequent sequence $Z_m, \ldots, Z_{2^{\tilde{k}} - 1}$ is searched. This reduction in the sequence continues in a bisection-like manner until T_C is located, in a process we will refer to as 'backtracking'.

1.6.1 The backtracking algorithm

Let T_{k+1} denote the $k + 1^{st}$ 'doubled' jump increment. Then:

$$T_{k+1} \sim R_1 + R_2 + \ldots + R_{2^k}, \tag{1.26}$$

where $R_i \sim \exp(\lambda)$ for $\lambda \in \mathbb{R}^+$. It follows that $T_{k+1} \sim \Gamma(2^k, \lambda)$, and the update rule for state $Z_{2^{k+1} - 1}$ is given by:

$$Z_{2^{k+1} - 1} = Z_{2^k - 1} + T_{k+1} \Phi. \tag{1.27}$$

If $Z_{2^k-1} \in C^c$ with T_k and $Z_{2^{k+1}-1} \in C$ with T_{k+1}, then in the backtracking step, we first simulate $Z' = Z_{2^{k+1}-1} - T_{k+1}^{(1)}\Phi$ where $T_{k+1}^{(1)} \sim R_1 + R_2 + ... + R_{2^{k-1}}$. We now need the distribution of $T_{k+1}^{(1)}$ conditional on the result of T_{k+1}. As the unconditional distribution of $T_{k+1}^{(1)}$ is a Gamma distribution with parameters $\Gamma(2^{k-1}, \lambda)$, then by Bayes' Theorem:

$$P(T_{k+1}^{(1)} \in dx | T_{k+1} \in dy) = \frac{P(T_{k+1}^{(1)} \in dx, T_{k+1} \in dy)}{P(T_{k+1} \in dy)}$$

$$= \frac{P(T_{k+1}^{(1)} \in dx, \tilde{T}_{k+1}^{(1)} \in d(y-x))}{P(T_{k+1} \in dy)}, \qquad (1.28)$$

where $\tilde{T}_{k+1}^{(1)} \sim R_{2^{k-1}+1} + ... + R_{2^k}$ has the $\Gamma(2^{k-1}, \lambda)$ distribution (independent of $T_{k+1}^{(1)}$). Letting $m = 2^{k-1}$, then (1.28) can be written as:

$$P(T_{n+1}^{(1)} \in dx | T_{n+1} \in dy) = \frac{\Gamma(2m)}{\Gamma(m)^2} \frac{(\lambda^m x^{m-1} e^{-\lambda x})(\lambda^m (y-x)^{m-1}) e^{-\lambda(y-x)}}{\lambda^{2m} y^{2m-1} e^{-\lambda y}}$$
$$= \frac{\Gamma(2m)}{y\Gamma(m)^2} \left(\frac{x}{y}\right)^{m-1} \left(1 - \frac{x}{y}\right)^{m-1}.$$

It follows that $\frac{x}{y}$ follows the $\beta(2^{k-1}, 2^{k-1})$ distribution. The backtracking algorithm which conducts the logarithmic search of $Z_{2^{\tilde{k}-1}-1}, \ldots, Z_{2^{\tilde{k}}-1}$ is formalised in Algorithm 3.

Algorithm 3: Backtracking Algorithm

1 **INPUT:** k, $Z_{2^k-1} \in C$ and $T_k \sim R_1 + \cdots + R_{2^{k-1}}$;
2 Set $T^{(0)} = T_k$ and $Z' = Z_{2^k-1}$;
3 Set $i = 1$;
4 **while** $i \leq k$ **do**
5   Sample $\omega \sim \beta(2^{k-i}, 2^{k-i})$ and calculate $T^{(i)} = \omega T^{(i-1)}$;
6   **if** $Z' \in C$ **then**
7   Set $Z' = Z' - T^{(i)}\Phi$;
8       $Z = Z'$;
9   **else**
10       $Z' = Z' + T^{(i)}\Phi$;
11   **end**
12   Set $i = i + 1$;
13 **end**
14 Set $Y = Z$;
15 **OUTPUT:** Y

With this, it is possible to specify an algorithm for a doubling sampler MCMC in Algorithm 4:

Algorithm 4: Doubling skipping sampler

---

1 Initialisation: The nth sample: $X_n \in \mathbb{R}^d$; $R \sim \exp(\lambda)$ for $\lambda \in \mathbb{R}^+$;

2 Set $X = X_n$;

3 Generate the initial proposal Y according to the density $q(u - X)du$;

4 Calculate the direction $\phi = \frac{Y-X}{|Y-X|}$;

5 Set $k = 1$ and $Z_1 = Y$;

6 **while** $Z_{2^k-1} \in C^c$ **do**

7 $\quad$ Generate 'doubled' jump increment $T_{k+1} \sim R_1 + R_2 + ...R_{2^k}$ distributed according to $T_{k+1} \sim \Gamma(2^k, \lambda)$;

8 $\quad$ Set $Z_{2^{k+1}-1} = Z_{2^k-1} + T_{k+1}\Phi$;

9 $\quad$ $k = k + 1$;

10 **end**

11 **Backtracking algorithm:(Input: k, $Z_{2^{k+1}-1}$ and T_{k+1}. Output: Y);**

12 Compute the acceptance probability α:

13

$$\alpha(X, Y) = \begin{cases} \min\left(1, \frac{\pi(Y)}{\pi(X)}\right) & \text{if } \pi(X) \neq 0; \\ 1, & \text{otherwise} \end{cases}$$

14 Generate a uniform random variable $v \sim U[0, 1]$;

15 **if** $v \leq \alpha(X, Y)$ **then**

16 $\quad$ $X_{n+1} = Y$;

17 **else**

18 $\quad$ $X_{n+1} = X_n$;

19 **end**

20 Set $n = n + 1$ and repeat from step 1;

Proposition 2.2: *The doubling proposal described in Algorithm 4 has the same density as the skipping proposal q_κ.*

Proof. Let q_d denote the doubling proposal density and q_κ be the skipping proposal density. The authors of [75] have already proven that q_κ is reversible.

Consider both the state D proposed by the doubling proposal after $K_d = k$ steps, and the state of the S proposed by the skipping proposal when the same 'doubled' steps $K_s = k$ are inspected, that is, we inspect the state of the skipping chain after the $(2^{k-1} - 1)^{st}$ steps. Following from the convexity of C^c, $\mathbb{P}(S \in dx)$ is given by Equation (1.29):

$$\mathbb{P}(S \in dx) = \sum_{k=1}^{\infty} \mathbb{P}(S \in dx | K_s = k)\mathbb{P}(K_s = k). \tag{1.29}$$

Likewise, $\mathbb{P}(D \in dx)$ is given by Equation (1.30):

$$\mathbb{P}(D \in dx) = \sum_{k=1}^{\infty} \mathbb{P}(D \in dx | K_d = k)\mathbb{P}(K_d = k). \tag{1.30}$$

By construction of each proposal, it follows that $\mathbb{P}(K_d = k) = \mathbb{P}(K_s = k)$. To

demonstrate that $\mathbb{P}(D \in dx | K_d = k) = \mathbb{P}(S \in dx | K_s = k)$, first consider the proposed state S generated by the skipping proposal, observed at the $(2^{k-1} - 1^{st})$ step. Let $M = \{1, ..., 2^{k-1}\}$ denote the number of distance increments R_i between the $2^{k-1} - 1^{st}$ and the $2^k - 1^{st}$ observations. For each $M = m$, the event $\{S \in dx | K_d = k\}$ is given by (1.31):

$$
\{S \in dx | K_s = k\} = \begin{cases} \{S \in dx, M = 1 | K_d = k\}, \\ \{S \in dx, M = 2 | K_d = k\}, \\ \qquad \vdots \\ \{S \in dx, M = 2^{k-1} | K_d = k\} \end{cases} . \tag{1.31}
$$

It follows:

$$
\mathbb{P}(S \in dx | K_d = k) = \sum_{i=1}^{2^{k-1}} \mathbb{P}(S \in dx, M = i | K_d = k).
$$

Similarly, denote the steps of the skipping sampler associated with the event defined in Equation (1.31) by $Z_{k,m}$ for $m = 1, ..., 2^{k-1}$. The convexity of C^c induces an ordering of $Z_{k,m}$ such that there exists $M = m^*$ such that:

$$
\begin{aligned} Z_{k,m^*} &\in C \\ Z_{k,m<m^*} &\in C^c \\ Z_{k,m>m^*} &\in C \end{aligned} .
$$

It follows that, by construction, $Z_{k,m*}$ is the proposal of the skipping chain.

Now we investigate the proposal of the doubling sampler D and the event:

$$
\{D \in dx | K_d = k_d\}.
$$

Further denoting steps of the doubling proposal D_k, if $K_d = k$, it follows that $D_{k-1} \in C^c$.

When $K_d = k$, $D_k = D_{k-1} + T_k \Phi$ where $T_k \sim R_1 + \cdots + R_{2^{k-1}}$. Let $X_{k,N}$ represent simulated partial sums between states D_{k-1} and D_k using R_i. Thus $X_{k,N} = D_{k-1} + \sum_{i=1}^{N} R_i$, where $\max(N) = 2^{k-1}$ and $X_{k,2^{k-1}} = D_k$. The backtracking algorithm conducts a logarithmic search of this sequence of partial sums to locate D. The event $\{D \in dx | K_d = k\}$, can therefore be expressed as:

$$\{D \in dx | K_d = k\} = \begin{cases} \{D \in dx, N = 1 | K_d = k\} \\ \{D \in dx, N = 2 | K_d = k\} \\ \quad \vdots \\ \{D \in dx, N = 2^{k-1} | K_d = k\} \end{cases}.$$

It follows:

$$P(D \in dx | K_D = k) = \sum_{i=1}^{2^{k-1}} P(D \in dx, N = i | K_D = k).$$

Additional, by the convexity of C^c and the construction of the backtracking algorithm, there exists n^* such that:

$$\begin{cases} X_{k,n^*} \in C; \\ X_{k,N<n^*} \in C^c \\ X_{k,N>n^*} \in C \end{cases}$$

It follows that, $D = X_{k,n^*}$, and $S|K_d \sim D|K_d$. Thus $P(D \in dx | K_d = k) = P(S \in dx | K_s = k)$ and $P(S \in dx) = P(D \in dx)$.

$\square$

1.6.2 Numerical Results

The *doubling sampler* algorithm was used to draw samples from the distribution $\pi = \frac{\rho \mathbb{1}_C}{\rho(C)}$, where $\rho \sim \mathcal{N}([0,0,0], 100 \times \mathbb{I}_3)$ where $\mathbb{I}_3$ is the 3×3 identity matrix. The support of π is $C := \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 > 50^2\}$, constructed so that C^c is convex. Additionally, the choice of ρ and C are such that the target density π is symmetric about the origin, with $E(X_i) = [0,0,0]$. The doubling sampler algorithm was initiated at $X_0 = [-50, -50, -50]$ and conducted $N = 50,000$ proposals with unconditional jump density $R \sim \exp(10)$. Thus, by the construction of Algorithm 4, $T_k \sim \Gamma(2^{k-1}, 10)$.

For comparison, a random-walk Metropolis algorithm and a skipping sampler algorithm were also applied to draw samples from π. Both were initiated at the same state X_0 as the doubling sampler and conducted 50,000 proposals with underlying density $q(X_n, X_{n+1}) \sim \mathcal{N}(X_n, 5 \times \mathbb{I}_3)$. The skipping sampler employed a halting regime of $K = 6$. Each algorithm was run for 200 independent simulations initiated at X_0. This allows for inferences to be made about the distribution of sample means.

Table 1.1 presents summary statistics for the samples generated by the three

algorithms described above. The average of the sample means generated by the random-walk Metropolis simulations was [-9.8,-21.7,-26.6], while the average distance between the sample means and $E(X)$ was 34.9 units. These results suggest limited exploration of the state space, with the signs of the components of the initial state $X_0 = [-50, -50, -50]$ correlating with those of the average of the sample means. The average MSJD also reflects the limited exploration of the state space, as the average displacement provided by the proposal was 2.76 units, the least of all algorithms. This can be understood as a consequence of the local moves conducted by the random-walk proposal.

In contrast, the doubling sampler produced samples whose sample means had an average Euclidean distance from $E(X)$ of 3.1 units, the least of the three algorithms, as well as the largest average MSJD of 8.9 units. These suggest the doubling sampler was able to conduct global exploration of the state space by traversing C^c, allowing the algorithm to generate samples with averages closer to the expected value. Additionally, given that C^c was convex and $\mathbb{P}[K_\phi = \infty] = 1$, the doubling sampler was expected have proposals with larger displacements than the skipping sampler. This is reflected by the higher MSJD of the doubling sampler over that of the skipping sampler. The average time taken also highlights the expected benefit of the doubling sampler, taking less time, on average, than the skipping sampler to complete the simulation, exhibiting the computational advantage over the skipping sampler as desired by the authors of [76].

Table 1.1: The performance metrics of each algorithm: RW-M: a random walk Metropolis-Hastings algorithm; Skip: the skipping sampler; D-Skip: the doubling sampler. For the metrics: Avg. $\hat{\mu}$ average sample means from simulations, Avg ED: average Euclidean distance between sample means and $E(X)$; Avg. time: average execution time for each algorithm; Avg MSJD: average Mean Squared Jump Distance.

Metric	$RW - M$	$Skip$	$D - Skip$
Avg. $\hat{\mu}$	$[-9.8, -21.7, -26.6]$	$[-5.6, -23.2, -17.0]$	$[-2.9, 0.4, 0.7]$
Avg. ED	34.9	28.1	3.1
Avg. time (s)	5.23	7.84	6.08
Avg. MSJD	2.76	7.78	8.90

The reduction in execution time provided by the doubling sampler has already been noted in [76] and is illustrated in the comparative results in Table 1.1. Through its construction, the reduced time spent in C^c is especially beneficial when elements from the sample space are intended to be input parameters of a computationally demanding physical model of a natural or engineered system. While this method requires *a priori* knowledge that C^c is convex, this geometry is not uncommon in reliability studies. For example, in the case of random changes in power injections from renewable energy sources, destabilising effects on power system operations were observed for random power injections with large ($\gg 0$) deviations from equilibrium

conditions (see chapters 3 and 4). If it is desired to sample such random power injections, the doubling sampler may be suitable as C^c may be a convex subset of the state space of power disturbances.

Chapter 2

Power system modelling

NOMENCLATURE

Machine and line parameters	
A_i	Governor droop response;
ω_{sys}	Steady state frequency of rotating reference frame;
δ_{ij}	Phase angle difference: $(\delta_i - \delta_j)$;
M	System angular momentum;
δ_i	Electrical phase angle of machine i with respect to ω_{sys};
$\dot{\delta}_i$	Electrical angular velocity. This is 0 during steady state;
$\ddot{\delta}_i$	Rate of change of frequency (RoCoF);
D	System load damping factor;
χ_i^G	Mechanical power of machine i;
χ_i^L	Power consumption (Load) at node i
χ^e	Electrical power generated/consumed at machine i;
N	Total number of machines in the network;
ν_i, ν_j	q-axis transient voltages at machines i and j; $i, j \leq N$;
$E_{f,i}$	Rotor's field voltage of machine i;
B_{ij}	Line susceptance connecting nodes i and j;
$X_{d,i}$	d-axis reactance of machine i;
$X'_{d,i}$	d-axis transient reactance of machine i;
ρ_i	Mechanical power provided by governor at machine i ;
P_i^m	Maximum power output of generator i;
$P_{ne,i}$	Rated active power of machine i p.u.;
P_i^G	Initial equilibrium power of machine i;
$s_{ne,i}$	Primary frequency control droop of machine i ;
$T_{se,i}$	Relative servomotor time constant of machine i;
$T_{d,i}$	Transient time constant of the d-axis of machine i;
η_i	Power disturbance at node i;
Protection system parameters	
C	Load shedding increments (%)
F^+	Over Frequency Generation Shedding (OFGS) Threshold, Hz
F^-	Set of Progressive Under Frequency Load Shedding (UFLS) Thresholds, Hz
G	RoCoF Trip Threshold, Hz s^{-2}
P^ϕ	Line Trip Power Flow Deviation Threshold, MW
T^ϕ	Line Trip Relay Delay
Ω_{ij}	Indicator for line protection system
Ξ_i	Indicator for generation shedding emergency response
Γ_i	Indicator for load shedding emergency response

A power system refers to the collection of components responsible for producing electrical energy and delivering this to consumers. While such systems vary in size and complexity, they share some basic characteristics.

Synchronous generators are the main source of electrical energy in power systems, with power system stability philosophies largely focused on keeping inter-connected machines in synchronism [59]. These generators also play an important role in both frequency and voltage regulation. The descriptor "synchronous" refers to the synchronising torque between generators which aims to make all generators' rotors spin with the same angular velocity, known as the *system frequency*, which is maintained very close to a particular value depending on the geographical region, normally either 50 or 60 Hz [43].

The *prime movers* of synchronous machines are the turbines which convert primary energy sources (gas, coal, nuclear fuels or renewable sources) to mechanical energy, which is then converted to electrical energy by synchronous generators [46]. The *network* is comprised of three-phase alternating current (AC) lines connecting components of the power system, sub-divided into the transmission, sub-transmission and distribution networks which operate at different but approximately constant voltages. Also found on the network are the transformers which regulate voltage between the different networks. *Loads*, including households, industry and commercial activities, convert electrical energy from the network to other forms for final consumption [57].

However, the rising prevalence of weather-dependent wind and solar-based renewable energy systems (RES) increases the challenge of maintaining frequency and voltage stability. Wind and solar RES are considered *low inertial* generation systems, i.e. they do not employ large spinning masses to generate electricity [110], and can therefore induce rapid "power disturbances"- random changes in power injected to the network. If power disturbances are sufficiently large, these can lead to *contingencies*- a significant deviation in frequency from its desired value, which can be damaging to sensitive equipment [42]. While all electrical networks employ *regulatory systems* (section 2.4) to maintain system frequency and *emergency responses* (section 2.5) which disconnect power system components to safeguard sensitive equipment, RES-induced contingencies may increase the risk of emergency responses and the loss of service to customers.

The risk of RES-induced power disturbances to frequency stability is revealed by empirical studies which have shown wind energy feed-in to have a measurable effect on frequency deviations on short time-scales (< 1s), below the response time of frequency regulation systems [39]. While the effects of RES intermittency may be expected to diminish with the aggregation of RES sites, in networks with high prevalences of RES, intermittency is still observed even under country-wide averaging, thus impacting frequency stability. Furthermore, the distribution of short-term

frequency deviations are non-Gaussian when subject to intermittent wind power generation, possessing heavier tails than those predicted by Gaussian models, representing a higher rate of large deviations, particularly in smaller networks. In larger networks the non-Gaussian behaviour in frequency deviations vanish on time scales longer than 1s, but in smaller more isolated networks, such heavy-tailed behaviour persists and remain detectable on time scales up to 10s [97].

Drawing a sample of power disturbances conditioned on the event of the activation of an emergency response is a key research objective of this thesis. Power systems are designed to be resilient against the occurrence of contingencies, thus the set of power disturbances sufficiently large and well-located to trigger emergency responses occur infrequently, and can be considered a *rare event*. In the context of power systems and this thesis, rare events are those whose probability of occurrence are very small, but which can have an adverse impact on the operation of the network and a loss of service to customers (i.e. a blackout) [16]. Rare events of interest in power systems investigated by this thesis include frequency deviations beyond the nominal operating band and the cascading disconnection of generating units, transmission lines or loads. However, studying such phenomena via simulation-sampling coupling can be computationally challenging - due to their infrequent occurrence, Monte Carlo sampling requires a large number of realisations to generate even a single instance of a rare event [11, 22]. One approach to this problem is to apply specialised mathematical techniques for the study of rare events, such as those reviewed in Chapter 1. In this thesis we focus on the application of the *skipping sampler* described in Section 1.4 to draw samples of power disturbances conditioned on the occurrence of specific contingencies in the network. These samples can be used to understand the potential vulnerabilities of a power system.

While detailed power system models which describe frequency dynamics are available, their complexity would consume significant resources to conduct the required number of simulations. Conversely, simplified models for frequency dynamics also exist, usually using the *swing equation* [58]. The evaluation of these models requires less time, but produce results which are only valid for small disturbances, and only for approximately one second ($t = 1$) following a simulated power disturbance on the network [57]. To mediate between the qualities of simplicity and reliability, this chapter will detail a novel power system model which can generate frequency dynamics for $t > 1$ seconds following a simulated power disturbance, but which can also be evaluated sufficiently quickly to allow the rare-event sampling methods discussed in Chapter 1 to be feasibly applied.

This chapter is organised as follows: a model for the electrical network and its transmission lines is provided in Section 2.1. The *classical model* and *third order model* for the synchronous generator and motor are presented Sections 2.2 and 2.3. Frequency regulation is introduced in Section 2.4 along with the model for governor

action following a power disturbance on the network. Section 2.5 presents the models for power system emergency responses intended to safeguard sensitive equipment, including generation shedding, load shedding and line disconnection. The impact of voltage regulation on frequency stability is explored in Section 2.6, and details a simple model for automatic voltage regulation. Finally, Section 2.7 presents the proposed power system model for rare event analysis which will be used in Chapters 3 and 4.

2.1 Network Model

Electrical power systems can be modelled as a graph $G(X, E)$ where the set of nodes X with $|X| = N$ represent the components of the network such as generators and loads. The set of edges E corresponds to the transmission lines connecting these nodes [100]. Transmission lines have an *impedance* calculated as $Z_{ij} = R_{ij} + jX_{ij}$ [57], where R_{ij} and X_{ij} are resistance and reactance respectively for the line connecting nodes i and j.

Each edge in the network also has a complex valued *admittance* given by $Y_{ij} = \frac{1}{Z_{ij}} = G_{ij} + jB_{ij}$, where G_{ij} is the conductance and B_{ij} the susceptance of line ij [100]. An *admittance matrix*, $Y \in \mathbb{C}^{N \times N}$ records the line admittances of each line, which are used in Kirchoff's and Ohms laws [59]:

$$I = Y\nu, \tag{2.1}$$

where $I, \nu \in \mathbb{C}^n$ are vectors of nodal currents and voltages respectively.

The electrical network is responsible for transmitting the power released by generators- their *apparent power*, to loads across the network. A generator's apparent power can be decomposed into two components- active power and reactive power, whose flows through the network are influenced by different control actions and have effects on different components of the network. *Active power* is closely related to frequency regulation and control, while *reactive power* is associated with voltage regulation [57]. As our focus is to understand rare phenomena in frequency regulation, this section will focus on modelling active power flows in the network, and its interaction with frequency regulation systems and emergency responses. The apparent power S at each node j is given by:

$$S_j = \nu_j I_j^*, \tag{2.2}$$

where I^* is the complex conjugate of nodal currents, $\nu_j = |\nu_j|e^{i\delta_j}$, $I_j = \sum_k Y_{jk}\nu_k$ and δ_j is the *electrical phase angle*. Substituting $y_{ij} = |y_{ij}|e^{i\delta_{ij}} = G_{ij} + iB_{ij}$ where y_{ij} is the ij^{th} component of the admittance matrix, then the active power at node j is $\chi_j = \mathcal{R}e(S_j)$:

$$\chi_j = \sum_{i=1}^{N} |\nu_j||\nu_i|[G_{ij}cos(\delta_i - \delta_j) + B_{ij}\sin(\delta_i - \delta_j)]. \tag{2.3}$$

Under the assumption the network is lossless, i.e. the lines experience no loss of power, then $G_{ij} = 0 \forall i, j$ and the electrical power flow can simplified as:

$$\chi_j = \sum_{i=1}^{N} |\nu_j||\nu_i|B_{ij}\sin(\delta_i - \delta_j), \tag{2.4}$$

where $B \in \mathbb{R}^{N \times N}$ is the matrix of susceptances. Equation (2.4) will be used in Section 2.2 to derive the model for the synchronous generator and motor.

2.1.1 Kron reduction

We conclude this section by providing a discussion on *Kron reduction*, a network reduction method from graph theory [24]. Assume the set of nodes X can be partitioned into *active nodes* which produce current and *passive nodes* which do not. Kron reduction can be used to eliminate passive nodes and form a reduced network, while retaining the characteristics of the original. Kron reduction of electrical networks creates a reduced admittance matrix $Y_{red} := Y \in \mathbb{C}^{N \times N}$ derived from the network's admittance matrix Y_{ij} as follows.

Each line connecting node i and j is included in the Y matrix. Only nodes with generators will be sources of current, thus the current vector can be partitioned:

$$\mathbf{I} = \left[\frac{I_n}{\mathbf{0}} \right].$$

Following equation (2.1), matrices Y and ν can also be partitioned :

$$\left[\frac{I_n}{\mathbf{0}} \right] = \left[\begin{array}{c|c} Y_{nn} & Y_{nr} \\ \hline Y_{rn} & Y_{rr} \end{array} \right] \left[\frac{\nu_n}{\nu_r} \right],$$

where the subscript n is used to denote generator nodes and r denotes remaining nodes [4]. Expansion yields $I_n = Y_{nn}\nu_n + Y_{nr}\nu_r$ and $0 = Y_{rn}\nu_n + Y_{rr}\nu_r$. Eliminating ν_r yields:

$$I_n = (Y_{nn} - Y_{nr}Y_{rr}^{-1}Y_{rn})\nu_n. \tag{2.5}$$

The matrix $Y_{red} = (Y_{nn} - Y_{nr}Y_{rr}^{-1}Y_{rn})$ is the reduced admittance matrix associated with an *effective network model* or *reduced network*, which represents the interactions between generators where loads are treated as constant impedances.

2.2 Second order model

A key novelty of the power system modelling in this thesis is its cyber-physical dynamics: in addition to the electro-mechanical (that is, physical) dynamics, we also model the action of automated protective (that is, cyber) equipment. While the physical dynamics are smooth and described below by differential equations, the cyber dynamics introduce discontinuous changes to the parameters of these equations. In order to justify these parameter changes, it is necessary to understand the derivation of the electro-mechanical dynamics, and that is the aim of the present section.

There are several models for generators with differing levels of complexity and accuracy. The simplest model is the *swing equation*, a second-order differential equation based on Newton's equations of motion [43]. We develop this model below following from [59]:

The torques acting on the rotor of the generator can be classified as the *mechanical torque* T_m and the *electromagnetic torque* T_e. When there is an imbalance between the torques acting on the rotor, the net torque results in an acceleration:

$$T_a = T_m - T_e, \tag{2.6}$$

where:

T_a is the accelerating torque in Newton metres Nm;

T_m is the mechanical torque acting on the rotor in Nm

T_e is the electromagnetic torque in Nm.

When modelling active power generation, we set $T_m > 0$, while $T_m < 0$ for motors which consume active power. The equation of motion for the rotor is given by Equation (2.7):

$$J\frac{d\omega_m}{dt} = T_a = T_m - T_e, \tag{2.7}$$

where:

J is the combined moment of inertia of the generator and turbine in $kg\ m^2$

ω_m is the angular velocity of the rotor, in mechanical rad/s;

t is time in seconds;

To simplify power system analysis, units of measurement for the parameters of power, voltage, current, impedance and admittance parameters are converted to the *per unit (p.u)* system, with all parameters expressed as multiples of selected base

units. Base units normally used include power, measured in Mega-Volt Ampere (MVA) and voltage, measured usually in kilovolts (kV). This allows the parameters used to describe network components to be specified in per unit values to ease calculations.

The *inertia constant H* is used to convert the equation of motion for the rotor into the *per-unit* system:

$$H = \frac{\text{stored energy at rated speed in MW.s}}{\text{MVA rating}}, \tag{2.8}$$

$$H = \frac{1}{2}\frac{J\omega_{0m}^2}{\text{MVA}_{base}}, \tag{2.9}$$

where ω_{0m} is the angular velocity in mechanical rad/s which the machine is rated to operate, and MVA_{base} is the base power of the machine.

Rearranging, the moment of inertia is given by:

$$J = \frac{2H}{\omega_{0m}^2}\text{MVA}_{base}. \tag{2.10}$$

Substituting (2.10) in equation (2.9) yields:

$$\frac{2H}{\omega_{0m}^2}\text{MVA}_{base}\frac{d\omega_m}{dt} = T_m - T_e. \tag{2.11}$$

Dividing throughout by $T_{base} = \frac{MVA_{base}}{\omega_{0m}}$ and rearranging yields:

$$2H\frac{d}{dt}\left(\frac{\omega_m}{\omega_{0m}}\right) = \frac{T_m - T_e}{T_{base}} \tag{2.12}$$

This gives the equation of motion for the rotor in *per-unit* form as:

$$2H\frac{d\bar{\omega}_r}{dt} = \bar{T}_m - \bar{T}_e. \tag{2.13}$$

$$\bar{\omega}_r = \frac{\omega_m}{\omega_{0m}} = \frac{\omega_r/p}{\omega_0/p} = \frac{\omega_r}{\omega_0}.$$

where:

ω_r is the angular velocity of the rotor in electrical rad/s

ω_0 is the rated speed of the generator;

p is the number of field poles.

Let the mechanical rotor angle δ_m denote the angular difference between the rotor and a reference axis rotating with synchronous angular velocity ω_{sys}. The electrical phase angle δ, measured in electrical radians with respect to the rotating reference axis, is given by $\delta = \frac{2\delta_m}{p}$, where p is the number of magnetic poles of the generator. The electrical phase angle at time t is given by:

$$\delta(t) = \omega_r t - \omega_0 t + \delta_0. \tag{2.14}$$

where δ_0 is the δ when $t = 0$. Taking the time derivative yields:

$$\frac{d\delta}{dt} = \omega_r - \omega_0 = \Delta\omega_r. \tag{2.15}$$

Taking the second derivative

$$\frac{d^2\delta}{dt^2} = \frac{d\omega_r}{dt} = \frac{d(\Delta\omega_r)}{dt}$$

$$= \omega_0 \frac{d\bar{\omega}_r}{dt} = \omega_0 \frac{d(\Delta\bar{\omega}_r)}{dt}. \tag{2.16}$$

Substituting for $d\bar{\omega}_r/dt$ in the above equation yields:

$$\frac{2H}{\omega_0} \frac{d^2\delta}{dt^2} = \bar{T}_m - \bar{T}_e. \tag{2.17}$$

An additional separate component for damping torque not accounted for in the calculation of T_e is usually included in the equation by adding a term K_D proportional to the speed of the rotor:

$$\frac{2H}{\omega_0} \frac{d^2\delta}{dt^2} = \bar{T}_m - \bar{T}_e - K_D \Delta\bar{\omega}_r, \tag{2.18}$$

where

$$\Delta\bar{\omega}_r = \frac{\Delta\omega_r}{\omega_0} = \frac{1}{\omega_0} \frac{d\delta}{dt};$$

yielding:

$$\frac{2H}{\omega_0} \frac{d^2\delta}{dt^2} = \bar{T}_m - \bar{T}_e - \frac{K_D}{\omega_0} \frac{d\delta}{dt}, \tag{2.19}$$

where K_D is the damping factor/coefficient in per unit torque/per unit speed deviation and $\omega_0 = 2\pi f$ where f is rated system frequency.

Many sources use angular momentum in the equation of motion for the generator [100]. Defined in relation to the inertia constant H:

$$M = \frac{2H}{2\pi f}\text{MVA}_G, \tag{2.20}$$

where MVA_G is the machine rated power in megavolt-amperes [54]. If the power is expressed in *per unit* form instead of megawatts, this result must be divided by the base power MVA_{base}, giving:

$$M = \frac{2H}{2\pi f}\frac{\text{MVA}_G}{\text{MVA}_{base}}. \tag{2.21}$$

In reality the angular momentum is not strictly constant as rotor speed varies during the swing following a disturbance. However, the change in speed due to the disturbance is so small in comparison to the regular speed of the rotor that it is reasonable to assume that M be treated as constant, known as the *inertia constant* [54].

This gives the form (in per unit basis):

$$M\frac{d^2\delta}{dt^2} = T_m - T_e - \frac{K_D}{\omega_0}\frac{d\delta}{dt}. \tag{2.22}$$

In the literature, the mechanical torque is provided by power from generators is usually denoted P_g and the electrical power extracted from the generator denoted P_e (see, for example [100]). However, for consistency with subsequent chapters of this thesis, we instead denote these quantities as χ^G and χ^e respectively. Thus, for a N-machine network with synchronous generators, this equation gives changes in the electrical angular velocity of the rotor of machine i and yields the *swing equation*:

$$M\ddot{\delta}_i = -D_i\dot{\delta}_i + \chi_i^G - \chi_i^e. \tag{2.23}$$

At each node, the active electrical power flow is given by $\chi_i^e = \sum_{j=1}^N \nu_i\nu_j B_{ij}\sin\delta_{ij}$, where $\delta_{ij} := \delta_i - \delta_j$ [98, 100]. The *classical model* for the generator's frequency dynamics can be presented:

$$M\ddot{\delta}_i = -D_i\dot{\delta}_i + \chi_i^G - \sum_{j=1}^N \nu_i\nu_j B_{ij}\sin\delta_{ij}. \tag{2.24}$$

The classical model in Equation (2.24) assumes constant nodal voltages ν_i, and is the reason why the reliability of results of transient stability analysis is limited to less than one second after a simulated disturbance [57]. Instead, the third order model given in Section 2.3 includes transient voltage of the generator, and will be used in this thesis to describe frequency dynamics of machines in the network.

2.3 Third order model

Because it includes the transient voltage of the generator, the third order model presented in this section is suitable for the modelling of larger power disturbances over timescales of multiple seconds. Although it nevertheless still neglects damper winding effects and the sub-transient equations found in the more complete *Park's model*, these omissions are often compensated by increasing the mechanical damping coefficient (D) of each machine [100].

To account for the temporal variations in loads in a network during response to a disturbance, loads will be modelled as *synchronous motors*, analogous to synchronous generators [100]. When modelling loads in this way, the sign of electrical power χ_i^e and mechanical power χ_i^G inputs are reversed in the swing equation (2.23), signifying that loads convert electrical power to mechanical power [100]. As we may desire to model both loads and generation at a single node, we instead use the term χ_i^L to represent the load at node i.

Power disturbances, as previously mentioned, represent changes in the power injected at any location in the network. We will model them as random changes in the net generation (or load) at each node. If the initial conditions of the system of differential equations are chosen such that the system is in equilibrium at $t = 0^-$ with system frequency equal to the desired frequency, then any non-zero disturbance $|\eta_i| > 0$ for $t \geq 0$ will lead to transient behaviour of the frequency.

The distribution of the vector of random power disturbances $\underline{\eta} \in \mathbb{R}^N$ for a N-node network is of importance to understanding their relationship with power system emergency responses, and will be discussed in greater detail in Chapters 3 and 4. Including load dynamics and the random power injection terms, the third order model is given below in equation (2.25):

$$
\begin{cases}
M\ddot{\delta}_i + D\dot{\delta}_i = \chi_i^G - \chi_i^L - \nu_i \sum_{j=1}^{N+L} B_{ij}\nu_j \sin(\delta_i - \delta_j) + \eta_i & \text{(2.25a)} \\[2em]
T_{d,i}\dot{\nu}_i = E_{f,i} - \nu_i + (X_{d,i} - X_{d,i}') \sum_{j=1}^{N+L} B_{ij}\nu_j \cos(\delta_i - \delta_j), & \text{(2.25b)}
\end{cases}
$$

where $T_{d,i}$ is the transient time constant of machine i, $X_{d,i}$ and $X_{d,i}'$ are the d-axis reactance and transient reactance respectively of machine i.

2.4 Frequency regulation

Frequency regulation refers to the systems and services which maintain network frequencies within pre-determined tolerances following a mismatch between active power generation and load demand. Such mismatches occur due to load fluctuations or random power disturbances. Frequency regulation is conducted by altering

active power output of generators to ensure grid stability [54]. These systems and services are normally classified as *primary, secondary* or *tertiary frequency control* depending on the delay of the response following a contingency [54].

- *Primary frequency control* involves rapid local response (first few seconds) of automatic systems to deviations from system frequency, for example through generator governor units, controllable loads [57] or more recently grid-scale batteries, with the aim of halting any frequency excursions and establishing a temporary steady-state.

- *Secondary frequency control* attempts to restore the system frequency to pre-disturbance values through a centralised network approach-i.e. by changing the electrical outputs of generators across the network coordinated by an automatic generation control (AGC) unit. This occurs between the first 5 - 20 seconds following a contingency.

- *Tertiary frequency control* refers to network operators dispatching additional generation to the network [54]. Like secondary control, tertiary control is centralised, but does not require as fast a response as secondary control, occurring on the order of several seconds to minutes following a contingency.

Synchronous generators, with their large rotating masses, already provide a measure of frequency control through the inertia response following a contingency [27]. This acts to slow the rate of change in frequency and is modelled by the inertia constant M. The following section describes the generator's governor, which actively regulates frequency following a contingency.

2.4.1 Governor model

Generator governors adjust active power output of the generator in order to maintain the desired system frequency value [59]. Governors are therefore a key component of primary frequency control [78]. While the prime movers of generators convert the thermal energy to mechanical energy, the governing unit of the prime mover can adjust its power in response to changes in frequency, and also forms an integral component of a centralised automatic generation control (AGC) mechanism which regulates system frequency across the network. A simple model to evaluate system frequency dynamics under governor action during primary frequency control is derived from [64]:

$$\dot{\rho}_i = -A_i \dot{\delta}_i, \tag{2.26}$$

49

where ρ_i is the mechanical power commanded by the governor of generator i and the constant A_i determines the response of the turbine governor to deviations in frequency from the desired system frequency at node i. The deviation from desired system frequency at generator i is denoted by $\dot{\delta}$, whose dynamics are provided by the third order model. A more detailed governor action presented by [65] includes damping action based on the generator's mechanical power:

$$\dot{\rho}_i = -A_i\dot{\delta}_i - \frac{1}{T_{se,i}}(\chi_i^G - P_i^G), \tag{2.27}$$

where χ_i^G is generator i's mechanical power; P_i^G is the expected equilibrium mechanical power of generator i; T_{se} is the servomotor time constant and the constant $\frac{1}{T_{se,i}}$ controls the response of the governor to changes in mechanical power of the generator at node i.

Additionally, we model a deadband frequency range in which the governor does not respond to deviations from system frequency to prevent excess oscillations in generator power:

$$\dot{\rho}_i = \left(-A_i\dot{\delta}_i - \frac{1}{T_{se,i}}(\chi_i^G - P_i^G) \right)\left(1 - 1_{\mathcal{W}}[\dot{\delta}_i] \right), \tag{2.28}$$

where $\mathcal{W}$ is the governor deadband frequency range. An example of such a deadband frequency range can be found in Appendix 8.1, which is applied in Chapter 4.

2.5 Power system emergency responses

This section follows the content of the supplementary document submitted along with the manuscript [37], which is presented in Chapter 4.

In the event of significant deviation in frequency beyond pre-set thresholds following a contingency, protection systems may be activated to disconnect generators, loads, lines or combinations of these with the dual purpose of protecting sensitive equipment and to halt the deviation in system frequency [29, 1]. These emergency responses include:

- Rate of change of frequency (RoCoF) generation shedding;

- Over-frequency generation shedding (OFGS);

- Under-frequency load shedding (UFLS);

- Line disconnection.

2.5.1 Generation shedding

Rate of change of frequency (RoCoF) is the time derivative $\ddot{\delta}$ of the power system frequency. Previously, this quantity was a less significant concern as the inertia of large synchronous generators would counter increases in RoCoF. However, as low-inertia, inverter-based, non-synchronous renewable energy generation increases in prevalence, larger RoCoF values may be observed in networks. Larger RoCoF values may endanger system security due to mechanical limitations of individual synchronous machines [29].

A sufficiently large disturbance, such as a system split or the loss of a large generator in a small network may result in a high rate of change of frequency. To protect against this, transmission system operators (TSOs) employ protection systems for generators against large RoCoF excursions, with different thresholds used among TSOs. For example, [1] specifies RoCoF protection settings of 1.0 Hzs^{-1} over a 0.5 second moving average for synchronous generators greater than 5MW.

Over-frequency generation shedding (OFGS) is triggered when local frequency measured at the generator exceeds a preset threshold [1]. OFGS involves the intentional disconnection of generators in a systematic approach with the objective to contain the system frequency rise during an over-frequency event [2].

OFGS is typically implemented progressively over multiple thresholds: for example, [1] recommends a two-staged scheme for over-frequency protection of generating units.

Modelling generation shedding

To model generation shedding, we monitor whether the frequency $\dot{\delta}(t) > F^+$, where $F^+ \in \mathbb{R}^+$ is a fixed threshold. That is, we consider the over-frequency indicators:

$$\lambda_i^+(t) := \Theta[\dot{\delta}_i(t) - F^+], \tag{2.29}$$

where, $i = 1, \ldots, N$ and Θ is the Heaviside function. Also, given a closed interval $[G^-, G^+]$ with $G^-, G^+ \in \mathbb{R}^+$ and $G^- < G^+$, the RoCoF violation indicator can be written as

$$\psi_i(t) := \Theta[\ddot{\delta}_i(t) - G^+] + \Theta[G^- - \ddot{\delta}_i(t)], \tag{2.30}$$

During the observation time window $[0, T]$, we can count the number of times the emergency schemes are activated in the power grid using the following time-integrated counters (whose definitions are motivated below):

$$\Lambda_i^+(t) := \min\left\{1; \int_0^t \delta[1 + (\lambda_i^+(s_\varepsilon) - \lambda_i^+(s))]ds\right\}, \tag{2.31}$$

and

$$\Psi_i(t) := \min\left\{1; \int_0^t \delta[1 + (\psi_i(s_\varepsilon) - \psi_i(s))]ds\right\}, \tag{2.32}$$

where $s_\varepsilon := s - \varepsilon$ with $\varepsilon > 0$ arbitrarily small and, with a little abuse of notation, here δ indicates the Dirac delta function[1]. The GS emergency response is activated by either an over-frequency or a RoCoF violation, and thus can be written as an indicator function:

$$\Xi_i := (1 - \Theta[\Psi_i])(1 - \Theta[\Lambda_i^+]), \tag{2.33}$$

since once GS is triggered at generator node i we have $\Theta[\Psi_i]$ and/or $\Theta[\Lambda_i^+]$ equal to 1 which gives in turn $\Xi_i = 0$. The GS scheme does not remove the relevant node from the graph, but only disconnects the generator from the network once a violation occurs (which motivates the min terms in (2.31) and (2.32)).

2.5.2 Under-frequency load shedding

Power systems can potentially avoid a frequency collapse following a contingency by implementing automatic load shedding during primary frequency control [78]. This action is known as *under frequency load shedding* (UFLS) and is triggered using under-frequency relays, which detect when frequency falls below preset thresholds and automatically sheds an appropriate amount of controllable load to restore the power balance across the network. These relays are usually installed at distribution and transmission substations [78]. In our case studies, we focus on fixed UFLS schemes, which shed a pre-defined set of load when system frequencies fall below a particular threshold, with further shedding occurring if frequency continues to degrade after the initial activation. This leads to a multi-stage, progressive load shedding scheme employed to combat under-frequency contingencies, see for example [68]. An example of such an UFLS scheme is presented in Table 2.1.

Table 2.1: An example of a multi-stage under-frequency load shed scheme. Adapted from [85].

Frequency (Hz)	Load Shed %
49.2	10%
48.8	20%
48.4	30%
48	40%

---

[1]We assume that the integration domain is consistent with the chosen time discretisation and large enough to allow for the Dirac delta to be well defined, so that (2.31) and (2.32) are natural numbers

Modelling under-frequency load shed events

We model the load shed (LS) scheme as shedding load progressively as frequency decreases. This progression follows a step function, where each jump (LS activation) corresponds to a lower fixed frequency threshold. In formulas, given an ordered set of decreasing frequency thresholds $\underline{F}^- := \{F_1^-, F_2^-, \ldots, F_\Delta^-\}$ with $\Delta \in \mathbb{N}$ and $F_1^- < F^+$, we define a matrix $\hat{\lambda}^-$ of under-frequency indicators element-wise as

$$\lambda_{ij}^-(t) := \Theta[F_j^- - \dot{\delta}_i(t)], \tag{2.34}$$

and the corresponding time-integrated counter matrix Λ^- as

$$\Lambda_{ij}^-(t) := \min\left\{1; \int_0^t \delta[1 + (\lambda_{ij}^-(s_\varepsilon) - \lambda_{ij}^-(s))]ds\right\}, \tag{2.35}$$

where the first index runs over nodes $i = 1, \ldots, N$ and the second over the under frequency set $j = 1, \ldots, \Delta$. The proportion of load shed is then specified as

$$\Gamma_i(\Lambda^-) := \sum_{j=1}^{\Delta} \Lambda_{ij}^- \in \{0, 1, \ldots, \Delta\}. \tag{2.36}$$

This value adjusts the load through χ_i^L by a desired amount at step intervals when a load shedding thresholds are met. Also, recalling (2.25), the damping coefficient which appears in the third order model adjusts dynamically according to:

$$D(\underline{\Gamma}) := D\left[1 - \left(\frac{C}{\sum_{i=1}^{N+L} \chi_i^L}\right) \sum_{i=1}^{N+L} \Gamma_i \chi_i^L\right], \tag{2.37}$$

where buses without the LS scheme have the relevant Γ_i set equal to 0. This results in the difference between the total load before and after a single load shed event as $C\Gamma_i \chi_i^L$ where C is proportion of load being shed at each activation.

2.5.3 Line disconnections

As presented in [27], the line connecting nodes i and j trips when two conditions are concurrently met. The first is verified when the power flowing through the line ij, namely $\phi_{ij} := B_{ij}\nu_i\nu_j \sin(\delta_i - \delta_j)$, exceeds a threshold value P^ϕ for the entire duration of a fixed time interval T^ϕ. So, following the same structure as for the preceding emergency responses, we define the *excess power flow indicator* as

$$\omega_{ij}(t) := \Theta\left[\int_{t-T^\phi}^t \Theta[\phi_{ij}(s) - P^\phi]ds - T^\phi\right]. \tag{2.38}$$

The second controls if each bus frequency deviation exceeds a threshold $F^\phi < F^+$ during $[0, T^\phi]$. This results for the generic bus i in an over-frequency indicator and its corresponding time-integrated counter

$$\lambda_i^\phi := \Theta[\dot{\delta}_i(t) - F^\phi], \quad \Lambda_i^\phi(t) := \min\left\{1; \int_{t-T^\phi}^t \delta[1 + (\lambda_i^\phi(s_\varepsilon) - \lambda_i^\phi(s))]ds\right\}. \quad (2.39)$$

These conditions are combined in a single *line trip indicator* which reads as

$$\Omega_{ij}(t) := 1 - \min\left\{1; \int_{T^\phi}^t \delta[\omega_{ij}(s)\Lambda_i^\phi(s)\Lambda_j^\phi(s) - 1]ds\right\}. \quad (2.40)$$

When a line trip occurs between nodes i and j, the indicator Ω_{ij} switches from 1 to 0.

2.6 Automatic voltage regulation (AVR)

Beyond frequency regulation, voltage stability is a further critical requirement for power system operation. Automatic voltage regulators (AVR) are used to maintain the terminal voltage of a generator at a desired level by controlling the excitation voltage of the generator [30]. The AVR functions when there is a voltage error greater than a specified dead-band threshold between measured terminal voltage and reference voltage due to a disturbance in the network [45]. The simplest such system consists of four main components: the amplifier, exciter, generator and sensor. The basic components and associated transfer function for a simple AVR model is developed below (adapted from [45]).

The sensor detects the terminal voltage for the generator and compares this to the reference voltage set-point stipulated for safe operation, generating an error signal. A sensor can be represented by the first order transfer function:

$$\frac{\nu^s(s)}{\nu(s)} = \frac{K^s}{1 + sT^s},$$

with input ν, the terminal voltage and output ν^s. The parameters K^s and T^s represent the gain and time constant of the sensor model. The amplifier rectifies the input signal from the sensor to a suitable form to control the exciter. A simple amplifier model is given by the first order transfer function:

$$\frac{\epsilon^a(s)}{\nu^e(s)} = \frac{K^a}{1 + sT^a},$$

with inputs $\nu^e = V^r - \nu^s$ and output ϵ^a, where V^r is the reference voltage while ν^s is the voltage signal. K^a and T^a represent the gain and time constants of the amplifier respectively.

The exciter provides DC current to the field winding of the generator, which in turn affects the generated electromagnetic field. The exciter, based on the rectified error signal, adjusts reactive power generation and the generator's field winding current, regulating the generated electro-magnetic field (emf) and field voltage E_i. This in turn regulates terminal voltage. The transfer function for the exciter model takes the form:

$$\frac{\epsilon^e(s)}{\epsilon^a(s)} = \frac{K^e}{1 + sT^e},$$

with input rectified voltage ϵ^a and output change in field voltage ϵ^e, along with gain and time constant K^e and T^e. The components of the transfer function are summarised in the block diagram given in Figure 2.1.
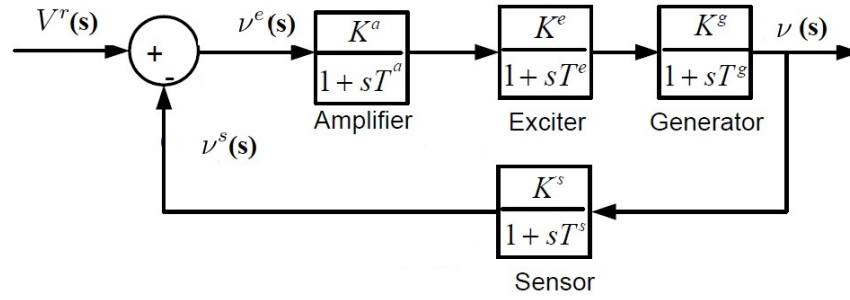
Figure 2.1: Block Diagram for AVR System (adapted from [30])

Common values for the parameters listed for the various components of the AVR system are provided below in Table 2.2.

Table 2.2: Table Showing values of AVR parameters used in the the model; adapted from [30]

Quality	Transfer Function	Gain Constant	Time Constant
Amplifier	$\frac{K^A}{1+T^A}$	$10 \leq K^A \leq 40$	$0.02 \leq T^A \leq 0.1$
Exciter	$\frac{K^E}{1+T^E}$	$1 \leq K^E \leq 10$	$0.4 \leq T^E \leq 1$
Generator	$\frac{K^g}{1+T^g s}$	$0.7 \leq K^g \leq 1$	$1 \leq T^A \leq 2$
Sensor	$\frac{K^s}{1+T^s s}$	$K^s = 1$	$0.001 \leq T^s \leq 0.06$

The following time domain differential equations may be obtained from the transfer functions of the AVR system using the Laplace transform:

$$\begin{cases} \dot{\nu}_i^s(t) = \dfrac{1}{T_i^s}\big(K_i^s \nu_i(t) - \nu_i^s(t)\big) & \text{(2.41a)} \\[2mm] \dot{\epsilon}_i^a(t) = \dfrac{1}{T_i^a}\Big(K_i^a\big(V_i^r - \nu_i^s(t)\big) - \epsilon_i^a(t)\Big). & \text{(2.41b)} \\[2mm] \dot{\epsilon}_i^e(t) = \dfrac{1}{T_i^e}\big(K_i^e \epsilon_i^a(t) - \epsilon_i^e(t)\big) & \text{(2.41c)} \end{cases}$$

2.7 Power system model for rare event analysis

This section summarises the power system model intended for rapid evaluation of frequency dynamics and power system emergency responses following a power disturbance on a specified network. This model is suitable to be used in conjunction with the skipping sampler discussed in Chapter 1 as it takes approximately 2-5 seconds to simulate frequency dynamics for 15 seconds following a simulated power disturbance using MATLAB. Starting with the third order model for the generator, the power system model for rare-event sampling simulations will also include a model for governor action, automatic voltage regulation and a model of protection system emergency responses if the frequency or RoCoF exceeds pre-defined thresholds.

$$\begin{cases} M(\Xi)\ddot{\delta}_i + D(\underline{\Gamma})\dot{\delta}_i = \Xi_i \chi_i^G - \chi_i^L - \nu_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \sin(\delta_i - \delta_j) + \eta_i & \text{(2.42a)} \\[4mm] T_{d,i}\dot{\nu}_i = \Xi_i(E_{f,i} - \epsilon_i^e) - \nu_i + (X_{d,i} - X'_{d,i})\sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \cos(\delta_i - \delta_j) & \text{(2.42b)} \\[4mm] \dot{\rho}_i = \Big(- A_i\dot{\delta}_i - \dfrac{1}{T_{se,i}}(\chi_i^G - P_i^G)\Big)\big(1 - \mathbb{1}_{\mathcal{W}}[\dot{\delta}_i]\big) & \text{(2.42c)} \\[3mm] \dot{\nu}_i^s(t) = \dfrac{1}{T_i^s}\big(K_i^s \nu_i(t) - \nu_i^s(t)\big) & \text{(2.42d)} \\[3mm] \dot{\epsilon}_i^a(t) = \dfrac{1}{T_i^a}\Big(K_i^a\big(V_i^r - \nu_i^s(t)\big) - \epsilon_i^a(t)\Big) & \text{(2.42e)} \\[3mm] \dot{\epsilon}_i^e(t) = \dfrac{1}{T_i^e}\big(K_i^e \epsilon_i^a(t) - \epsilon_i^e(t)\big). & \text{(2.42f)} \end{cases}$$

Equation 2.42a is the Swing Equation, which describes the dynamics of machine i's frequency, where equation (2.42a), $\chi_i^G := \min\{\rho_i + P_i^G, P_i^m\}$ is generator i's mechanical power, P_i^G is the equilibrium mechanical power output of the generator and P_i^m is the generator's nominal maximum power output. The dynamics of each machine's voltage is given by equation (2.42b) (Section 2.3), while the dynamics for governor action which adjust generator's power output in response to frequency deviations are described in (2.42c), where ρ_i is the change in generator power commanded by the governor, A_i is the governor's response to frequency deviation and $\mathcal{W}$ is the governor deadband (Section 2.4.1). Equations (2.42d) - (2.42f) describes the dynamics and action of each generator's voltage sensor, amplifier and exciter respectively, key

components of an automatic voltage regulators (AVR) intended to maintain voltage deviations within prescribed limits (Section 2.6).

The corresponding equations for each load bus $i = N + 1, \ldots, N + L$ are

$$
\begin{cases}
M(\Xi)\ddot{\delta}_i + D(\underline{\Gamma})\dot{\delta}_i = -\chi_i^L + \nu_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \sin(\delta_i - \delta_j) & \text{(2.43a)} \\[2em]
T_{d,i}\dot{\nu}_i = E_{f,i} - \nu_i + (X_{d,i} - X'_{d,i}) \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \cos(\delta_i - \delta_j). & \text{(2.43b)}
\end{cases}
$$

In Chapter 3, a simplified version of this model without the AVR components will be applied to a number of test networks to analyse how power disturbances from low inertial renewable sources can lead to cascading disconnection of loads and generators. In Chapter 4, the full model is applied to the Kundur Two Area Network, an IEEE test network based on the Australian power system. In this application, the model will be used to analyse the effect of power disturbances on a more realistic network, and evaluate how fast response battery storage systems can mitigate the effects of correlated power disturbances.

Chapter 3

Distributions of cascade sizes in power system emergency response

This chapter is derived from a manuscript with the same name published in the 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). All authors have been credited.

Maldon Patrice Goodridge[1] John Moriarty [2]
Andrea Pizzoferrato[3]

Abstract

Following disturbances to a power system triggering emergency responses such as protection or load/generation shedding, several factors affect the way in which these responses may cascade through the network. Beyond deterministic factors such as network topology, in this paper we aim to quantify the effect of correlations in power disturbances. These arise, for example, from common weather patterns causing correlated forecast errors in renewable generation. Our results suggest that for highly connected networks, the cascade size distribution is bimodal and positively correlated disturbances tend to reduce cascade size. The latter relationship is stronger in more reliable networks.

keywords: Cascade size, Correlated disturbances, Emergency response, Protection schemes, Rare events, Network topology

3.1   Introduction

The design and operation of power system emergency response schemes such as protection or load/generation shedding carries increasing importance as countries shift their energy provision from fossil to renewable sources. This is because a move towards more stochastic forms of generation poses new challenges for power system reliability. Renewable generation is subject to exogenous shocks due to forecast errors, implying increased exposure of electrical networks to power fluctuations. These fluctuations in turn can activate power system emergency responses, causing component disconnections in the power grid which may propagate in a cascading fashion [41]. Such cascading failures are the primary cause of most large-scale network outages, and can occur in in both large and small networks.

In the study of more general networks, the probability distribution of cascade sizes has been found to depend in a systematic way on network topology. From a theoretical perspective, in [112] it was shown that in highly connected networks the distribution of cascade sizes is bimodal, while in sparse networks the distribution obeys a power law. This result has been supported by simulation studies [15].

In this paper we view a power system as the combination of the network and a set of correlated random power disturbances. These disturbances are relative to the

[1]School of Mathematical Sciences, Queen Mary University of London, London, UK, m.p.goodridge@qmul.ac.uk

[2]School of Mathematical Sciences, Queen Mary University of London, The Alan Turing Institute, London, UK, j.moriarty@qmul.ac.uk

[3]School of Mathematical Sciences Queen Mary University of London, The Alan Turing Institute, London, UK, ap@andreapizzoferrato.com

planned operating equilibrium and arise, for example, from errors in generation or demand forecasts. It is potentially important to consider correlation in these disturbances since renewable generation can experience common weather patterns over significant spatial areas. From an electrical perspective, power system transients are coupled via the network topology. Our aim in this work is to explore the implications of this correlation and coupling for the cascading of power system emergency responses.

In particular, in contrast with quasi-static or quasi-dynamic studies of cascade sizes [114, 116, 51], we simulate power system dynamics using an appropriate system of differential equations. To represent opposite extremes of connectivity we consider both a fully-connected and a ring topology. The authors of [99] used the *swing equation* to analyse the propagation of cascade failures in network. However, this work focused only on the cascades of line disconnections. We extend this framework by considering the activation of generator and load disconnection schemes following a change in equilibrium nodal power injections, or power disturbance. To model high penetration of distributed renewable generation, power disturbances are applied at the majority of nodes (that is, buses). Transient dynamics are modelled by the *third-order model* (see for example [100, 64]), i.e. differential equations for an AC system taking into account both power and voltage, as advocated by [99] as necessary to evaluate system dynamics on longer time periods.

The power disturbances, which perturb these dynamics, are sampled from correlated Gaussian distributions with zero mean to represent forecast errors. While empirical studies have noted that wind generation has been measured to induce frequency deviations with non-Gaussian, heavier tailed distributions on timescales of approximately 1 second, frequency deviations were found to be approximately Gaussian on timescales greater than 1s. As we model a static disturbance over the simulation period, we consider a Gaussian model suitable for this investigation. Future work will endeavour to study the effect power disturbances with heavier-tailed distributions. Emergency response schemes, which monitor the local frequency at each node, are activated either by this frequency moving outside a certain range or by the RoCoF (*Rate of Change of Frequency*) exceeding a given magnitude.

Our results suggest that in more highly connected networks, positively correlated disturbances tend to reduce cascade size. Further, this relationship is stronger in more reliable networks. We confirm findings from more general network models that cascade size distribution in highly connected networks is bimodal. The results also suggest that in highly resilient systems, not only do emergency response cascades occur less frequently, but their size distribution is also different when compared to less resilient systems.

The rest of the paper is structured as follows. In Section 3.2 we provide the transient model, while Section 3.3 details the emergency response schemes considered.

Section 3.4 details the statistical model, while Section 3.5 presents a Monte Carlo study and discussion and Section 3.6 outlines potential future directions.

3.2 Transient Model

We consider an unweighted and undirected graph $G(N + 1, V)$, having $N + 1$ nodes (buses) and V edges (lines). Each node $1, \ldots, N$ represents a bus with renewable generation, conventional generation and load, and is modelled as an equivalent generator, while node $N + 1$ represents a pure load. The equivalent generators at nodes $1, \ldots, N$ may be electrically disconnected from the network by the activation of the emergency response schemes, while the load at node $N + 1$ may be progressively shed.

The equivalent generator dynamics are given by the third-order model, regulated by a governor. As in [64], each governor is itself modelled as an equivalent generator. The corresponding set of differential equations for node $i = 1, \ldots, N$ is (where constant and time-evolving variables are indicated using Latin upper-case and Greek lower-case letters, respectively):

$$
\begin{cases}
M_i \ddot{\delta}_i + D_i \dot{\delta}_i = \alpha_i \left(\chi_i - \nu_i \sum_{j=1}^{N+1} B_{ij} \nu_j \sin(\delta_i - \delta_j) \right), \\
T_{d,i} \dot{\nu}_i = \alpha_i E_i - \nu_i + L_i \sum_{j=1}^{N+1} B_{ij} \nu_j \cos(\delta_i - \delta_j), \\
\dot{\rho}_i = -A_i \dot{\delta}_i - H_i \rho_i,
\end{cases}
\tag{3.1}
$$

where at node i, $\chi_i = \min\{\rho_i + P_i^0 + \eta_i, P_i^m\}$ is the power injection, η_i accounts for a power disturbance, and α_i is an indicator function encoding the emergency scheme operation, as specified below. The matrix B encodes the topology of the graph, while the remaining quantities are specified in Table 3.1 (see [100] for further details). Node $N + 1$ is modelled as a synchronous motor load via the differential equations

$$
\begin{cases}
M_{N+1} \ddot{\delta}_{N+1} + D_{N+1} \dot{\delta}_{N+1} = \quad -(P_{N+1}^0 - P^s \Gamma(\Lambda^-)) \\
\quad + \nu_{N+1} \sum_{j=1}^{N+1} B_{N+1j} \nu_j \sin(\delta_{N+1} - \delta_j), \\
T_{d,N+1} \dot{\nu}_{N+1} = \\
\quad E_{N+1} - \nu_{N+1} + L_{N+1} \sum_{j=1}^{N+1} B_{N+1j} \nu_j \cos(\delta_{N+1} - \delta_j),
\end{cases}
\tag{3.2}
$$

where P_{N+1}^0 is the nominal value of the pure load and $P^s \Gamma(\Lambda^-)$ is the amount of load shed, as detailed in Section 3.3. (For convenience we take ρ_{N+1} identically

Table 3.1: Variables labels used in (3.1).

Label	Meaning	Unit of measure
$\delta_i \in \mathbb{R}^+$	Phase angle	[rad]
$\dot{\delta}_i \in \mathbb{R}$	Frequency	[s^{-1}]
$\ddot{\delta}_i \in \mathbb{R}$	Rate of Change of Frequency (RoCoF)	[s^{-2}]
$\nu_i \in \mathbb{R}$	Voltage	[WA^{-1}]
$\rho_i \in \mathbb{R}^+$	Mechanical Power	[W]
$A_i \in \mathbb{R}^+$	Power-frequency response	[W]
$B \in \mathbb{R}$	Susceptance matrix	[W^{-1}A^2]
$D_i \in \mathbb{R}^+$	Damping Torque	[Ws]
$E_i \in \mathbb{R}^+$	Rotor field	[WA^{-1}]
$H_i \in \mathbb{R}^+$	Damping Power	[s^{-1}]
$L_i \in \mathbb{R}^+$	Reactance	[WA^{-2}]
$M_i \in \mathbb{R}^+$	Angular Momentum	[Ws^2]
$P_i^m \in \mathbb{R}$	Maximum power output	[W]
$T_{d,i} \in \mathbb{R}^+$	Transient Time d-axis	[s]

equal to 0.) The above mentioned studies [112, 15] found a relationship between graph connectivity and the distribution of cascade sizes in general graph models. We explore the extent to which this translates to power systems contexts by choosing two topologies which are extremes of connectivity, namely the fully connected graph and ring graph. (Note that the Monte Carlo simulation study of Section 3.5 was also run with five symmetrically spaced nodes capable of load shedding, and this was found to make no qualitative difference to the results. As a result, in this study all load shedding is located at node $N + 1$). We note that the choice of network topologies was not intended to reflect realistic power system networks, but to allow inferences to be drawn about the relationship between connectivity and cascade sizes induced by random power disturbances. We examine a more realistic network topology in Chapter 4 where we apply the power system model to the Kundur Two Area network.

In practice we sample the observable evolution (3.1)–(3.2) in discrete time, acknowledging that this is subject to the particular choice of time discretisation, which we assume to be optimal (no loss of sensitivity, see [74] for a detailed discussion).

3.3 Disturbances and emergency responses

In contrast to the swing equation, the third-order model captures voltage dynamics and angle-voltage interplay and has the advantage that disturbances are not required to be small. We model disturbances as constant over the timescale of our study,

which is on the order of seconds. In particular we have

$$\underline{\eta}(t) := \underline{U}\Theta[t] \in \mathbb{R}^N, \quad [\text{W}] \tag{3.3}$$

where $\underline{U} = (U_i)_{i=1,\ldots,N}$ is the vector of disturbance magnitudes and Θ is the Heaviside step function.

An equilibrium $(\delta_i(0), \nu_i(0))_{i=1,\ldots,N+1}$ is first calculated for equations (3.1)–(3.2) and taken as the initial conditions at time $0-$ (that is, just before time 0). From time 0 onwards, the constant power disturbance (3.3) perturbs this equilibrium according to (3.1)–(3.2), creating transient dynamics. These dynamics are simulated over an observation time window $[0, T]$ with $T \in \mathbb{R}^+$ [s].

We model emergency response from both the generators (generation shedding, GS) and load (load shedding, LS). The equivalent generator at each node is disconnected from the network when either RoCoF is observed to lie outside its normal operating band, or when frequency is observed to lie above a given limit. According to this scheme, the power injected at node i is specified as follows. To detect when the frequency $\dot{\delta}(t) > F^+$, where $F^+ \in \mathbb{R}^+$ [s^{-1}], we introduce over-frequency indicators:

$$\lambda_i^+(t) := \Theta[\dot{\delta}_i(t) - F^+], \tag{3.4}$$

where, $i = 1, \ldots, N$. Similarly, the RoCoF indicators are

$$\psi_i(t) := \Theta[\ddot{\delta}_i(t) - G^+] + \Theta[G^- - \ddot{\delta}_i(t)], \tag{3.5}$$

where $[G^-, G^+]$ is a closed interval with $G^-, G^+ \in \mathbb{R}^+$ [s^{-2}] and $G^- < G^+$. The following time-integrated counters record the number of times each protection scheme is activated during the observation window:

$$\Lambda_i^+(t) := \min\left\{1, \int_0^t \delta[1 + (\lambda_i^+(s-) - \lambda_i^+(s))]ds\right\}, \tag{3.6}$$

$$\Psi_i(t) := \min\left\{1, \int_0^t \delta[1 + (\psi_i(s-) - \psi_i(s))]ds\right\}, \tag{3.7}$$

where δ denotes the Dirac delta function. The indicator function α_i is then given by

$$\alpha_i = (1 - \Theta[\Psi_i])(1 - \Theta[\Lambda_i^+]). \tag{3.8}$$

Thus the activation of GS at node i disconnects its equivalent generator electrically from the network without altering the network topology.

In contrast to the GS scheme, load is shed progressively as frequency falls below a sequence of thresholds. Given a decreasing set of frequencies $\underline{F}^- := \left\{F_1^-, F_2^-, \ldots, F_\Delta^-\right\}$

with $\Delta \in \mathbb{N}$ and $F_1^- < F^+$, we define a matrix λ of under-frequency indicators by

$$\lambda_{ij}^-(t) := \Theta[F_j^- - \dot{\delta}_i(t)], \tag{3.9}$$

and the corresponding time-integrated counter matrix Λ^- as

$$\Lambda_{ij}^-(t) := \min\left\{1, \int_0^t \delta[1 + (\lambda_{ij}^-(s-) - \lambda_{ij}^-(s))]ds\right\}, \tag{3.10}$$

where the first index runs over nodes $i = 1, \ldots, N$ and the second over the under frequency set $j = 1, \ldots, \Delta$. The amount of load shed is then specified as

$$\Gamma(\Lambda^-) := P^s \max_{i=1,2,\ldots,N} \sum_{j=1}^{\Delta} \Lambda_{ij}^-, \tag{3.11}$$

where $P^s \in \mathbb{R}^+$ $[MW]$. Thus load is shed in equal steps (activations), driven by the lowest frequency observed in the network. At any time t, the number of activations which have occurred by time t is equal to the index j of the lowest threshold F_j in $\underline{F}^- := \left\{F_1^-, F_2^-, \ldots, F_\Delta^-\right\}$ which has been observed.

3.4   Statistical model

Beyond the initial power disturbance vector $\underline{U}$ at time 0, we also model the dynamic effect on transient behaviour of the above emergency responses. In particular we are interested in the distribution of the total number of responses during the observation time window, which we call 'cascade size'. In order to explore the effect of correlation in the initial disturbances on cascade size we use the following convenient statistical model.

The interpretation of the model is as follows. As discussed above, each node $1, \ldots, N$ is an equivalent generator representing a bus with renewable generation, conventional generation and load. Further the system is assumed to be operating in equilibrium at time $0-$, and at time 0 random power disturbances are applied at each generator node. These disturbances can therefore be understood to take account of errors in the generation schedule, due for example to imperfect forecasts of renewable production. The initial power disturbances are therefore taken to have mean zero. A covariance matrix Σ parametrised by two positive parameters σ, α is taken:

$$\Sigma_{ij} = \begin{cases} \sigma^2, & i = j, \\ \alpha\sigma^2, & i \neq j, \end{cases} \quad i, j \in \{1, \ldots, N\}. \tag{3.12}$$

In addition to its convenience, this particular choice of the covariance matrix guarantees its positive semi-definiteness provided that $\alpha \in [0, 1]$. The random power disturbance U_i at generator node i is also scaled linearly according to the equilib-

rium power injection P_0^i at that node. That is, the initial power disturbance $\underline{U}$ in (4.11) is taken to be the Hadamard product

$$\underline{U} = \underline{u} \circ \underline{P}^0. \tag{3.13}$$

3.5   Monte Carlo study

The Monte Carlo simulation study in this section aims to explore the influence of correlation in the initial power disturbances $\underline{U}$ on the cascade size distribution. We also aim to distinguish its effect from that of two other factors: the rareness of emergency scheme activations, and the connectivity of the network.

Correlations between the initial power disturbances $\underline{U} = (U_1, \ldots, U_N)$ arise, for example, from common weather patterns causing correlated forecast errors. This is accounted for in our framework by varying the correlation parameter α of the covariance matrix Σ.

Rareness has recently been addressed by applying the *large deviations* framework to power grids [79]. Large deviations theory concerns events with exponentially small probabilities, and provides insight into how such events typically occur. We account for the rareness of emergency responses by varying the standard deviation parameter σ of the covariance matrix Σ, as smaller values of σ tend to make emergency scheme activations less frequent.

Connectivity is also an important driver for cascade sizes in more general models of network dynamics [112, 15]. We account for its influence by simulating cascades on two different network topologies. For clarity we choose two topologies at the extremes of connectivity, namely the complete graph and ring graph.

For each considered network and pair of values α, σ, the disturbance vector $\underline{U}$ is independently sampled $N_{\mathrm{mc}} = 10^5$ times. For each sample $\underline{U}_j$ the dynamics (4.3)–(4.4), which take account of the emergency responses described in Section 3.3, are run over the observation time window and the number of emergency responses as encoded by (3.6), (3.7), and (3.10) is recorded. For instance in the case of over-frequency response, this results in a matrix $D_{\mathrm{OF}} \in \mathbb{N}^{N_{\mathrm{mc}} \times (N+1)}$ whose (i, k)th entry for $k = 1, \ldots, N$ is equal to the number of times generator k is shed under the initial disturbance $\underline{U}_i$ (and whose $(i, N+1)^{th}$ entries are zero since there is no generator at node $N+1$). Note that the entries of D_{OF} can only take the value 0 or 1 since, under the scheme of Section 3.3, generator disconnection is permanent during the time observation window. Similarly, matrices D_{UF} and D_{R} can be respectively defined for UF responses and RoCoF responses. The data matrix for all sampled responses is thus

$$D := D_{\mathrm{UF}} + D_{\mathrm{OF}} + D_{\mathrm{R}} \in \mathbb{N}^{N_{\mathrm{mc}} \times (N+1)}, \tag{3.14}$$

while the vector of cascade sizes is

$$\underline{d} := \{d_1, d_2, \ldots, d_{N_{\mathrm{mc}}}\} \text{ with } d_i := \sum_{k=1}^{N+1} D_{ik}. \tag{3.15}$$

Table 3.2 records the parameter values used.

Table 3.2: Parameter values used. Semicolons separate differing values for generators (left) and the load (right).

Variable	Value
A_i	1
B	$\mathrm{diag}(-237) + \mathrm{offdiag}(15.8)$
D_i	1; 2
Δ	4
E_i	1
$\underline{F}^-$	$-0.1, -0.2, -0.4, -0.8$
F^+	0.1
$G^+ = -G^-$	2
H_i	4
L_i	0.17
M_i	$29.0 \cdot 10^{-4};\ 2.95 \cdot 10^{-1}$
P_i^0	0.5; -750
P_i^m	100 ;
P^s	50
$T_{d,i}$	1
N_{mc}	10^5
N	15
σ^2	$\{0.01, 0.03, 0.05\}$

Figure 3.1 presents the effect of the correlation parameter α on the empirical distribution of cascade sizes, for both the fully connected and ring networks, when the disturbance standard deviation is $\sigma^2 = 0.03$ and $\sigma^2 = 0.01$ respectively. These standard deviations were chosen so that the proportion of samples with at least one emergency response was on the order of 10% in both cases. For better legibility, all plots exclude samples with 0 emergency responses. Figure 3.2 indicates the split between activations of the three different protection schemes. In both cases the results shown are for a single value of α, but these splits are representative of the other values for α (data not shown). Figure 3.3 plots the mean cascade size as α varies for several different values of the disturbance standard deviation σ.

Figure 3.1 confirms the effect of network topology. Firstly, the fully connected network required disturbances with significantly higher standard deviation ($\sigma^2 = 0.03$) than those of the ring network ($\sigma^2 = 0.01$) in order to generate the same proportion of samples with emergency responses. In this sense, the fully connected network exhibits greater resilience. Secondly, consistent with the more general models of cascades on graphs explored in [15, 112], the fully connected network gives rise to a bimodal distribution in which the cascade size is almost always small or
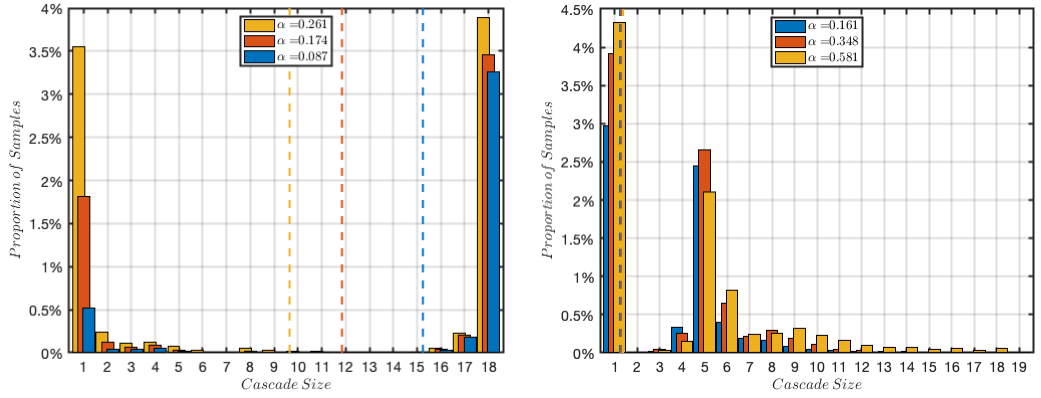
Figure 3.1: Histogram of empirical cascade sizes for the fully-connected network with $\sigma^2 = 0.03$ (left) and ring network with $\sigma^2 = 0.01$ (right), and mean cascade size (dashed vertical lines), as the disturbance correlation α varies. For better legibility, cascades of size 0 are excluded; the remaining samples constitute on the order of 10% of the samples in each case. Bi-modality, and the decreasing relationship between correlation and mean cascade size, are clearly seen in the top panel.

large. This bi-modality is clarified by Figure 3.2, which plots the split between the under-frequency load shedding, over-frequency generation shedding and RoCoF emergency schemes. RoCoF responses are shown to play a significant role only in the fully connected network, and they typically occur at all 15 generation nodes. This is consistent with the high connectivity of the latter network, since a rapid transient at any node influences every node. (Note that the two modes at 3 and 15 activations in the top panel of Figure 3.2 correspond to the single mode at 18 activations in the top panel of Figure 3.1.)

For the fully connected network, from Figure 3.1 the mode at 1 activation is driven by increasing values in the correlation parameter α. Further, Figure 3.2 shows that this mode consists of either under- or over- frequency emergency responses. As the correlation parameter α increases, the initial power disturbances $\underline{U} = \{U_1, \ldots, U_N\}$ are then more homogeneous. In the fully connected network, a single generation or load shedding respectively is therefore more capable of correcting the initial disturbance $\underline{U}$. Note also from Figure 3.1 that for the ring network, the mode at 1 activation is again driven by increasing values in the correlation parameter α, although the cascade size distribution varies much less as a function of the disturbance correlation. This observation is consistent with the ring network's low connectivity, meaning that disturbances have only local influence, making their global correlation less important.

As put on a rigorous basis by large deviations theory [79], the rareness of an event influences the way in which it typically occurs. It follows that as the disturbance standard deviation parameter σ decreases and emergency responses become more rare, so the statistical pattern of the disturbances $\underline{U}$ causing them may also change. This is borne out in the top panel of Figure 3.3, where the effect of the correlation parameter α is in general observed more clearly for smaller values of σ. The bottom

panel of Figure 3.3 confirms that for the ring network there is little relationship between α and the mean cascade size.
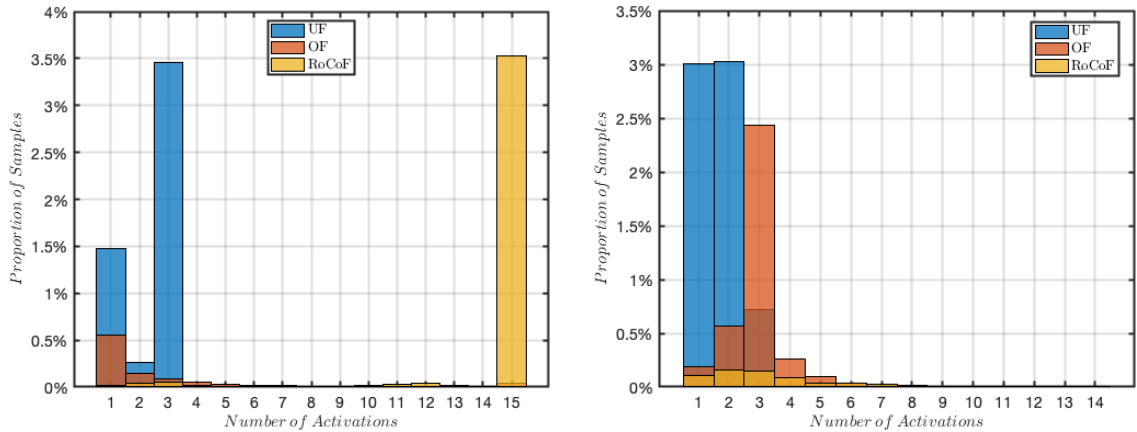

Figure 3.2: Split between activations of under frequency load shedding (UF), over frequency generation shedding (OF) and RoCoF emergency schemes. Left: fully connected network, $\alpha = 0.174$, $\sigma^2 = 0.03$. Right: ring network, $\alpha = 0.16149$, $\sigma^2 = 0.01$.
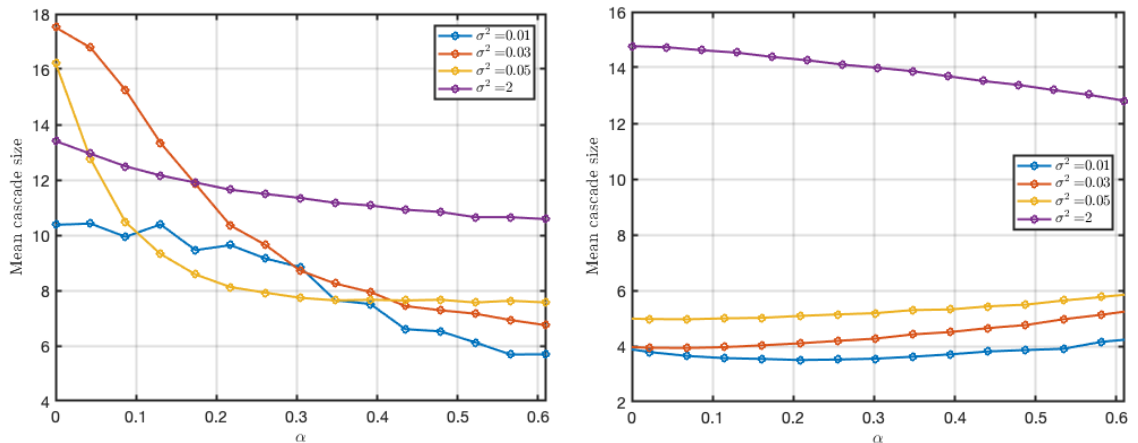
Figure 3.3: Plot of mean cascade size versus disturbance correlation parameter α for the fully-connected network (Left) and ring network (Right), for different values of the disturbance standard deviation σ. The corresponding curves for intermediate values of σ are intermediate between those plotted.

3.6 Conclusion and Outlook

In this paper we have simulated cascades in power system protection and load/generation shedding when the network is perturbed by multiple exogenous shocks. Considering these shocks as errors in the generation schedule, due for example to imperfect forecasts of renewable production, we have modelled them using Gaussian random vector $\underline{U}$ with zero mean and a covariance matrix Σ parametrised by two parameters σ, α. We have distinguished the effect of correlation in the initial power disturbances

$\underline{U}$ from that of two other factors which are known to have material effects, namely the rareness of emergency scheme activations and the connectivity of the network.

The results indicate that these correlations have a material effect if the network is highly connected. In this case, increases in the correlation parameter α drive an increase in the number of small cascades and, in particular, single emergency responses. This may be explained by the fact that in highly connected networks, a single generation or load shedding is capable of correcting multiple similar disturbances $\underline{U} = (U_1, \dots, U_N)$. Interestingly this effect is observed clearly only when emergency responses are rare, which is the operational paradigm for many power systems globally. This observation is in line with large deviations theory [79]: under our Gaussian model, theory predicts that a large overall disturbance typically arises as an aggregate of small but consistent disturbances. The results also confirm findings from more general models of cascades on graphs explored in [15, 112] concerning the effect of network topology on cascade size.

In order to draw qualitative conclusions on the effects of correlation, rareness and topology, the networks we have considered are highly stylized and there is clearly scope to explore more realistic systems. Our modelling framework may also prove useful for the study of innovative emergency schemes, such as the use of grid-connected battery storage systems. Further kinds of disturbance may also be considered, for example line faults which result in dynamic changes to network topology.

Considering the bi-modality observed in Figure 3.1 (top), care may be needed in the design of simulation and sampling schemes for cascade size distributions. In particular, when large cascades are rare, bias could arise related to under-sampling of large cascades. More sophisticated sampling schemes than the Monte Carlo approach of the present paper may therefore be required, in particular methods designed for rare events (see for example [12, 74]). We employ such a specialised rare event sampler- the *skipping sampler*, in Chapter 4.

Acknowledgments

Chapter 4

A rare-event study of frequency regulation and contingency services from grid-scale batteries

This chapter presents a manuscript published on 7 June 2021 with the same name in the 2021 issue of the Philosophical Transactions of the Royal Society A. All authors have been credited. Some variable symbols have been change from the published version to maintain consistency with other chapters in the thesis.

Maldon Patrice Goodridge[1], John Moriarty[2] Andrea Pizzoferrato[3]

Abstract

We perform a rare-event study on a simulated power system in which grid-scale batteries provide both regulation and emergency frequency control ancillary services. Using a model of random power disturbances at each bus, we employ the skipping sampler, a Markov Chain Monte Carlo algorithm for rare-event sampling, to build conditional distributions of the power disturbances leading to two kinds of instability: frequency excursions outside the normal operating band, and load shedding. Potential saturation in the benefits, and competition between the two services, are explored as the battery maximum power output increases.

4.1 Introduction

A growing number of countries are transitioning their energy generation from fossil to renewable energy sources (RES), due to an increased desire to mitigate the effects of climate change [36]. While on one side this shift positively impacts the environment [35] (for instance, by reducing carbon emissions), it also poses new challenges in power system stability and security. Renewable resources are vulnerable to exogenous shocks, such as from weather conditions [87], resulting in increased exposure of the electricity network to power fluctuations. When combined with endogenous power disturbances due to various other system contingencies, these disturbances could in the worst case activate emergency responses in the power grid [48, 27] which may propagate in a cascading fashion [41, 77, 5, 92, 63, 113].

There is increasing interest in embedding battery energy storage systems (BESS) in power grids [83] to improve grid stability and resilience to disturbances [115, 73, 8, 91, 26, 70] by virtue of their rapid response [52]. In this study we employ the skipping sampler, a Markov Chain Monte Carlo (MCMC) rare-event sampler, to examine the benefits of BESS as both the battery maximum power output (MPO) and the distribution of power disturbances vary. Taking a probabilistic model for the potential power disturbances at all buses (interpreted as both exogenous and endogenous), we build conditional distributions of the combinations of disturbances which lead to rare instabilities of two kinds. Firstly, to examine the frequency regulation benefits of BESS we study frequency excursions beyond the normal operating band. Then, to explore the emergency response provided by BESS, we condition on a load shedding event. Potential saturation in the benefits of BESS is examined

---

[1]Queen Mary University of London
[2]Queen Mary University of London, The Alan Turing Institute, London
[3]University of Bath

as MPO increases, and also potential interactions between the frequency regulation and contingency services which they provide.

In this line of research we aim to combine time-domain simulation with random sampling, in order to study rare events occurring in power systems. The main challenges are that standard Monte Carlo simulation methods are unreliable in rare event settings [88], a problem compounded by the heavy computational expense of time-domain, simulation-based assessment of power system robustness [62]. In the face of these challenges, one approach is to apply carefully reformulated methods for random sampling (see for example [84]), while another is to explore alternatives to random sampling, such as those studied in [79, 18, 62]. However, these methods typically involve simplifying assumptions of varying strength, which may make these studies challenging to transfer to other related contexts.

This paper continues recent work based on Markov Chain Monte Carlo (MCMC) random sampling which was begun in [74, 38]. While the latter papers also made strong simplifications, our goal in the present work is to demonstrate the incorporation of a power system model which is both detailed and adaptable. That is, the model and case study below incorporate network topology and nodal frequencies, third-order dynamics, automatic voltage regulation, primary and secondary frequency control, battery storage, frequency-based load and generation shedding, and rate of change of frequency (RoCoF) generator protection. We also emphasise that, as in the latter papers, any probability density may be taken as the model of power disturbances.

In [74] a simpler MCMC rare-event sampler was used to build conditional distributions of random disturbances leading to excessive RoCoF at any bus. One limitation in that work was the use of dynamics which neglect system voltage transients and are valid only for small disturbances. In order to study cascades of emergency responses caused by larger disturbances, [38] employed third-order dynamics. The latter work established the effect of network connectivity on the conditional distribution of cascade sizes (that is, the number of emergency responses arising from the initial disturbances), namely that these distributions can be highly bimodal in more highly connected networks. However, these findings were obtained through the use of toy power system models at the extremes of connectivity, namely the ring and fully connected networks. In contrast, we perform a case study based on Kundur's two-area system [57], including the additional features of BESS, line disconnection and automatic voltage regulation. This enables study of the benefits of BESS for system robustness and resilience under uncertainty, taking into account potential interactions with emergency protection schemes.

In the spirit of the present theme issue of *Philosophical Transactions A*, we aim as far as possible for a self-contained presentation suitable for an interdisciplinary audience, with further technical detail provided in Chapter 2. The results may be

reproduced using the code available at https://github.com/ahw493/ .

4.2   Simulation model

Our simulation model is cyber-physical, in the sense that we model physical observables (the physical layer) in continuous time while simultaneously modelling emergency protection schemes (the cyber layer) in discrete time. Between interventions from the cyber layer, the physical layer evolves via the differential equations (4.3) and (4.4). The cyber layer, which consists of the emergency protection schemes described in Chapter 2 Section 2.5, inspects the physical layer at regular intervals. On this discrete-time lattice, if any activation criterion is met then the corresponding protection system activates. Each activation creates a discontinuity in the physical layer, resulting either from an instantaneous reduction in generation or load, or an instantaneous change to the graph topology through the loss of an edge (line).

Table 3.1 collects notation used in the paper. Constant and time-evolving variables are indicated using Latin and Greek letters, respectively. Lower-case Greek letters correspond to physical observables (rotor angles, frequencies, powers and voltages), while upper-case Greek letters are used to denote the state of the emergency response schemes. Vectors are underlined e.g $\underline{v}$.

4.2.1   Network model

Consider an undirected graph $G\left(N+L, W\right)$, where $N+L$ is the set of nodes and W a set of weighted edges. The set of nodes $1, \ldots, N$ represents generator buses, while the remaining $N+1, \ldots, N+L$ nodes represent load buses. Edges represent power lines, and the weight W_{ij} of line ij (that is, the line between buses i and j) is the triple of characteristics $W_{ij} = (p_{ij}, e_{ij}, l_{ij})$, where p_{ij} is the reactance per unit, e_{ij} the resistance per unit and l_{ij} the length of line ij. The line indicator variable Ω_{ij} switches from 1 to 0 upon disconnection of line ij by the protection scheme. The system's susceptance matrix can then be written as a Laplacian-like matrix (see [20]), expressed in turn as a function $B(\Omega)$ of the line indicator variables:

$$B_{ij}(\Omega) := \begin{cases} \left(\sum_{k=1}^{N+L} w_{ik}(\Omega_{ik}) \right) - w_{ij}(\Omega_{ij}) & \text{if } i = j, \\ -w_{ij}(\Omega_{ij}) & \text{if } i \neq j, \end{cases} \tag{4.1}$$

where

$$w_{ij}(\Omega_{ij}) := \begin{cases} 0 & \text{if } i = j \\ -\dfrac{p_{ij}}{e_{ij}^2 + p_{ij}^2} \dfrac{\Omega_{ij}}{l_{ij}} & \text{if } i \neq j. \end{cases} \tag{4.2}$$

The network is assumed to be lossless, so that the susceptance matrix coincides with the imaginary part of the admittance matrix [57, 101].

Table 4.1: Summary of notation. Note: per unit quantities are denoted p.u.

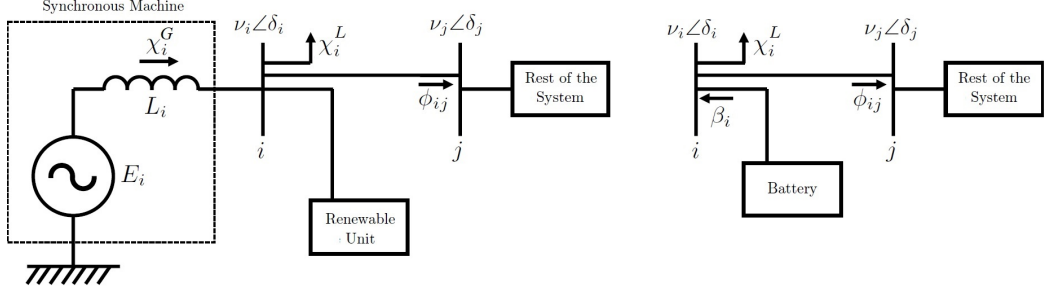| Sym.	Meaning	Units
Equivalent generator and line parameters		
A_i	Governor droop response	MW/rad
B_{ij}	Susceptance matrix	p.u
D	Load damping factor	%
δ_i	Voltage angle	rad
$\dot{\delta}_i$	Frequency	p.u
$\ddot{\delta}_i$	Rate of change of frequency (RoCoF)	p.u
E_i	Rotor field voltage	p.u
ϵ_i^e	Automatic voltage regulation	p.u
η_i	Random power injection or drain	p.u
L_i	Equivalent machine reactance (see Appendix 8.1)	ohms
$M(\underline{\Xi})$	System angular momentum	Ws^2
M_i	Generator angular momentum	Ws^2
ν_i, ν_j	Nodal voltage	p.u
P_i^m	Maximum power output	p.u
P_i^G	Initial generator power (at t= 0)	p.u
P_i^L	Initial loads (at t= 0)	p.u
ϕ_{ij}	Power flow from bus i to j	p.u
ρ_i	Governor mechanical power	p.u
$T_{d,i}$	Transient time constant d-axis	s
$\mathcal{W}$	Governor deadband frequency range	Hz
χ_i^L, χ_i^G	Active bus loads/generation	p.u
Battery parameters		
β_i	Battery power injection/drain	p.u
b_i^+	Initial battery state parameter	%
B_i^0	Initial battery power	MW
B^m	Maximum battery power	MW
B^r	Maximum battery power for regulation FCAS	MW
B^g	Battery response to global frequency	MW
B^l	Battery response to local frequency	MW
F^d	Battery deadband frequency deviation	Hz
F^n	Emergency FCAS frequency deviation	Hz
F^m	Frequency deviation associated with maximum battery power	Hz
$\mathcal{E}$	Frequency interval for local battery response	Hz
$\mathcal{R}$	Frequency interval for AGC commanded battery response	Hz
$\mathcal{D}$	Battery's deadband frequency interval	Hz
T^b	AGC signal interval	s
Protection system parameters		
C	Load shedding increments	%
Ω_{ij}	Indicator for line protection system	-
Ξ_i	Indicator for generation shedding emergency response	-
Γ_i	Indicator for load shedding emergency response	-

Figure 4.1: (Left) Generator bus including synchronous machine, local load, and renewable unit (Right) Load bus including load (modelled as a motor) and BESS. ϕ_{ij} is the power flow through the line from bus i to bus j, see Appendix 8.1 for parameter choices, Table 3.1 and main text for other notation.

4.2.2 System dynamics

Each generation bus $i = 1, \ldots, N$ includes a synchronous machine, load, and renewable unit, as shown in Figure 4.1 (left) and equations (4.3):

$$
\begin{cases}
M(\Xi)\ddot{\delta}_i + D(\underline{\Gamma})\dot{\delta}_i = \Xi_i \chi_i^G - \chi_i^L - \nu_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \sin(\delta_i - \delta_j) + \eta_i & \text{(4.3a)} \\[3mm]
T_{d,i}\dot{\nu}_i = \Xi_i(E_i - \epsilon_i^e) - \nu_i + L_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \cos(\delta_i - \delta_j) & \text{(4.3b)} \\[3mm]
\dot{\rho}_i = -A_i \dot{\delta}_i(1 - 1_{\mathcal{W}}[\dot{\delta}_i]) & \text{(4.3c)}
\end{cases}
$$

In equation (4.3a), $\chi_i^G := \min\{\rho_i + P_i^G, P_i^m\}$ is generator $i's$ mechanical power, P_i^G is the equilibrium mechanical power output of the generator, P_i^m is the generator's nominal maximum power output and ρ_i is the power contribution of a governor, whose dynamics are described in (4.3c) where $\mathcal{W}$ is the governor deadband (see Appendix 8.1); voltage dynamics are given by (4.3b); and the terms $M(\Xi)$, $D(\underline{\Gamma})$ and ϵ_i^e are detailed below.

Each load bus $i = N+1, \ldots, N+L$ includes a load and a battery, as shown in Figure 4.1 (right), whose corresponding equations are

$$
\begin{cases}
M(\Xi)\ddot{\delta}_i + D(\underline{\Gamma})\dot{\delta}_i = -\chi_i^L + \nu_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \sin(\delta_i - \delta_j) + \beta_i & \text{(4.4a)} \\[3mm]
T_{d,i}\dot{\nu}_i = E_i - \nu_i + L_i \sum_{j=1}^{N+L} B_{ij}(\Omega_{ij})\nu_j \cos(\delta_i - \delta_j). & \text{(4.4b)}
\end{cases}
$$

The initial conditions of the above system of equations, denoted $\delta_i(0)$, $\nu_i(0)$, $\rho_i(0)$ and P_i^G are set equal to equilibrium states which can be determined numerically. [4]

---

[4] see, for example, [101] for a study of the existence of these states and Appendix 8.1 for specific values.

Recalling that the models of generation and load shedding are presented in Chapter 2, the variables $\underline{\Xi} := \{\Xi_1, \Xi_2, \ldots, \Xi_N\}$ are used to record activations of generation shedding emergency responses. That is, Ξ_i is initially 1 and switches to 0 when generation shedding occurs at generator node i. The global variable $M(\underline{\Xi}) := \sum_{i=1}^{N+L} \Xi_i M_i$ then represents the total angular momentum of the system, while accounting dynamically for generation shedding through its dependence on $\underline{\Xi}$. For the load buses $i = N+1, \ldots, N+L$ the values $\chi_i^L := (1 - C\Gamma_i) P_i^L$ are also dynamic, due to the action of the load shedding emergency scheme. Here $0 < C < 1$ is the additional amount of load shed per activation of this emergency scheme and Γ_i counts the number of load shedding events at node i. In the same way, the global damping coefficient $D(\underline{\Gamma})$ accounts dynamically for load shedding events. The variable η_i representing the power disturbance at bus i is detailed in Section 4.3. The term ϵ_i^e accounts for the action of automatic voltage regulation (see Section 2.6), while β_i is the battery power injection/drain defined in (4.5). Note that the sign in front of the voltage ν_i on the right-hand side of (4.4a) is the opposite of that in (4.3a), as explained by the opposite direction of the mechanical and electrical energy conversions.

4.2.3 Battery model

Each BESS has two reference response strategies, one for response to local frequency and one for response to an Automatic Generation Control (AGC) signal based on the average global frequency. The response to local frequency takes precedence. Each of these reference strategies has a deadband region with no response, a region of linear response, and a region of fixed response. Thus the battery response is given by:

$$\beta_i := 1_{\mathcal{E}}[\dot{\delta}_i](B^l(\dot{\delta}_i) - B_i^0) + (1 - 1_{\mathcal{E}}[\dot{\delta}_i])(B^g(\underline{\dot{\delta}}) - B_i^0). \tag{4.5}$$

Here, the indicator function $1_{\mathcal{E}}[\dot{\delta}_i]$ determines whether the local frequency at bus i lies in the range $\mathcal{E} \equiv \mathcal{E}^1 \cup \mathcal{E}^2$, where $\mathcal{E}^1 \equiv (-\infty, -F^m] \cup [F^m, \infty)$ and $\mathcal{E}^2 \equiv (-F^m, -F^n] \cup [F^n, F^m)$ (see Figure 4.2). The terms $(B^l(\dot{\delta}_i) - B_i^0)$ and $(B^g(\underline{\dot{\delta}}) - B_i^0)$, where the functions $B^l(\delta_i))$ and $B^g(\underline{\dot{\delta}}))$ are defined below, take account of the fact that the BESS will in general be responding at a pre-contingency level B_i^0 just prior to time 0. These terms therefore model the *change* in power output when the BESS delivers its reference response. Since we do not explicitly model the state of the system prior to time 0, the values of B_i^0 are randomly sampled initial conditions, see Section 4.3 for more details.

BESS response to local frequency δ_i

According to (4.5), if the local frequency δ_i lies in the range $\mathcal{E}$ then the BESS response is determined by this local frequency deviation. In this case the power-frequency

relation of the response is given by the blue graph in Figure 4.2:[5]

$$
B^l(\dot{\delta}_i) := \begin{cases} B^m \mathrm{sgn}[\dot{\delta}_i] & \dot{\delta}_i \in \mathcal{E}^1 \\ \left(M^l \dot{\delta}_i + Q^l \mathrm{sgn}[\dot{\delta}_i] \right) & \dot{\delta}_i \in \mathcal{E}^2 \\ 0 & \dot{\delta}_i \in \mathcal{R} \cup \mathcal{D} \end{cases}, \tag{4.6}
$$

where $\mathcal{R} \equiv (-F^n, -F^d] \cup [F^d, F^n)$, $\mathcal{D} \equiv (-F^d, F^d)$, B^m is the battery maximum power output (MPO) and

$$
M^l := \left(\frac{B^m - B^0}{F^n - F^m} \right) < 0 \text{ and } Q^l := \left| \frac{B^m F^n - B^0 F^m}{F^n - F^m} \right| \tag{4.7}
$$

are respectively the slope and intercept of the diagonal blue lines in Figure 4.2.

BESS response to system average frequency $\dot{\underline{\delta}}$

Again from (4.5), if the local frequency δ_i lies outside the range $\mathcal{E}$ then the BESS response is determined by the system average frequency deviation through the AGC signal. In this case the power-frequency relation of this response is given by the red graph in Figure 4.2:

$$
B^g(\dot{\underline{\delta}}) := \begin{cases} B^r \mathrm{sgn}[\dot{\underline{\delta}}] & \dot{\underline{\delta}} \in \mathcal{E} \\ (M^g \dot{\underline{\delta}} + Q^g \mathrm{sgn}[\dot{\underline{\delta}}]) & \dot{\underline{\delta}} \in \mathcal{R} \\ 0 & \dot{\underline{\delta}} \in \mathcal{D} \end{cases}, \tag{4.8}
$$

where
$$
M^g := \left(\frac{B^r - B^0}{F^d - F^n} \right) < 0, \text{ and } Q^g := \left| \frac{B^r F^d - B^0 F^n}{F^d - F^n} \right| \tag{4.9}
$$

are respectively the slope and intercept of the diagonal red lines in Figure 4.2 and the system average frequency deviation is calculated as

$$
\dot{\underline{\delta}} = \dot{\underline{\delta}}(t) := \frac{1}{N} \sum_{i=1}^{N} \dot{\delta}_i(T^b \lfloor t/T^b \rfloor). \tag{4.10}
$$

The form of (4.10) takes into account time delay in the calculation, broadcast and reception of the system-wide AGC signal. The discretised time index $T^b \lfloor t/T^b \rfloor$ corresponds to reception of the AGC signal at the BESS every T^b seconds.

---

[5]The superscript l stands for "local", superscript m stands for "maximum", superscript g stands for "global", superscript r stands for "regulation".
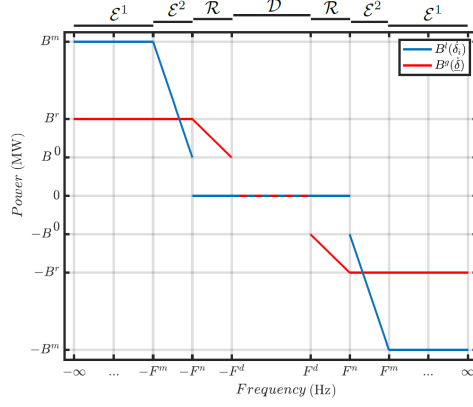
Figure 4.2: Reference power-frequency response strategies for the BESS response to local frequency (blue line) and the AGC signal (red line). As in (4.5), response to local frequency takes precedence. The common deadband $\mathcal{D}$ is indicated by the dashed central line.

4.3 Statistical Model

4.3.1 Unconditional distribution

At each bus i we sample a random power injection or drain η_i. These random disturbances are modelled agnostically and may, for example, represent exogenous shocks and/or endogenous contingencies. They are applied as impulses at time 0 and, as in [38], are modelled as constant power disturbances over the timescale of our simulations. Thus for example, if the model is used to explore the effect of renewable generation forecast errors then the joint distribution of these disturbances should reflect the historical joint distribution of such errors, and the timescale of the simulation should be comparable to the characteristic timescale of these errors. In the case study of Section 4.4, to enable a straightforward parametric exploration of the effect of correlations between disturbances, the vector $\underline{\eta}$ of disturbances is modelled as

$$\underline{\eta}(t) := \underline{U}\Theta[t] \in \mathbb{R}^{N+L} \tag{4.11}$$

where Θ is the Heaviside step function and $\underline{U}$ is the vector of disturbance magnitudes. More generally, if it is not judged reasonable to model disturbances as static over the timescale of the simulation, then dynamic disturbances can be accounted for by augmenting the simulation model with appropriate dynamics for the disturbances η_i in equation (4.3a).

Each U_i is normalised by setting it proportional to the stationary power injection at node i. Thus $\underline{U} = \underline{u} \circ \underline{P}^L$, where $\underline{P}^L$ is the vector of equilibrium nodal power loads, $\circ$ is the component-wise product and $\underline{u} \in \mathbb{R}^{N+L}$ is taken to be a Gaussian vector

$$\underline{u} \sim \mathcal{N}(\underline{0}, \Sigma) \quad \text{where} \quad \Sigma_{ij} := \sigma^2 \cdot \begin{cases} 1 & i = j \\ \alpha & i \neq j \end{cases} \quad i, j \in \{1, \ldots, N + L\}, \tag{4.12}$$

79

with $0 \leq \alpha < 1$ to ensure positive semidefiniteness.

The pre-contingency response of the battery just prior to time 0 is sampled from a uniform distribution

$$B_i^0 \sim \mathbb{U}\left[-\frac{b^+ B^m}{2}, \frac{b^+ B^m}{2}\right] \tag{4.13}$$

(see Appendix 8.1 for parameter specifications).

It is assumed that time 0 in our simulations can occur at any time in the interval between AGC signals. To reflect this, the time of the first AGC signal in the simulation is sampled uniformly:

$$T_0^b \sim \mathcal{U}[0, \ cT^b], \tag{4.14}$$

and subsequent AGC signals occur every T^b seconds.

4.3.2 Rare event sampler

Rare event sampling is performed using the skipping sampler, a Markov Chain Monte Carlo (MCMC) algorithm developed for this purpose. The sampler belongs to the class of Metropolis-Hastings (MH) algorithms and, as proved in [75] and demonstrated in the case study of Section 4.4, improves performance relative to the random walk Metropolis algorithm. Starting from any state (in the present context, a vector $\underline{u}_0 \in \mathbb{R}^{N+L}$ of power disturbances), a proposed new state $\underline{\tilde{u}} \in \mathbb{R}^{N+L}$ is sampled from a so-called proposal density. The proposal $\underline{\tilde{u}}$ is either accepted or rejected according to a specified acceptance probability. If it is accepted, the proposal is added to the output sample and becomes the new state $\underline{u}_1$. This procedure is repeated a desired number of times and the output sample $\{\underline{u}_1, \underline{u}_2, \ldots\}$ is returned. The skipping sampler is dedicated to sampling from any rare event Υ of interest, since its proposal density 'skips' over the unwanted region Υ^c until the rare event Υ is sampled (or until the skipping process is halted for reasons of computational efficiency, if this occurs first).

More precisely, in rare event sampling we are given an unconditional density μ on $\mathbb{R}^{N+L}$ and a rare event $\Upsilon \subset \mathbb{R}^{N+L}$ of interest and the task is to sample from this distribution conditional on Υ. The density of this conditional distribution at the point $\underline{u} \in \mathbb{R}^{N+L}$ is

$$\pi\left(\underline{u}\right) := \frac{\mu\left(\underline{u}\right) \mathbf{1}_\Upsilon\left(\underline{u}\right)}{\mu\left(\Upsilon\right)}, \tag{4.15}$$

where $\mu\left(\Upsilon\right)$ is the probability of the event Υ.

Pseudocode for the skipping sampler is given in Algorithm 5. Given the current state $\underline{u}$, the proposal $\underline{\tilde{u}}$ is a random walk proposal. If $\underline{\tilde{u}} \notin \Upsilon$, we calculate the direction $\underline{\Phi} = \left(\underline{\tilde{u}} - \underline{u}\right) / \|\underline{\tilde{u}} - \underline{u}\|$ between these two points, and move ('skip') a further independent random distance R_1 in this direction, where R_1 has the conditional

distribution of $\|\tilde{\underline{u}} - \underline{u}\|$ when conditioned on the observed value of $\underline{\Phi}$. If the modified proposal $\underline{Z} = \underline{u} + R_1 \underline{\Phi}$ lies in Υ then it is either accepted or rejected according to the acceptance probability $\alpha(\underline{u}, \underline{Z})$, otherwise the process skips again in the same direction $\underline{\Phi}$, by an independent random distance R_2 having the same distribution as R_1, and so on, until either Υ is entered or the budget for skipping is exhausted.

We take the unconditional distribution μ to be the Gaussian distribution in (4.12) above, and let the set Υ be a particular system instability. While the underlying sampler is the same as that employed in [38], in the case study of Section 4.4 below the system model and simulator aim to be more realistic and we are interested in instabilities relating to two different services provided by BESS (that is, regulation and emergency responses) and their interaction.

Algorithm 5: Skipping sampler (n-th step). Here q is a symmetric random walk proposal density and $q_{r|\Phi}(r|\Phi)$ is the conditional density of its polar radial coordinate r given the polar angle Φ, see [75] for full details.

Input : The n-th sample $\underline{u}_n$

Set $\underline{u} \coloneqq \underline{u}_n$;
Generate an initial proposal $\tilde{\underline{u}}$ distributed according to the density
 $q(\underline{y} - \underline{u}) d\underline{y}$;
Calculate the direction $\Phi = (\tilde{\underline{u}} - \underline{u}) / \|\tilde{\underline{u}} - \underline{u}\|$;
Generate a halting index $K \sim K_\varphi$;
Set $k = 1$ and $\underline{Z}_1 \coloneqq \tilde{\underline{u}}$;
while $\underline{Z}_k \notin \Upsilon$ **and** $k < K$ **do**
 | Generate a distance increment R distributed according to $q_{r|\Phi}(r|\Phi)$;
 | Set $\underline{Z}_{k+1} = \underline{Z}_k + \Phi R$;
 | Increase w by one;
end
Set $\underline{Z} \coloneqq \underline{Z}_k$;
Evaluate the acceptance probability:

$$
\alpha(\underline{u}, \underline{Z}) = \begin{cases} \min\left(1, \frac{\pi(\underline{Z})}{\pi(\underline{u})}\right) & \text{if } \pi(\underline{u}) \neq 0, \\ 1, & \text{otherwise,} \end{cases} \tag{4.16}
$$

Generate a uniform random variable V on $(0,1)$;
if $V \leq \alpha(\underline{u}, \underline{Z})$ **then**
 | $\underline{u}_{n+1} = \underline{Z}$;
else
 | $\underline{u}_{n+1} = \underline{u}$;
end
return $\underline{u}_{n+1}$.

4.4 Case Study

4.4.1 Kundur Two-Area System

Our case study is based on the Kundur two-area system (KTAS) [57]. In particular we take a Kron reduced version (see for example [24, 23]) consisting of $N = 4$ generation buses and $L = 2$ load buses as in Figure 4.3. For simplicity the load buses are equipped with equally specified batteries. In equilibrium, power flows from Area 1 to Area 2 through the line connecting nodes 5 and 6. This is modelled as a weak tie line (see for example [90]) with a disconnection ('line tripping') scheme. The system parameters (whose values can be found in Appendix 8.1) are such that the system is $N - 1$ secure, in the sense that the loss of a generator (in the absence of any other disturbance) does not trigger an emergency response.
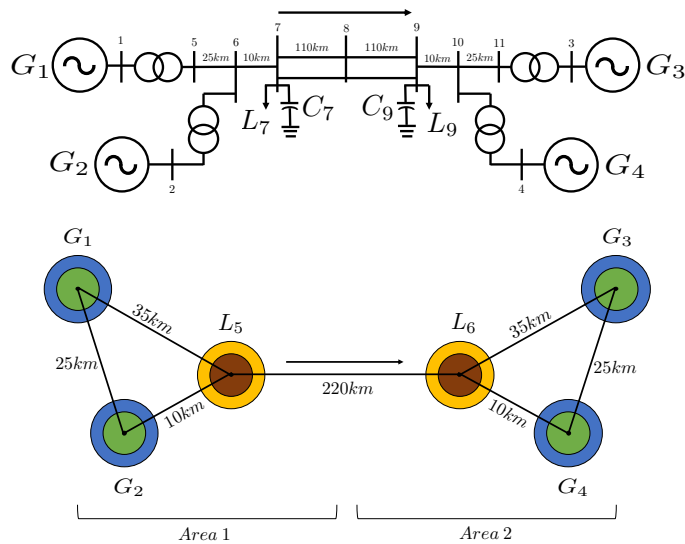

Figure 4.3: (Top) Single line diagram of the Kundur two-area 4 node network, reproduced from [57], before Kron reduction. (Bottom) Schematic drawing of the same system after Kron reduction. Generator buses (green circles) correspond to nodes $i = 1, \ldots, 4$ and load buses (brown circles) correspond to nodes $i = 5, 6$. Governors (blue annuli) are attached to generator buses while BESS (yellow annuli) are located at load buses. Line lengths are indicated.

4.4.2 Metrics for instability

In order to explore the frequency regulation benefits of BESS, the first metric we apply is the frequency excursion area (FEA) which may be defined per bus as

$$\mathcal{F}_i(t) := \int_0^t \left\{ \Theta[\dot{\delta}_i(s) - \mathcal{F}^+](\dot{\delta}_i(s) - \mathcal{F}^+)ds + \Theta[\mathcal{F}^- - \dot{\delta}_i(s)](\mathcal{F}^- - \dot{\delta}_i(s)) \right\} ds, \quad (4.17)$$

where $\mathcal{F}^+$ and $\mathcal{F}^-$ are constant thresholds. Thus the FEA is the area lying between the graph of frequency at bus i and the frequency band $[\mathcal{F}^+, \mathcal{F}^-]$.

To measure the emergency response benefits of BESS, the metric of interest is the amount of load shed (equivalently, the numbers Γ_i of load shed events at the load buses, as defined in Section 2.5.2 of Chapter 2). The system averages of these quantities are then

$$\mathscr{F}(t) := \frac{1}{N+L} \sum_{i=1}^{N+L} \mathcal{F}_i(t) \quad \text{and} \quad \mathscr{G}(t) := \frac{1}{L} \sum_{i=N+1}^{N+L} \Gamma_i(t) \tag{4.18}$$

In equilibrium we have $\mathscr{F}(t) = 0$ and $\mathscr{G}(t) = 0$. Recalling Section 34.3.2, the rare events of interest are the sets $\Upsilon^{\mathscr{F}}$ and $\Upsilon^{\mathscr{G}}$, corresponding to frequency excursions and load shedding respectively, where:

$$\Upsilon^{\mathscr{F}} := \{ \underline{u} \in \mathbb{R}^{N+L} : \mathscr{F}(T) > 0 \} \quad \text{and} \quad \Upsilon^{\mathscr{G}} := \{ \underline{u} \in \mathbb{R}^{N+L} : \mathscr{G}(T) > 0 \}. \tag{4.19}$$

Here $T > 0$ is the length of the observation time window considered. In direct analogy with the well-known conditional value at risk (CVaR) metric, we may then define the conditional average load shed as

$$\langle \mathscr{G}(T) \rangle := \sum_{i=1}^{L\Delta} \mathscr{G}(T) \cdot \mathbb{P}[\mathscr{G}(T) = i | \Upsilon^{\mathscr{G}}], \tag{4.20}$$

where Δ is the maximum possible number of load shed events per node (Section 2.5.2 of Chapter 2). Thus $\langle \mathscr{G}(T) \rangle$ measures the severity of load shedding, given that it occurs. We may similarly define the conditional average FEA $\langle \mathscr{F}(T) \rangle$, measuring the severity of frequency excursions when they occur. The conditional samples output by the skipping sampler of Section 4.3.2 may be used to calculate the simulation values of these conditional metrics.

4.4.3 Skipping sampler implementation

In Algorithm 5, the halting index is taken to be constant and the proposal q to be Gaussian, so that the conditional distribution $q_{r|\varphi}$ of the polar radial coordinate r of the initial proposal is the generalised Gamma distribution

$$q_{r|\varphi}(r|\varphi) = \frac{(\varphi^T \Sigma^{-1} \varphi)^{\frac{N+L}{2}}}{2^{\frac{N+L}{2}-1} \Gamma(\frac{N+L}{2})} e^{-(\varphi^T \Sigma^{-1} \varphi)\frac{r^2}{2}} r^{N+L-1}. \tag{4.21}$$

4.5 Results

In this section we explore how the frequency regulation and emergency response benefits of BESS vary with both the BESS Maximum Power Output (MPO) B^m and the correlation parameter α for disturbances. Based on BESS characteristics anticipated in the near future [3] we take B^m from 0 (no BESS) to 1000MW (the

maximum BESS MPO considered). Simulations were conducted using MATLAB®, each taking between 2 and 5 seconds to execute for a given vector of initial conditions using a desktop machine. To compare the computational complexity of the skipping sampler to standard Metropolis-Hastings and Monte Carlo sampling, in the study of Section 4.5.1 with 1GW battery, independent 15 minute runs generated 112, 63, and 9 samples from the skipping sampler, MH, and Monte Carlo samplers respectively. In the study of Section 4.5.2 with 1GW battery, independent 15 minute runs generated 95, 53 and 3 samples from the skipping sampler, MH, and Monte Carlo samplers respectively.

4.5.1 Frequency regulation benefits of BESS

Figure 4.4 (Left) illustrates the relationship between battery MPO and $\langle \mathscr{F}(T) \rangle$, the conditional average frequency excursion area, for different values of α. The results reveal a clear decreasing trend in $\langle \mathscr{F}(T) \rangle$, approximately in the interval $[0, 400]$ MW for MPO, while the frequency regulation benefit tends to saturate beyond this level.

Figure 4.4 (Centre) plots the conditional average magnitude of disturbances in the rare set $\Upsilon^{\mathscr{F}}$, showing an increasing trend. Recalling the 'bell-shaped curve' of the Gaussian distribution, this indicates that the event $\Upsilon^{\mathscr{F}}$ becomes increasingly rare as battery MPO increases. Thus higher MPOs correspond to a reduction in both the likelihood of frequency excursions outside the normal operating band and, when they occur, the frequency excursions are less severe on average. While this reduction in severity saturates as MPO increases to 1GW, the reduction in frequency appears to be maintained up to this level.

These measures of severity and likelihood may be combined to produce a normalised version of conditional Average FEA, calculated by dividing $\langle \mathscr{F}(T) \rangle$ by the average magnitude of disturbances in the set $\Upsilon^{\mathscr{F}}$. This quantity is illustrated in Figure 4.4 (Right). Although this shows the same trend as Figure 4.4 (Left) in the present case of FEA, it is of greater interest in the next section.

4.5.2 Emergency response benefits of BESS

Figure 4.5 replicates the plots introduced in Figure 4.4, this time for the conditional average load shed metric. Recalling the above discussion of Figure 4.4, the relationships are more complex when we consider load shedding, a rarer instability. In the left panel, the conditional average load shed is initially approximately constant as MPO increases, after which a significant decreasing relationship is seen. From the centre panel, the conditional average disturbance size initially increases, after which a significant decrease is observed. Together this implies that as MPO increases from 0 to 1GW, the initial benefit of BESS is in reducing the likelihood of load shedding while its severity remains approximately constant. After this initial
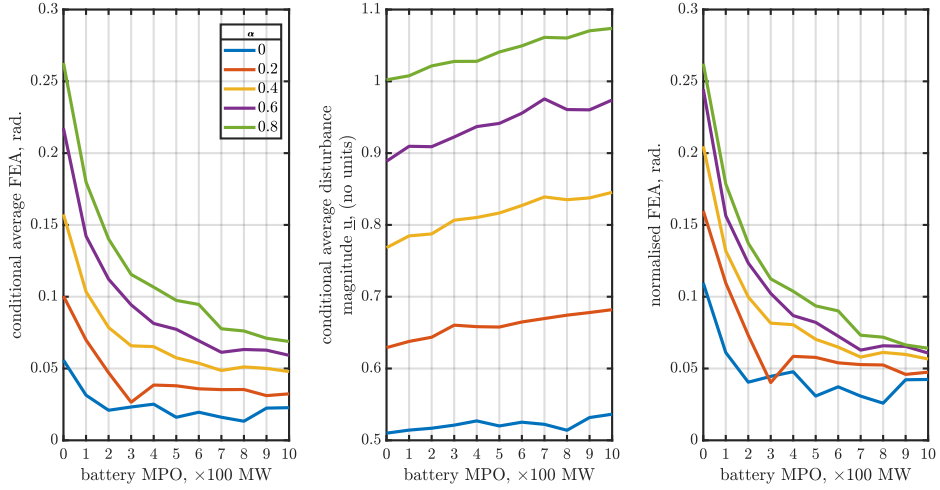
Figure 4.4: Conditioned on the rare event $\Upsilon^{\mathscr{F}}$ (occurrence of a frequency excursion outside normal operating band), the above plots illustrate (Left): Plot of the conditional average FEA; (Centre): Plot of conditional average disturbance magnitude against battery MPO; (Right): Plot of normalised FEA versus battery MPO.

trend, significant reductions in severity are balanced against significant increases in the likelihood of load shedding (as evidenced by the decrease in conditional average disturbance size). When these two aspects are weighed against each other in the right panel through normalisation, a consistent improvement in system resilience under this metric is observed with increasing MPO, with the exception of the cases of highest correlation between disturbances ($\alpha = 0.6, 0.8$).

Figure 4.6 investigates further the phenomena observed in Figure 4.5. As shown in the left panel, the increase in normalised load shedding seen for $\alpha = 0.6, 0.8$ coincides with an increase in disconnection of the weak tie line between nodes 5 and 6. Further investigation (data not shown) reveals that in this parameter range, load shedding was associated with the occurrence of large negative disturbances in Area 2 of the KTAS. A plausible explanation is that with this combination of parameters and disturbances, the BESS in Area 1 contributes to excessive power transfer to Area 2, resulting in disconnection of the weak tie line. The additional transient dynamics resulting from the consequent system separation exacerbate the on-going cascade of both load shedding and generation shedding (right panel), reducing system resilience. Additionally, the centre panel of Figure 4.5 reveals a different but similarly complex relationship between MPO and the probability of generation shedding due to RoCoF.

4.5.3 Interaction between frequency regulation and contingency services

Recall from Section 4.3 that to account for the frequency regulation service provided by BESS, the power injection or withdrawal B_i^0 from each battery just before time 0
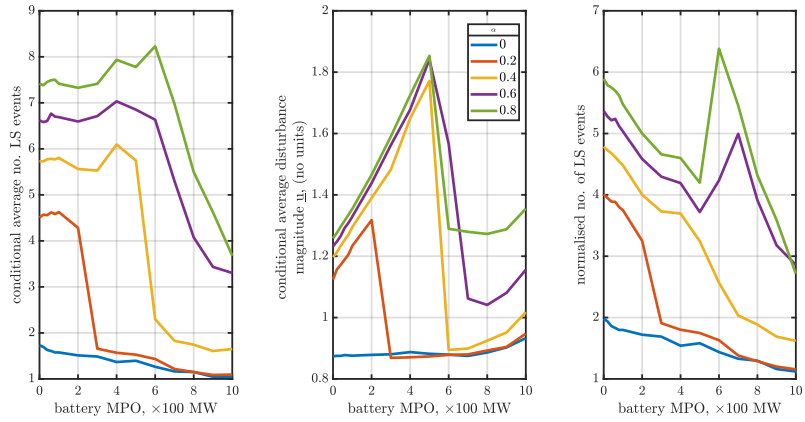
Figure 4.5: Conditioned on the rare event $\Upsilon^{\mathscr{G}}$ (occurrence of at least one load shed event), the above plots illustrate (Left): Plot of the conditional average load shed; (Centre): Plot of conditional average disturbance magnitude against battery MPO; (Right): Plot of normalised load shed versus battery MPO.
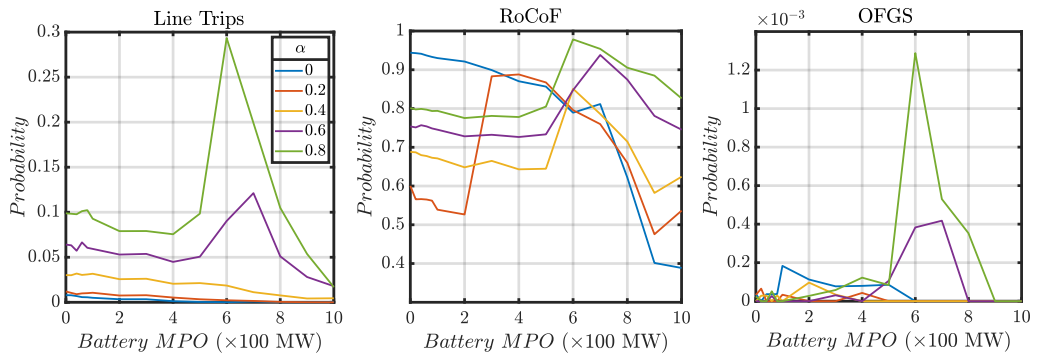

Figure 4.6: Conditioned on the rare event $\Upsilon^{\mathscr{G}}$ (occurrence of at least one load shed event), the above plots illustrate (Left): Conditional probability of line trip; (Centre): Conditional probability of generation shedding due to rate of change of frequency; (Right) Conditional probability of over frequency generation shedding.

is sampled from a uniform distribution. Also, from equation (4.5), this initial state determines the maximum power available to respond to the disturbances at time 0. To confirm that the model captures interaction of the frequency regulation and emergency responses, the top row of Figure 4.7 plots the full distribution of B_i^0 at the two buses with BESS ($i = 5, 6$), when conditioned on $\Upsilon^{\mathcal{G}}$ (that is, conditioned on load shedding). The top left pair plots results for MPO 200MW, and from visual inspection the conditional distribution is again approximately uniform. In contrast the top right pair of distributions, which plots results for MPO 1GW, are markedly skewed. This indicates interaction between the frequency regulation and contingency services, since load shedding is associated with high values of B_5^0 (the initial BESS power in Area 1) and low values of B_6^0 (the initial BESS power in Area 2).
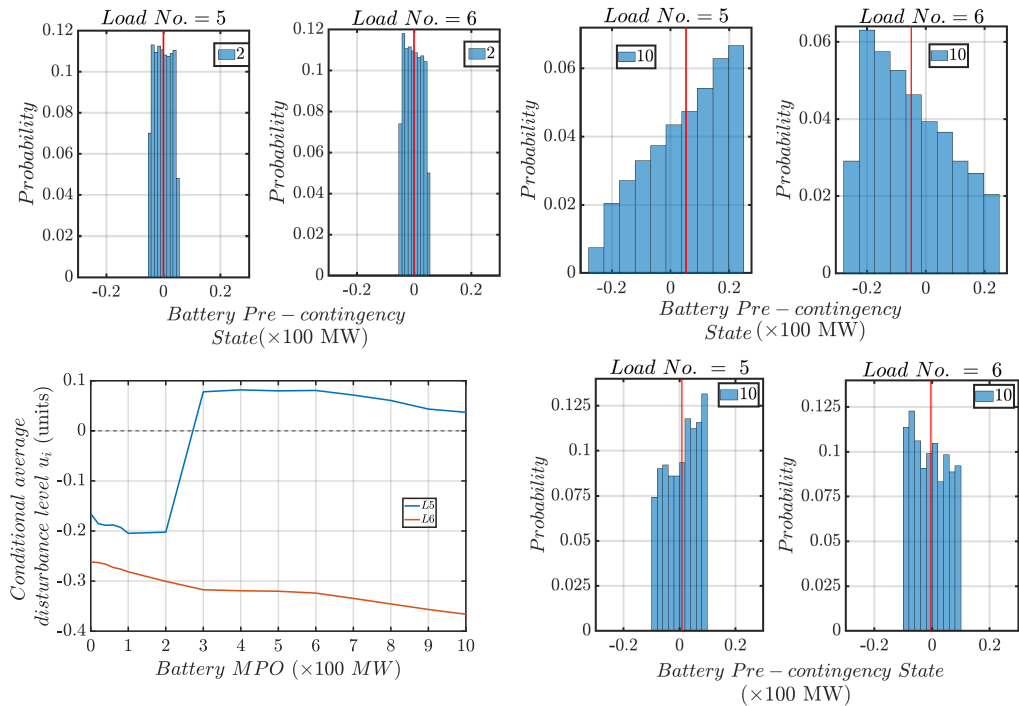

Figure 4.7: Empirical distributions of the BESS initial output power B_i^0 for $i = 5, 6$, conditional on the event $\Upsilon^{\mathcal{G}}$ (that is, conditional on load shedding), for $\sigma^2 = 0.003$ and $\alpha = 0.2$. Vertical red lines locate the conditional mean. (Top Left Pair): MPO 200 MW, $b^+ = 1$. (Top Right Pair): MPO 1GW, $b^+ = 1$. (Bottom Right Pair): MPO 1GW, $b^+ = 0.2$. (Bottom Left): Plot of the conditional average disturbance level (including sign) at buses 5 and 6 versus battery MPO.

Indeed the bottom left panel shows that, for MPO 1GW, the conditional average disturbance at node 5 is positive, while it is negative at node 6. Thus in this case study, load shedding is associated with disturbances which are in the same direction and thus reinforce the initial power outputs B_i^0 ($i = 5, 6$), rather than being in the opposite direction and thus partially cancelling them. The bottom right panel confirms that this potential issue may be mitigated by judicious choice of the parameter b^+ in (4.13): taking MPO 1GW and $b^+ = 0.2$, the conditional distribution of B_i^0 is markedly less skewed.

4.6 Conclusions and outlook

Our results demonstrate successful application of rare event sampling to a detailed power system dynamic simulation model including BESS. The method accounts for the feedback effects of emergency responses and is flexible, producing a wide variety of empirical statistics capable of distinguishing between the likelihood and severity of instabilities, and of identifying both saturation in these benefits as MPO increases and competition between the frequency regulation and emergency responses provided by BESS. Possible future research directions include extensions to further protection systems such as voltage regulation, more detailed sensitivity analyses including the influence of BESS speed of response, and investigation of the effect of different network topologies.

Chapter 5

Skipping between distant basins

This chapter presents the manuscript entitled 'Skipping between distant basins', submitted to the Journal of Global Optimization in August 2021. At the time of compiling this thesis the article has received positive reviews and the authors are amending the manuscript with the suggestions provided by the reviewers for publication.

Maldon Goodridge[1] John Moriarty [2] Jure Vogrinc[3] Alessandro Zocca[4]

Abstract

We present and analyse the Basin Hopping with Skipping (BH-S) algorithm for stochastic optimisation. This algorithm replaces the perturbation step of basin hopping (BH) with a so-called skipping mechanism from the rare-event sampling. Empirical results on benchmark optimisation surfaces demonstrate that BH-S can improve performance relative to BH by encouraging non-local exploration, that is, by hopping between distant basins.

keywords: Basin hopping, stochastic optimisation, skipping sampler, rare events, Markov chains

5.1   Introduction and background

In the literature on global optimisation of a non-convex energy landscape a source of inspiration has been methods from the theory of rare-event sampling. Examples include the methods of cross-entropy for combinatorial and continuous optimisation [96] and, more recently, splitting for optimisation [28]. In stochastic optimisation algorithms such as random search [102], basin hopping [61, 111], simulated annealing [55] and the multistart method [47, 69], one or more initial points X_0 are perturbed in order to discover new neighbourhoods (or 'basins') of lower energy, which may then be explored by a local procedure such as gradient descent. As such algorithms discover progressively smaller energy values, the remaining lower-energy basins form a decreasing sequence of sets. Viewing the optimisation domain heuristically as a probability space and these basins as events, the discovery of smaller energy values can then also be likened to rare-event sampling.

In this analogy, the local perturbation step plays a similar role to the proposal step in a Markov Chain Monte Carlo (MCMC) sampler (see [13], [93]). Thus in order to enhance performance, one may explore the use of alternative MCMC proposal distributions developed in the context of rare event sampling as alternative perturbation steps within stochastic optimisation routines. This is the approach we take in the present paper.

---

[1]School of Mathematics, Queen Mary, University of London, London, E1 4NS, UK, ahw493@qmul.ac.uk

[2]School of Mathematics, Queen Mary, University of London, London, E1 4NS, UK j.moriarty@qmul.ac.uk

[3]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK Jure.Vogrinc@warwick.ac.uk

[4]Department of Mathematics, Vrije Universiteit Amsterdam, 1081HV, Netherlands, a.zocca@vu.nl

To illustrate the potential benefit of this approach, consider the problem of metastability in the study of dynamical systems, where an energy function governs the evolution of a stochastic process. Metastable regions correspond to local minima of the energy landscape, which may be rare and well-separated by so-called *energy barriers*. Particles which follow the dynamics defined by the energy function may therefore remain in such local minima for extended periods of time, with the average switching time between metastable events being orders of magnitude longer than the timescale of the process itself [10]. If X_0 lies in one such minima, separation means that local perturbations are not well suited to the direct discovery of another basin. Instead, algorithms using local perturbations to minimise over such a landscape should be non-monotonic, accepting transitions from X_0 to states of higher energy in the hope of later reaching lower-energy basins. In contrast, since non-local perturbation steps offer the possibility of direct moves between distant low-energy basins, they may possibly be effective on such surfaces within a monotonic optimisation algorithm. In this paper we explore the use of a particular non-local perturbation, the 'skipping perturbation' of [76].

Although other non-local perturbations have been proposed in the literature (see for example [6, 89, 60, 104, 105, 108] in the context of MCMC), skipping has the advantage of being just as straightforward to implement as a local random walk perturbation. That is, it requires no additional information about the energy landscape beyond the ability to evaluate it pointwise.

We explore its use within the basin hopping (BH) algorithm [61, 111], which combines local optimisation with perturbation steps and requires only pointwise evaluations of the energy function f. The resulting 'basin hopping with skipping' (BH-S) algorithm is thus as generally applicable as the BH algorithm.

The BH algorithm works as follows: the current state X_n is perturbed via a random walk step to give Y_n which is, in turn, mapped via deterministic local minimisation to a local minimum X_{n+1}. This local minimum point is then either accepted or rejected as the new state with a probability given by the Metropolis acceptance ratio, and the procedure is repeated until a pre-determined stopping criterion is met. Due to its effectiveness and ease of implementation, the BH algorithm has been used to solve a wide variety of optimisation problems (see [44, 82, 86] for more details).

The BH-S replaces the random walk step of BH with a skipping perturbation over the sublevel set of the current state X_n. Like a flat stone skimming across water, this involves repeated perturbations in a straight line until either a point of lower energy is found, or the skipping process is halted. In contrast with the non-monotonic BH algorithm, BH-S is monotonic- that is, if the current state of the BH-S is X_n, the BH-S will only transition to a new state X_{n+1} if $f(X_{n+1}) < f(X_n)$, where f is the energy function to be minimised. The BH-S algorithm, which was first outlined in [76], thus provides a direct mechanism to escape local minima which contrasts

with the indirect approach taken by BH. Another perspective is that BH-S alters the balance between the computational effort expended on local optimisation versus the effort spent on perturbation, typically increasing the latter while decreasing the former (cf. Table 5.1 below).

This paper investigates the interaction between the choice of parameters and type of landscape by using a set of benchmark problems to present a systematic overview on the types of optimisation problem on which BH-S tends to outperform BH. The rest of the paper is structured as follows: Section 5.2 introduces the algorithms, empirical results are presented and discussed in Section 5.3, and Section 5.4 concludes.

5.2   The BH-S algorithm

Consider the box constrained global optimisation problem on a rectangular subdomain $D \subset \mathbb{R}^d$, of the form

$$\min \quad f(\boldsymbol{x}) \quad \text{s.t.} \ \ \boldsymbol{x} \in D := \prod_{i=1}^{d} [l_i, u_i], \tag{5.1}$$

for some scalars $l_i \leq u_i$, $i = 1, \ldots, d$. In the rest of the paper, we will often refer to f as the *energy function* and to its graph as the *energy landscape*. This terminology, which is similar to that of simulated annealing, is appropriate, since the BH algorithm was originally conceived as a method to find the lowest energy configuration of a molecular system [111]. In this section, we review the BH algorithm and then introduce basin hopping with skipping (BH-S).

5.2.1   Basin hopping algorithm

The core idea of the basin hopping algorithm [111], which is presented in Algorithm 6, is to supplement local deterministic optimisation by alternating it with a random perturbation step capable of escaping local minima. More specifically, inside the `RandomPerturbation` procedure at step 5 of Algorithm 6, a random perturbation $W \in \mathbb{R}^d$ is drawn and added to the current state X_n giving a state $Y_n = X_n + W$. Most commonly, the increment W is either spherically symmetric or has independent coordinates. The state Y_n becomes the starting point of a deterministic local minimisation routine. In our implementation of Algorithm 6, the `LocalMinimisation` procedure at step 6 is performed using the limited-memory BFGS algorithm [66], a quasi-Newton method capable of incorporating boundary constraints, although we note that other choices are possible. The resulting local minimum U_n is then either

accepted or rejected as the new state with probability equal to

$$\min\left(1, \exp\left(-\frac{f(U_n) - f(X_n)}{T}\right)\right),$$

where $T \geq 0$ is a fixed temperature parameter. This means, in particular, that downwards steps for which $f(U_n) < f(X_n)$ are always accepted. The BH algorithm prescribes to repeat this basic step until a pre-defined stopping criterion is satisfied. Commonly used stopping criteria for the BH algorithm include, among others, a limit on the number of evaluations of the function f or the absence of improvement over several consecutive iterations [82, 95]. The monotonic basin hopping method introduced in [61] is the BH variant corresponding to the limiting case $T = 0$, in which all steps that increase the energy are rejected.

Algorithm 6: Basin hopping

1 Generate a random initial state $Y_0 \in D$;
2 $X_0 = \texttt{LocalMinimisation}(Y_0)$;
3 $n = 0$;
4 **while** *Stopping criterion for $\{X_j\}_{j \leq n}$ is not satisfied:* **do**
5 $Y_n = \texttt{RandomPerturbation}(X_n)$;
6 $U_n = \texttt{LocalMinimisation}(Y_n)$;
7   Generate $V \sim \text{Uniform}([0,1])$;
8   **if** $V < \min\left(1, \exp\left(-\frac{f(U_n) - f(X_n)}{T}\right)\right)$ **then**
9 $X_{n+1} = U_n$;
10   **else**
11 $X_{n+1} = X_n$;
12   **end**
13   Increase n by 1;
14 **end**

Basin hopping can thus be viewed as a random walk on the set of local minima of the energy landscape, which, due to its transition probabilities favour downhill moves to lower minima, is capable of finding the global minimum and, hence, of solving global optimisation problems. Its transition probabilities depend in a complex way on the current position, the landscape, and the perturbation step. The BH-S algorithm introduced in the next section modifies these transition probabilities, aiming to accelerate optimisation.

5.2.2 Skipping perturbations and the BH-S algorithm

In this subsection we introduce the BH-S algorithm, which differs from BH only in the perturbation step of line 5 in Algorithm 6. Instead of the random walk perturbation described above, the `RandomPerturbation` procedure described in Algorithm 7 below is applied in order to obtain Y_n. The `LocalMinimisation` and acceptance steps remain identical to those in Algorithm 6.

Given the current state X_n and a fixed probability density q on $\mathbb{R}^d$, the random walk perturbation of the BH algorithm can be understood as drawing a state Y_n from the density $y \mapsto q(y - X_n)$. In contrast, the *skipping perturbation* of BH-S depends on both the current state X_n and a *target set* $C \subseteq \mathbb{R}^d$ of states. The target set C_n for the n-th skipping perturbation is the sublevel set of the energy function f at the current point X_n, i.e.,

$$C_n := \{x \in D : f(x) \leq f(X_n)\} \subset \mathbb{R}^d. \tag{5.2}$$

A state Z_1 is drawn according to the density q just as in the random walk perturbation and, if Z_1 does not lie in the target set C_n, further states $Z_2, Z_3, \ldots$ are drawn such that $X_n, Z_1, Z_2, \ldots$ lie in order on a straight line, with each distance increment $|Z_{j+1} - Z_j|$ having the same distribution as that of $|Z_1 - X_n|$ conditioned on the line's direction $\frac{Z_1 - X_n}{|Z_1 - X_n|}$. The first state of this sequence to land in the target set C_n becomes the state Y_n. If C_n is not entered before the skipping process is halted, then Y_n is set equal to X_n.

More precisely, let $x = (r, \varphi)$ be polar coordinates on $\mathbb{R}^d$ with the angular part φ lying on the $d - 1$ dimensional unit sphere $\mathbb{S}^{d-1}$. Write $\varphi \mapsto q_\varphi(\varphi)$ for the marginal density of q with respect to the angular part φ, which we may call the *directional density* (and which we assume is strictly positive). For each $\varphi \in \mathbb{S}^{d-1}$ denote by

$$q_{r|\varphi}(r|\varphi) := \frac{q_{r,\varphi}(r, \varphi)}{q_\varphi(\varphi)}$$

the *conditional jump density*, i.e., the conditional density of the radial part r given the direction φ. To construct the skipping perturbation, set $Z_0 = X_n$ and draw a random direction $\Phi \in \mathbb{S}^{d-1}$ from the directional density q_φ. A sequence of i.i.d. distances $R_1, R_2, \ldots$ is then drawn from the conditional jump density $q_{r|\Phi}$, defining a sequence of modified perturbations $\{Z_k\}_{k \geq 1}$ on $\mathbb{R}^d$ by

$$Z_{k+1} := Z_k + \Phi R_{k+1}, \qquad k = 0, 1, \ldots$$

Since this modification of the BH perturbation is more likely to generate states Z_k lying outside the optimisation domain D, we apply periodic boundary conditions.

If $Z_k \in C_n$ for some $k \leq K$, where K is a pre-defined maximum number of steps called the *halting index*, then we set $Y_n = Z_k$ in Algorithm 6 and continue to the `LocalMinimisation` and acceptance steps. Alternatively if $Z_k \notin C_n$ for all $k \leq K$ we set $Y_n = X_n$. Note that although in [76] the halting index K can be randomised, in the present setting with a known bounded domain D it is sufficient to consider only fixed halting indices.

For clarity, in the remainder of the paper we will understand the BH algorithm to mean setting $K = 1$ in Algorithm 7. In all simulations we set the perturbation

q to be a spherically symmetric and Gaussian with standard deviation σ, although other choices are possible (see the discussion in Section 4.1 of [76]). In the next section we explore for which types of energy function f BH-S offers an advantage over BH, and also discuss the choice of the halting index.

Algorithm 7: `RandomPerturbation` subroutine for BH-S

Input : State $X_n \in \mathbb{R}^d$
Output: Randomly perturbed state $Y_n \in \mathbb{R}^d$

1 Set $Z_0 = X_n$;
2 Generate an initial perturbation W distributed according to the density $w \mapsto q(w - X_n)$;
3 Calculate the direction

$$\Phi \quad = \quad \frac{(W - X_n)}{\|W - X_n\|} \; ;$$

Set $k = 1$ and $Z_1 := W$;
4 **while** $f(Z_k) > f(X_n)$ **and** $k < K$ **do**
5 $\quad$ Generate an independent distance increment R distributed as $\|W - X_n\|$ given Φ ;
6 $\quad$ Set $Z_{k+1} = Z_k + \Phi R$;
7 $\quad$ Increase k by one ;
8 **end**
9 Set $Y_n := Z_k$;

5.3 Empirical results

In this section, we aim to explore on which types of optimisation problem BH-S tends to outperform BH and *vice versa* using a set of benchmark energy landscapes with known global minima from [32, 49, 106]. To facilitate discussion of landscape geometry, we initially restrict attention to two-dimensional energy functions, before considering higher dimensions in Section 5.3.6.

In Subsection 5.3.3 we show that, if an energy landscape has *distant basins* (recall that with the word 'basin' we refer to the neighbourhood of a local minimum) then BH-S tends to offer an advantage. Otherwise, as described in Subsection 5.3.4, BH is to be preferred since any benefit from BH-S is then typically outweighed by its additional computational overhead. We also explore the effect of the state space dimension d on the performance of both algorithms and offer guidance on tuning the BH-S method, including strategies to improve exploration of challenging energy landscapes.

5.3.1 Methodology

For each benchmark energy landscape, we compare the performance of BH-S to that of BH with temperature $T = 1$, in both cases taking the density q of the initial perturbation as the centred Gaussian

$$q \sim \mathcal{N}(\mathbf{0}, \sigma^2 \cdot I_d), \tag{5.3}$$

where I_d is the $d \times d$ identity matrix and the parameter σ allows for tuning, as follows. Both the BH and BH-S algorithms are run on a set of uniformly distributed initial states $I := \{X_0^{(n)} \in D, \, n = 1, \ldots, |I|\}$. These initial states are used sequentially until the computational budget of 300 seconds of CPU time has elapsed, and the corresponding set of final states is recorded. To account for numerical tolerance, we consider a run to have successfully identified an element x^* in the set of global minimum points in D, if its final state lies in $\mathcal{G} := \{x \in \mathbb{R}^d \; : \; ||x - x^*|| \leq 10^{-5}\}$. This choice excludes all non-global minima for all benchmark landscapes.

The performance of each algorithm is then assessed with respect to two metrics:

- **Effectiveness**, defined as the proportion of runs terminating in $\mathcal{G}$,

- **Efficiency,** defined as the number of runs terminating in $\mathcal{G}$.

We write ρ_c and ρ_s for the effectiveness of the BH and BH-S algorithms respectively, while ϵ_c and ϵ_s denote their respective efficiencies. The BH and BH-S algorithms are individually tuned for each function by selecting σ and K to maximise their efficiency.

In order to understand the role played by the skipping perturbation, we also record diagnostics on the average size of perturbations. For each new state $X_{n+1} \neq X_n$ accepted in Algorithm 6, define the *perturbation distance J* as $||Y_n - X_n||$, the Euclidean distance between the state X_n at step n and its perturbation Y_n. For each run of an algorithm, the mean $\overline{J}$ of these perturbation distances is recorded. Then for each 300 second budget, the *expected mean jump distance υ* is the average $\upsilon := N^{-1} \sum_{n=1}^{N} \overline{J}^{(n)}$, where N is the number of runs realised within the time budget. For the BH-S algorithm, υ is calculated separately for the accepted random walk perturbations (that is, those for which $Y_n = Z_1$ in Algorithm 7) and the accepted skipping perturbations (those for which $Y_n = Z_k$ with $k \geq 2$ in Algorithm 7), denoting these by υ_1 and υ_s respectively.

The simulations were conducted on a single core using Python 3.7, using the `basinhopping` and limited-memory BFGS routines in SciPy version 1.6.2 for the BH algorithm. Results for all considered landscapes are presented in the Appendix.

5.3.2 Exploratory analysis

As an exploratory comparison between BH and BH-S, their relative effectiveness ρ_s/ρ_c and efficiency ϵ_s/ϵ_c were calculated for each benchmark energy landscape and plotted in Figure 5.1.
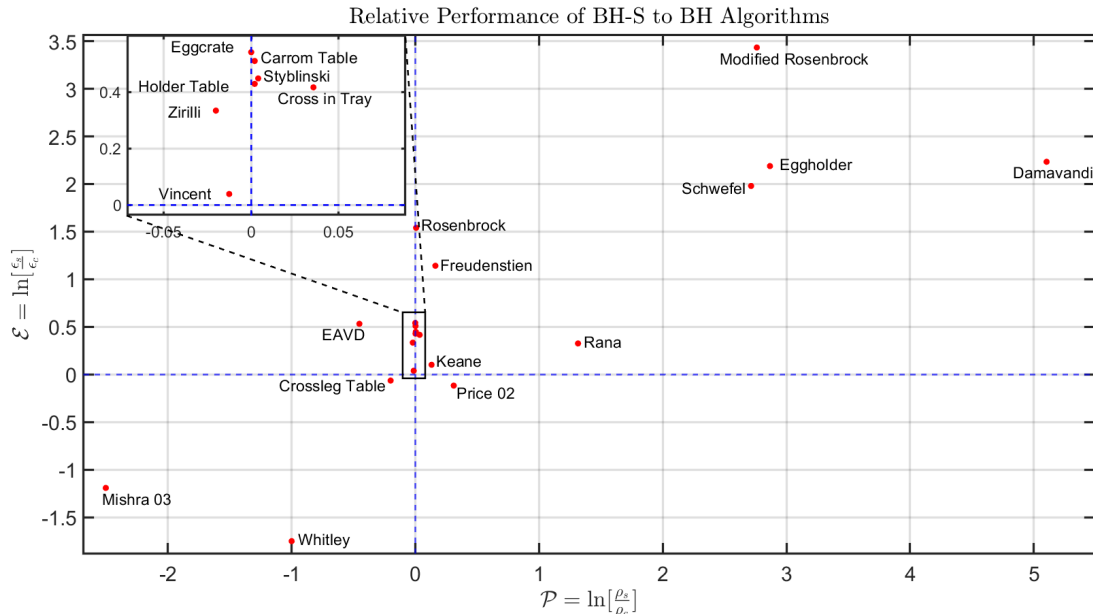

Figure 5.1: Scatter-plot of relative efficiency $\mathcal{E} = \ln(\epsilon_s/\epsilon_c)$ versus relative effectiveness $\mathcal{P} = \ln(\rho_s/\rho_c)$ for the BH and BH-S algorithms on benchmark energy landscapes. See Appendix 8.2 for detailed performance metrics of the BH-S and BH algorithms for each benchmark energy landscape.

Landscapes in the first quadrant of Figure 5.1 represent cases where the BH-S algorithm exhibits both greater effectiveness and greater efficiency than BH. The common feature among these landscapes, which are plotted in the Appendix, might be called *distant basins*: that is, basins separated by sufficient Euclidean distance that the random walk performed by the BH algorithm is unlikely to transition directly between them. While indirect transitions between such basins may be possible, they require a suitable combination of steps to be made. Such indirect transitions may carry significant computational expense, for example if suitable combinations of steps are long or relatively unlikely. In the BH-S algorithm, by contrast, the linear sequence of steps taken by the skipping perturbation enables direct transitions even between distant basins.

Conversely, landscapes lying in the lower-left quadrant of Figure 5.1 represent cases where the BH-S algorithm is both less reliable and less efficient than BH. As explored more extensively later in Subsection 5.3.4, for each of these landscapes, if the energy of the state X_n is close to the global minimum value $f(x^*)$ then the corresponding sublevel set C_n has almost zero volume. This means that even if the skipping perturbation traverses the distance between basins, the states $Z_1, \ldots, Z_k$ are unlikely to fall in C_n due to its small volume. Since the BH algorithm is non-

monotonic, it does not suffer from the same issue and outperforms BH-S for these landscapes.

Figure 5.1 displays a positive correlation between relative efficiency and relative effectiveness. However for several landscapes (which lie near the vertical axis) the performance of BH and BH-S cannot be clearly distinguished on the basis of effectiveness alone. As confirmed by the Appendix, this is typically because both algorithms have effectiveness close to 100%. Nevertheless the algorithms differ in their efficiency, with BH-S observed to be more efficient than BH for each such landscape. One surface also lies in each of the second and fourth quadrants.

Table 5.1: CPU time spent on the perturbation and local minimisation steps by the BH and BH-S algorithms for the test functions discussed in Sections 5.3.3–5.3.5, normalised by efficiency.

| | | $\frac{\text{Time spent}}{\text{Efficiency}}$, s | | | |
| | | BH | | BH-S | |
Location in Figure 5.1	Function	Perturbation	Local Minimisation	Skipping Perturbation	Local Minimisation
First quadrant	Egg-holder	0.72	6.76	0.73	0.19
(Section 5.3.3)	Modified Rosenbrock	0.46	13.53	0.22	0.10
Third quadrant	Mishra-03	0.02	1.75	1.74	3.21
(Section 5.3.4)	Whitley	0.01	0.68	4.92	1.58
Special cases	Rosenbrock	0.01	0.13	0.05	0.01
(Section 5.3.5)	Styblinksi	0.02	0.06	0.06	0.01

Further exploratory analysis is provided in Table 5.1, which indicates average CPU time spent on the perturbation versus the local minimisation steps for each algorithm. To facilitate comparison between the two algorithms, in each case the total time spent is normalised by the algorithm's efficiency, as defined in Section 5.3.1. This demonstrates that the BH algorithm spends most computing time on the local minimisation step, with relatively little devoted to the perturbation step. While the ratio between processor time spent on local minimisation and perturbation is more problem-dependent for BH-S, the balance appears to be shifted in favour of perturbation.

The BH-S perturbation step is more expensive by construction, since it requires between 1 and K evaluations of the energy function f (depending on the sublevel set of the current state), whereas each BH perturbation requires just one evaluation of f. However in Table 5.1, for the Damavandi, Schwefel, Modified Rosenbrock and Egg-holder functions for which BH-S works well (cf. Figure 5.1), after normalisation the BH-S algorithm spent approximately the same or less CPU time than BH on perturbation, in addition to spending less time on local minimisation. Thus for these landscapes which favour BH-S, perturbation steps were not only less frequent for BH-S (again, after normalisation by efficiency) than BH, but the monotonic BH-S perturbations also reduced the total computational burden arising from the local

minimisation step.

Conversely it was noted above for landscapes in the third quadrant of Figure 5.1, if the energy of the state X_n is close to the global minimum value $f(x^*)$, then the corresponding sublevel set C_n has almost zero volume. This represents the worst case for the BH-S perturbation: if the states $Z_1, \ldots, Z_k$ all lie outside the sublevel set, then the perturbation requires the maximum number k of evaluations of the energy function. However, the perturbed state Y_n is rejected and $X_{n+1} = X_n$, so the optimisation procedure does not advance. Indeed for the Mishra-03 and Whitley functions in Table 5.1, the efficiency normalised CPU time invested in perturbations is two orders of magnitude greater for BH-S than for BH. For these landscapes, the efficiency normalised computational burden from local minimisation is also observed to be greater for BH-S than for BH, although the reasons for this are less clear.

Guided by the exploratory analysis of Figure 5.2, in Sections 5.3.3–5.3.5 we study the performance of both algorithms on specific energy landscapes in greater detail.

5.3.3 Landscapes favouring the BH-S algorithm

(a) Modified Rosenbrock function

(b) Egg-Holder function

(c) Sublevel set $C = \{x \in \mathbb{R}^2 \ : \ f(x) < 100\}$ of the Modified Rosenbrock function

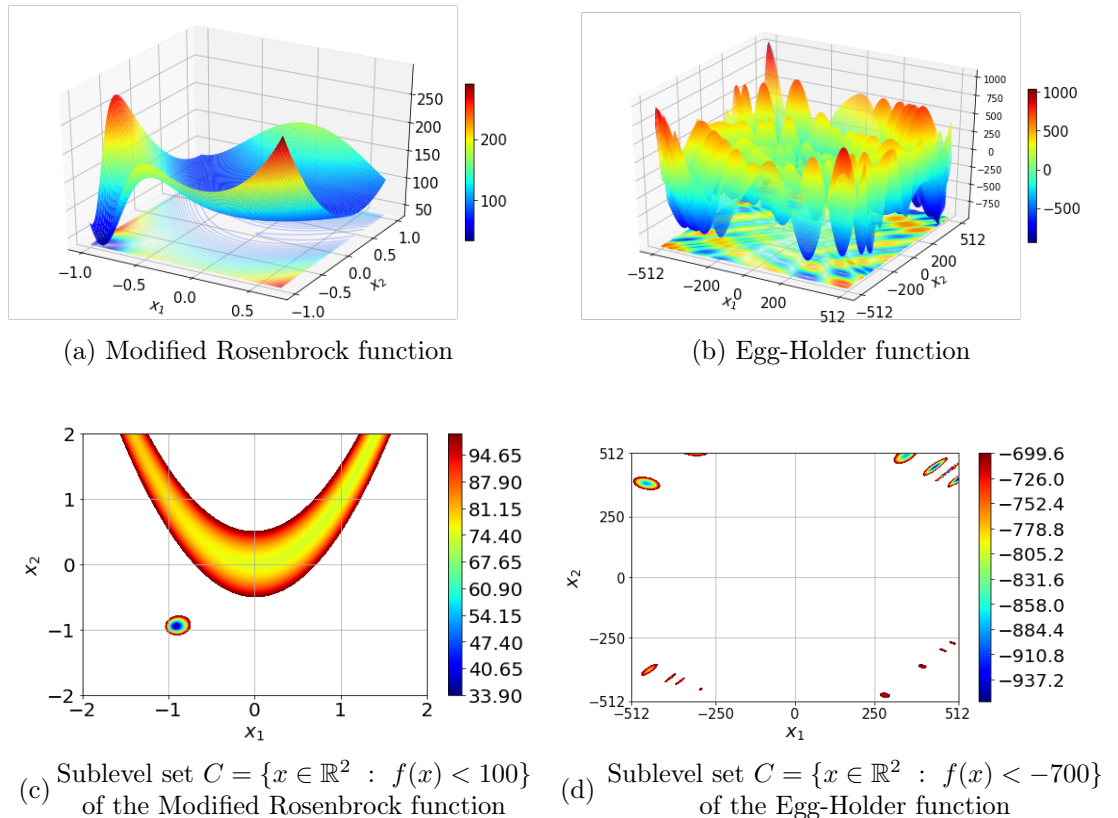(d) Sublevel set $C = \{x \in \mathbb{R}^2 \ : \ f(x) < -700\}$ of the Egg-Holder function

Figure 5.2: Examples of energy landscapes from the first quadrant of Figure 5.1

Figure 5.2 plots two landscapes from the first quadrant of Figure 5.1–that is, landscapes which favour the BH-S algorithm over BH. For each landscape, the sublevel set of a level above the global minimum $f(x^*)$ is also plotted. The Modified

Rosenbrock energy function is given by

$$f(x) = 74 + 100(x_2 - x_1^2)^2 + (1 - x_1)^2 - 400\exp\left(-\frac{(x_1 + 1)^2 + (x_2 + 1)^2}{0.1}\right),$$

and we take the domain $D = [-2, \ 2]^2$, with global minimum $x^* = (-0.95, \ -0.95)$ [32].

The Egg-Holder energy function is

$$f(x) = -(x_2 + 47)\sin\left(\sqrt{\left|x_2 + \frac{x_1}{2} + 47\right|}\right) - x_1 \sin\left(\sqrt{|x_1 - (x_2 + 47)|}\right),$$

and we take the domain $D = [-512, 512]^2$, with global minimum at $x^* = (512, 404.2319)$ [49].

Observing Figure 5.2a, the modified Rosenbrock function has two basins: a larger basin with a U-shaped valley and a smaller, well-shaped basin. To transition from the minimum of the valley to the minimum of the well, the BH algorithm would require a relatively large perturbation step landing directly in the well, otherwise the local optimisation procedure would take it back to the minimum of the valley. Even for an optimal choice of σ, which would require *a priori* knowledge about the landscape, such perturbations would be unlikely.

In contrast, if the initial point X_0 lies at the minimum of the valley, the BH-S algorithm aims to skip across the domain and enter its sublevel set C_0 as defined in (5.2). From Figure 5.2c, this will correspond to entering an approximately circular basin near the point $(-1, -1)$ in the domain. By Algorithm 7, the skipping perturbation has the potential to enter that basin provided that the straight line issuing from X_0 in the initial direction Φ in Algorithm 7 intersects it. In particular, this ability is robust to the choice of standard deviation σ provided that the halting index K is chosen appropriately (see the discussion on tuning in Subsection 5.3.7 below).

Table 5.2: Effectiveness ρ and efficiency ϵ of the BH and BH-S algorithms, with and without the application of periodic boundary conditions, for the test functions discussed in Sections 5.3.3–5.3.5.

Location in Figure 5.1	Function	Boundary Condition	ρ_c	ϵ_c	ρ_s	ϵ_s
First quadrant (Section 5.3.3)	Egg-Holder	Periodic	1.8%	77	29.4%	280
		Non-periodic	1.4%	58	81.7%	815
	Modified Rosenbrock	Periodic	18.8%	783	100%	1249
		Non-periodic	17.2%	597	96.7%	321
Third quadrant (Section 5.3.4)	Mishra-03	Periodic	77.9%	811	6.2%	78
		Non-periodic	49.6%	474	68.4%	576
	Whitley	Periodic	90.1%	137	13.9%	33
		Non-periodic	86.8%	138	13.4%	34
Special cases (Section 5.3.5)	Rosenbrock	Periodic	100%	942	100%	1376
		Non-periodic	100%	1006	100%	1524
	Styblinski	Periodic	99.7%	366	99.9%	932
		Non-periodic	99.4%	526	92.3%	784

Figure 5.2b illustrates that the Egg-Holder function has multiple basins, many of which have near-global minima. Figure 5.2d shows that the deepest basins lie in four groups, one group per corner of the domain. Within each group, the basins are close in the Euclidean distance and so perturbations are likely to enter different basins within that group. Also, the basins in each group have similar depths (that is, similar local minimum energies), making the acceptance ratio in Algorithm 6 high for such within-group perturbations. As a result the BH algorithm is likely to walk regularly between within-group local minima. Also from Figure 5.2b, the Egg-Holder function has shallower basins distributed throughout its domain. As discussed in Subsection 5.3.2 these provide an indirect, although potentially computationally expensive, route for BH to cross between the four groups of Figure 5.2d.

However between groups the Euclidean distance is large, creating the same challenge for BH as with the modified Rosenbrock function: even for optimally chosen σ, which would require *a priori* knowledge of the landscape, transitions between groups are relatively rare.

In contrast, the BH-S algorithm is capable of moving between the four groups in Figure 5.2d provided the initial direction Φ of its skipping perturbation intersects a different group. The likelihood of such an intersection is increased by both the length of the skipping chain and the use of periodic boundary conditions in the BH-S algorithm, and is again robust with respect to the choice of standard deviation σ.

Regarding the application of periodic boundary conditions to the domain D, we have argued that they are natural for BH-S, since otherwise long skipping chains would tend to exit the domain D. In contrast, they are not implemented for the BH algorithm in the results of Figure 5.1 and Table 8.4. One may therefore ask whether it is their use, rather than the skipping perturbation of BH-S, which yields any observed improvement. To explore this, Table 5.2 illustrates the effect of imposing periodic boundary conditions on the performance of both the BH and BH-S

algorithms. Interestingly the performance of BH-S on the Egg-Holder landscape is improved without their use. This tendency appears to be driven by the proximity of its global minimiser x^* to the boundary. In general, it is clear from Table 5.2 that for both algorithms their benefit or disbenefit is problem-dependent and the skipping perturbation explains a distinct and material part of the observed improvements relative to BH.

It can be observed from Table 8.4 in the Appendix that the expected mean jump distances v_s and v_c (defined in Subsection 5.3.1) typically satisfy $v_s \gg v_c$ for landscapes in the first quadrant of Figure 5.1. This confirms quantitatively the success of BH-S in hopping between distant basins. The cost of this feature is that the BH-S skipping perturbation is more computationally intensive than the random walk perturbation of BH.

Without skipping (that is, using the halting index $K = 1$ in Algorithm 7), BH-S would reduce to the monotonic basin hopping method of [61] and the *initial* perturbation W of Algorithm 7 would simply be either accepted or rejected. One may therefore also ask whether this increase in the expected mean jump distance is induced by the skipping mechanism of BH-S, or simply by its monotonicity. To address this, recall that Algorithm 7 first perturbs the current state X_n to give an initial perturbation $Z_1 := W$. Then if $f(W) > f(X_n)$, the initial perturbation is modified to Z_2, and so on, until either a state Z_k is generated with $f(Z_k) \leq f(X_n)$ or skipping is halted. If such a Z_k is found, then it may be accepted by setting $X_{n+1} = Z_k$ or rejected. The Appendix records the proportion of accepted BH-S perturbations $X_{n+1} = Z_k$ for which $k > 1$. Indeed, for many landscapes in the first quadrant of Figure 5.1 this proportion is 100%. That is, for such landscapes, each accepted perturbation X_{n+1} required the skipping mechanism since none of the initial perturbations had lower energy than the current state X_n.

5.3.4 Landscapes favouring the BH algorithm

Figure 5.3 plots two landscapes from the third quadrant of Figure 5.1, on which the BH algorithm outperforms BH-S, each with two sublevel sets above the global minimum $f(x^*)$. The Mishra-03 function is
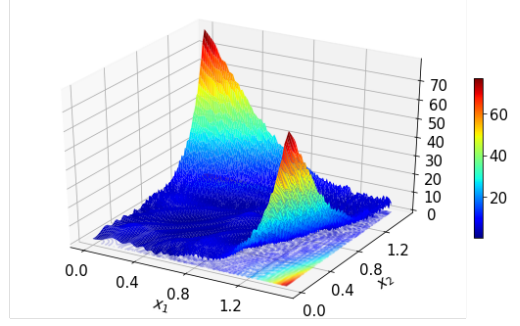
$$f(x) := \sqrt{|\cos\left(\sqrt{|x_1^2 + x_2^2|}\right)|} + 0.01(x_1 + x_2),$$

and the domain $D = [-10, 10]^2$ gives $x^* = (-8.466, -10)$ [49]. The Whitley function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

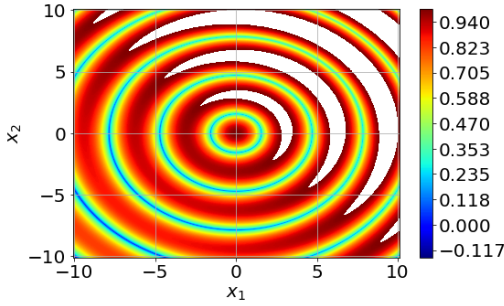$$f(x) := \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\left(100(x_i^2 - x_j)^2 + (1 - x_j)^2\right)^2}{4000} - \cos\left(100(x_i^2 - x_j)^2 + (1 - x_j)^2 + 1\right) \right),$$
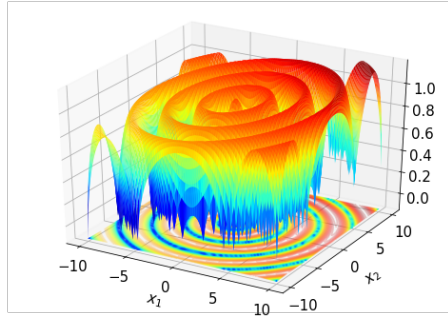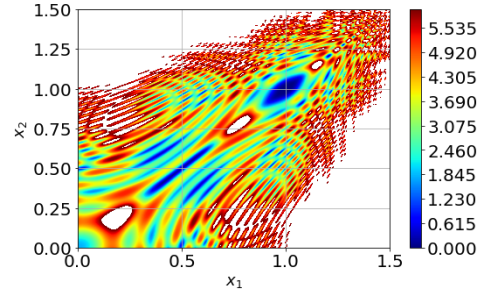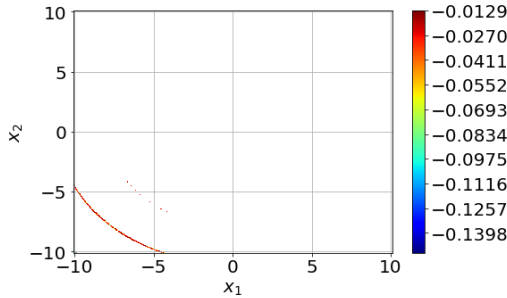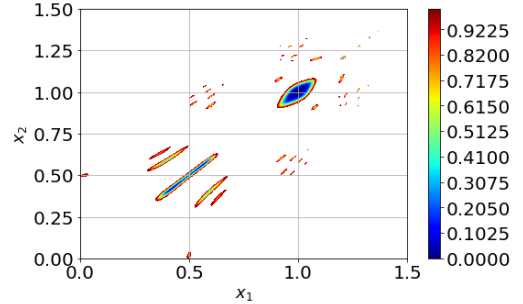
(a) Mishra-03 Function

(b) Whitley Function

(c) Sublevel set $C = \{x : f(x) < 1\}$ of the Mishra-03 function

(d) Sublevel set $C = \{x : f(x) < 5\}$ of the Whitley function

(e) Sublevel set $C = \{x : f(x) < -0.01\}$ of the Mishra-03 function

(f) Sublevel set $C = \{x : f(x) < 0.1\}$ of the Whitley function

Figure 5.3: Examples of energy landscapes favouring the BH algorithm

has global minimum $x^* = (1, 1)$ on the domain $[0, 1.5]^2$ [49].

From Figure 5.3a, the Mishra-03 function is highly irregular and has many basins which appear almost point-like. Figure 5.3e confirms that the situation outlined in Section 5.3.2 applies to this landscape. That is, for states X_n with energy close to the global minimum value $f(x^*)$, the corresponding sublevel set C_n has almost zero volume and the states $Z_1, Z_2, \ldots$, of Algorithm 7 are unlikely to fall in C_n.

The deepest basins of Mishra-03 form groups arranged in concentric circular arcs. Since the Euclidean distances both within and between these groups are relatively small, the BH algorithm is able to move frequently both within and between groups without requiring precise tuning of the standard deviation parameter σ. In particular, it outperforms BH-S on this landscape.

Similarly, by observing Figure 5.3d, the deepest basins of the Whitley function can be seen either as forming one group, or as a small number of groups close to each other. Thus, as for Mishra-03, the BH algorithm is able to move frequently between them while nevertheless being robust to the choice of the standard deviation parameter σ. As with Mishra-03, however, from Figure 5.3f the sublevel sets C_n corresponding to near-global minimum states X_n have low volume. Thus it is more challenging for BH-S to transition between the deepest basins, and BH outperforms BH-S on this landscape.

These limitations of the BH-S routine can be mitigated by alternating between a monotonic and non-monotonic perturbation step. In Subsection 5.3.8 we provide a discussion on how this alternating perturbation can be implemented.

5.3.5 Special cases

It was noted in Section 5.3.2 that for several landscapes lying near the vertical axis, both BH and BH-S algorithms have effectiveness close to 100%. For these surfaces BH-S typically has greater efficiency simply because of its monotonicity, since no further computational effort is expended on local optimisation once the global optimum is reached. The Holder Table and Carrom Table landscapes have multiple distant 'legs', each leg being the basin of a global minimum point. In this case, the ability of BH-S to skip between distant basins is not reflected in either its efficiency or its effectiveness, although it would clearly be beneficial if the goal was to identify the number of global minima in the landscape.

5.3.6 Scaling with dimension

In this section we aim to illustrate the performance of the BH-S algorithm as the dimension of the optimisation problem increases. For this we focus on Schwefel-07, a landscape with 'distant basins' which is also defined for higher dimensions. It is given by the function $f_d : \mathbb{R}^d \to \mathbb{R}$, where

$$f_d(x) = 418.9829 \times d - \sum_{i=1}^{d} x_i \sin(\sqrt{|x_i|}),$$

and has global minimum $(421.0)^d$ on the domain $D_d = [-500, 500]^d$ [49].

Figure 5.4a illustrates that the effectiveness and also the efficiency of both algorithms decrease approximately linearly with increased dimension. Recall that relative to BH, the strength of BH-S lies in its ability to transition directly between distant basins. With reference to Algorithm 7, in order to transition directly to the global minimum basin, it is necessary for the line from the current state in the random direction Φ to intersect that basin. As Φ is drawn from a space of dimension $d - 1$, heuristically this becomes less likely as d increases.

(a) Percent of samples in desired basin: ρ (b) Total trials in desired basin: ϵ
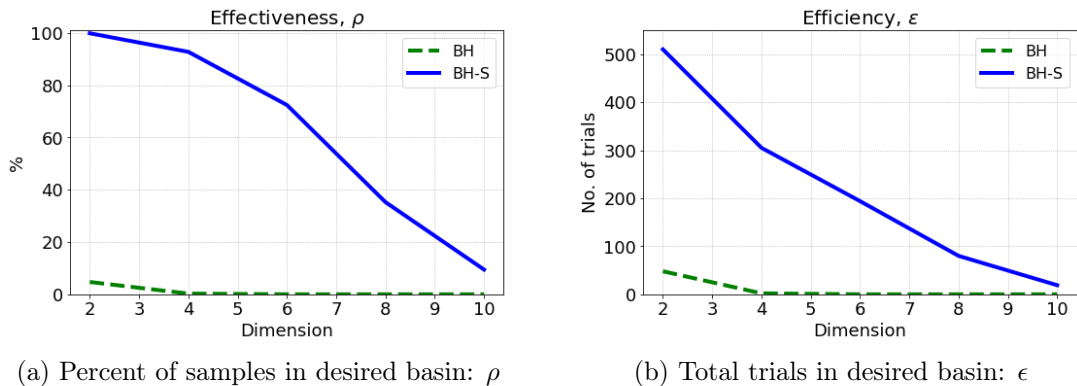
Figure 5.4: Comparison of BH and BH-S performance when applied to the Schwefel-07 function while varying the dimension d of the domain D. We set $\sigma = 20$ for both algorithms and the BH-S has halting index $K = 50$. These parameters were close to optimal for both algorithms. Each simulation used a CPU time budget of 300s.

In contrast, the BH algorithm should rely to a greater extent on indirect transitions from its current state to the global minimum. By statistical independence, the probability of a particular indirect transition is the product of the probabilities of its constituent steps. Since the probability of each step decays with dimension as discussed above for BH-S, this suggests that the performance of BH will degrade more rapidly with dimension than BH-S.

This is observed/illustrated in Figure 5.4a, where BH fails to locate x^* within the 300 second budget for any dimension $d \geq 4$, while BH-S continues to locate x^* (albeit with decreasing effectiveness and efficiency) until dimension $d = 11$. Indeed, the effectiveness of BH-S for this landscape is above 50% for dimensions $d \leq 7$.

5.3.7 Adapting parameter values to the landscape

Both BH and BH-S have the parameter σ, the standard deviation of the centred Gaussian density q used to generate the initial perturbation. As noted above, the initial perturbation is analogous to a Metropolis-Hastings (MH) proposal in MCMC. The MH literature highlights the importance of tuning such proposals, guided either by theory or by careful experimentation [31, 71]. Following this analogy, in this section we explore the choice of σ and also of the BH-S halting index K. To facilitate this discussion we restrict attention to the two-dimensional Egg-Holder function.

Figure 5.5 plots the effectiveness and efficiency of both BH and BH-S as σ varies between 0 and 300 (recall that the domain $D = [-512, 512]^2$; also, we set $K = 25$ for BH-S). Clearly, for both algorithms σ should not be very small (≤ 10). In that case the random walk step W is likely to land in the same basin as the current point X_n, so that the local optimisation step maps the perturbation back to X_n and the algorithms do not advance.

We note first from Figure 5.5 that both the effectiveness and efficiency of the

BH algorithm increase approximately linearly within this range as σ increases. As discussed in Section 5.3.3, this reflects the fact that as σ increases, direct transitions between the four groups of deepest basins become more likely. In contrast, and again confirming the discussion in Section 5.3.3, both the efficiency and effectiveness of BH-S appear to be rather robust to the choice of σ.
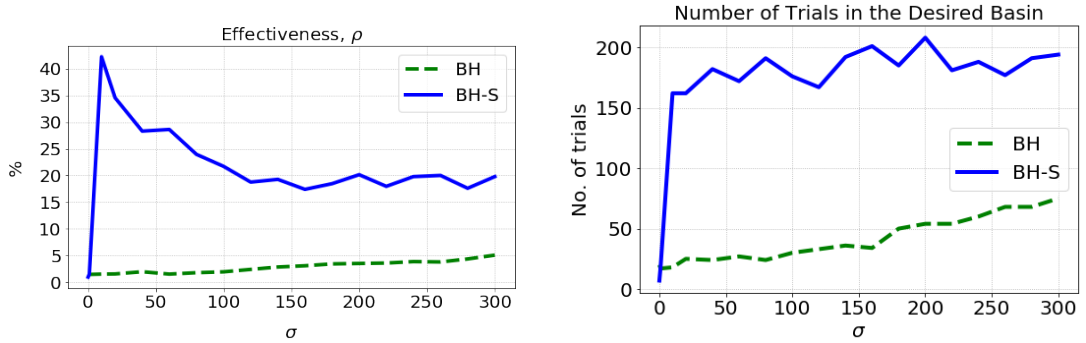
(a) Percentage of trials which successfully reported the correct global minimum

(b) Total number of trials which successfully reported the correct global minimum.
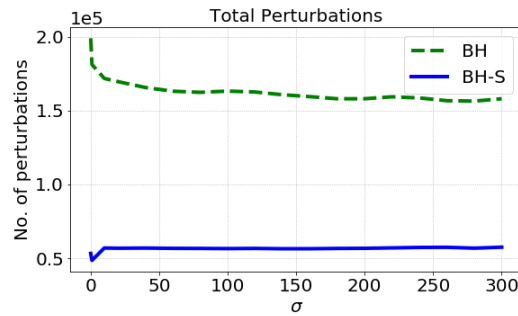
(c) Total perturbation steps conducted during the 300s time budget.

Figure 5.5: Comparison of individually tuned BH-S and BH performances on the Egg-holder function. Set-up: CPU time budget of 300 seconds; stopping criteria: 50 perturbations; the halting index for skipping perturbation is set to $K = 25$ for all simulations.
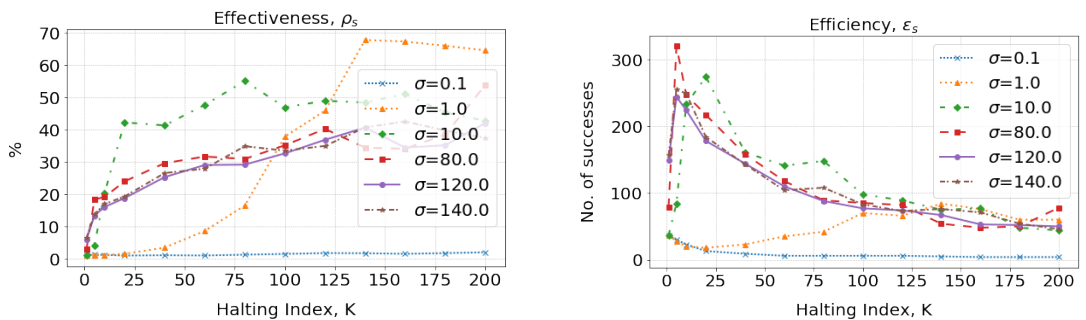
(a) Percent of samples in desired basin: ρ_s

(b) Total trials in desired basin: ϵ_s

Figure 5.6: Effectiveness and efficiency results for the BH-S applied to the Egg-holder function for various combinations of K and σ. A CPU time budget of 300s was applied to all simulations.

Figure 5.6 illustrates the impact on effectiveness and efficiency of the choice of halting index K. From Algorithm 7, the maximum linear distance covered by the skipping procedure is $\sum_{k=1}^{K} R_k$, where each R_k is distributed as the radial part of a centred Gaussian with standard deviation σ. This suggests that K should not be too small, and the plot of efficiency in Figure 5.6b indicates that K should be at least 5 in our example (by default we take $K = 25$).

It is seen that provided $(K \leq 5)$, increasing K tends to increase effectiveness while decreasing efficiency. This reflects the fact that larger K allows the skipping procedure to travel further, thus increasing the likelihood of a direct transition to the global minimum basin, after which the BH-S algorithm would stop due to its monotonicity. In this way, greater K increases effectiveness. On the other hand, greater K increases the length of unsuccessful skipping trajectories. That is, each time the perturbed state Y_n of Algorithm 7 is not accepted (after the local minimisation step of Algorithm 6), the landscape is evaluated up to K times without advancing the optimisation. This implies that increased K also typically leads to decreased efficiency.

The adaptation considerations discussed above for the BH-S algorithm can be summed up as follows. It should first be checked that σ is large enough that the initial perturbation regularly falls outside the basin of the current state X_n. Having selected σ, K should then be taken large enough that the skipping procedure regularly enters the sublevel set C_n. A practical suggestion here is to choose K so that $K\sigma$ exceeds the diameter of the domain D.

(a) Percent of samples in desired basin: ρ (b) Total trials in desired basin: ϵ
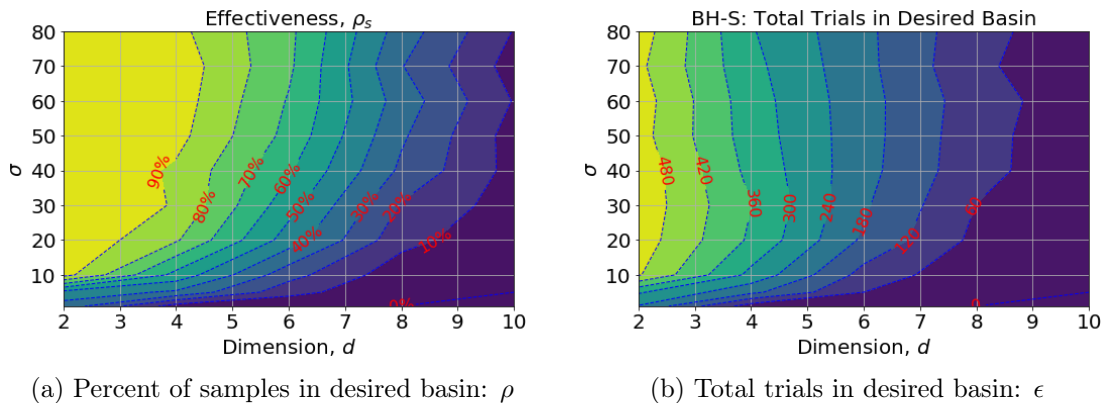
Figure 5.7: Performance of the BH-S algorithm on the Schwefel-07 function for different combinations of domain dimension d and perturbation variance σ. Note: the halting index was set to $K = 50$ with a CPU time budget of 300s for all simulations.

Figure 5.7 confirms these guidelines in higher dimensions, by plotting the BH-S effectiveness and efficiency in dimensions up to 10 as σ varies with the fixed choice $K = 50$. It confirms that these performance metrics are relatively robust to the value of σ, provided that σ is sufficiently large.

5.3.8 Alternating BH-S and BH

In this section we explore a hybrid approach which is intended to overcome the challenges identified in Section 5.3.4 for the monotonic BH-S algorithm by regularly including non-monotonic BH steps. Figure 5.8 plots the effectiveness and efficiency metrics for this hybrid algorithm on various landscapes, as the ratio between BH-S and BH steps varies.
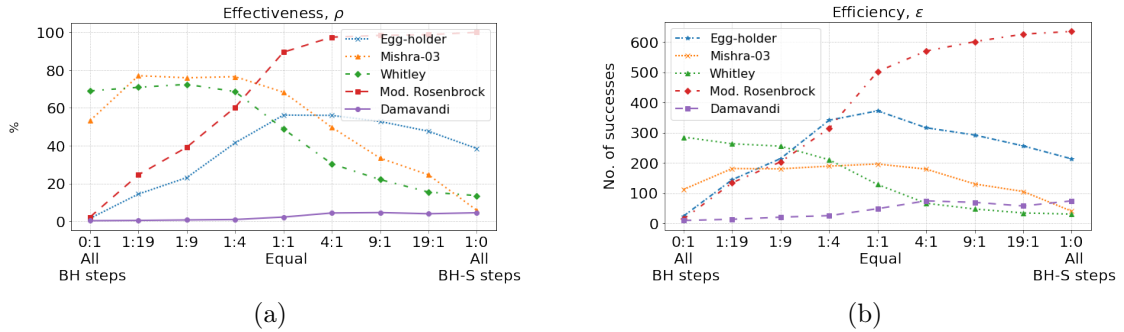

Figure 5.8: Performance of the hybrid algorithm with varying proportions of BH to BH-S steps.

It can be seen that for the Mishra-03 and Whitley functions of Section 5.3.4, this hybrid improves both effectiveness and efficiency compared to BH-S. Indeed, the performance of a 1:1 ratio of BH and BH-S steps is comparable to that of BH for these landscapes. Further, on the landscapes of Section 5.3.3, this 1:1 ratio achieves performance superior to that of BH and somewhat comparable to that of BH-S. Thus if little is known about the problem's energy landscape *a priori*, these results indicate that the 1:1 hybrid is to be preferred.

5.4 Discussion and future work

Basin hopping with skipping (BH-S) is a global optimisation algorithm inspired by both the basin hopping algorithm and the skipping sampler, an MCMC algorithm. As such, the MCMC literature also suggests potential extensions of this work. In adaptive MCMC, parameter tuning is an online procedure driven by the progress of the chain [7]. A similar idea has been proposed for BH in [33] and is part of the of the SciPy implementation of the BH method. We believe it could be interesting as future work to devise an adaptive scheme for the halting index K and the standard deviation σ, possibly reducing in this way the amount of tuning required to implement BH-S.

During our investigation of the research question, we also explored the idea of sampling several directions and skipping in all of them simultaneously. As a negative finding, we report that preliminary results indicated that the computational effort is best spent searching over a single, rather than multiple, directions. Our heuristic

explanation is that the line is the shortest route between two sets, and so is the most efficient way to cover distance. An alternative, more sophisticated approach would be to introduce multiple BH-S particles which explore the energy landscape in a coordinated way. This could for instance be inspired by selection-resampling procedures as in sequential Monte Carlo sampling [72], or by an optimisation procedure such as particle swarm optimisation [53]. Potential applications for the BH-S algorithm are to problems of metastability, where transitions between metastable energy states are rare due to separation by large energy barriers. The BH-S algorithm may improve global exploration to locate other stable energy-minima, and reduce the computational time required to transition between local minima of the energy landscape.

Chapter 6

Sequential Monte Carlo with skipping

Let π be a target density with support $C \in \mathbb{R}^m$ from which samples are desired. If the geometry of C is not known *a priori*, then sampling π when $m \gg 2$ poses a challenge for many classes of algorithms. This problem especially arises in the context of rare-event sampling when C is disjoint or in the context of global non-convex optimisation when local sub-level sets are well-separated. As potential trajectories increase with the dimension of the sample space, the chance that sampled paths intersect with any region of interest decreases, concentrating particles in local modes due to limited exploration of the sample space. Under such circumstances, even MCMC algorithms, including the skipping sampler, become increasingly ineffective [103], exhibiting low acceptance rates as it becomes increasingly challenging to propose states in $C \subset \mathbb{R}^m$.

Instead of directly sampling π, assume that samples can be drawn in a sequential manner from a series of distributions $\{\pi_n\}_{n \in \mathcal{N}}$ indexed by $\mathcal{N}$ and defined on a common measurable space E, where $\mathcal{N} = \{1, 2, ..., d\}$ or $\mathcal{N} = \mathbb{N}$. The sequential procedure means samples are to be drawn first from π_1, then π_2 and so on [21]. In practice, such sequential sampling problems arise commonly, especially in both Bayesian inference and optimisation contexts. MCMC algorithms such as the Metropolis-Hastings (MH) algorithm presented in Section 1.2 are not well adapted for such sequential sampling-type problems, since, at index n, one must wait until the Markov chain with kernel q_n achieves its associated stationary distribution π_n. In addition, it is difficult to determine when a MCMC chain has reached its stationary distribution, and chains can easily become trapped in local modes, limiting exploration of the sample space [21].

In the literature, sequential Monte Carlo SMC samplers have been employed to generate samples from such sequential-type problems [67, 21, 25]. At index n, the SMC procedure generates a large collection of $N \gg 1$ random variables $\{X_n^{(i)}\}_{i=1,...,N}$, referred to as particles, with associated weights $\{W_n^{(i)}\}_{i=1,...,N}$, whose empirical distribution converges asymptotically to the target sequence distribution

π_n at index n as $N \to \infty$ [21]. That is, for N weighted particles $\{W_n^{(i)}, x_n^{(i)}\}$, $(i = 1, ..., N; W_n^{(i)} > 0; \sum_{i=1}^{N} W_n^{(i)} = 1)$, and for any function $t : E \to \mathbb{R}$:

$$\sum_{i=1}^{N} W_n^{(i)} t(X_n^{(i)}) \to \mathbb{E}_{\pi_n}(t), \tag{6.1}$$

almost surely; where:

$$\mathbb{E}_{\pi_n}(t) = \int_E t(x)\pi_n(x)dx. \tag{6.2}$$

Particles are then evolved from density π_n to π_{n+1} by using a combination of sequential importance sampling (SIS) and resampling ideas found in [25].

SMC methods exhibit certain advantages over MCMC algorithms. Firstly, SMC algorithms do not require any burn-in period (see Section 1.2) inherent to many MCMC algorithms. Secondly, SMC samplers are considered a richer class of algorithm than MCMC as there is greater freedom in the specification of the transition kernels which can be used to move particles- in fact, in SMC algorithms, such transition kernels need not be reversible, nor even Markovian [80]. Thirdly, the performance of MCMC algorithms typically decreases as dimensionality of the state space increases. However, SMC algorithms have already been successfully applied to many high dimensional sequential problems in Bayesian inference, such as particle filter methods used in digital communication and target tracking [80].

Given the potential benefits of SMC methods for sampling from challenging distributions in higher dimensions, this chapter will present the *sequential Monte Carlo with skipping* (SMC-S) formulation of the SMC algorithm, which uses the skipping proposal from [76] (see Section 1.4) as the transition kernel to move a population of particles $\{X_n^{(i)}\}_{i=1,...,N}$ from π_n at index n to π_{n+1} at index $n+1$. That is, instead of constructing a sample incrementally by accepting/rejecting a single proposal at each iteration as is done in the MH algorithm, the SMC-S algorithm aims to sample a target density π by moving particles sequentially through a sequence of carefully designed artificial distributions $\pi_1, \ldots, \pi_{d-1}, \pi_d = \pi$ using the skipping sampler.

This chapter is structured as follows. Section 6.1 presents a sequential Monte Carlo (SMC) algorithm and a discussion on parameter choices. Subsequently, Section 6.2 presents the SMC-S algorithm for rare event sampling, while Section 6.2.1 benchmarks the performance of the SMC-S algorithm against other rare event sampling algorithms discussed in this thesis. Finally, Section 6.3 evaluates an SMC-S algorithm with non-uniform sampling of skipping trajectories, intended to improve sampling in higher dimensions.

6.1 Sequential Monte Carlo samplers

This section provides an algorithm for Sequential Monte Carlo framework, and follows the text of [21], which focused on presenting an algorithmic perspective of the SMC methodology. As mentioned, the aim of SMC methodologies is to move a set of particles $\{X_n^i\}_{i=1,...,N}$ sequentially through a series of distributions $\{\pi_n\}$ for $n = 1, \ldots, d$, where $\pi_d := \pi$ is our desired target distribution and $\{\pi_n\}_{n=1,...,d-1}$ are referred to as intermediate distributions. Algorithm 8 formalises the SMC routine.

Algorithm 8: Sequential Monte Carlo algorithm. Adapted from [21]

1 **Initialisation;** $n = 1$;

2 For $i = 1, ..., N$, sample $X_0^i \sim \nu_0(\cdot)$;

3 For $i = 1, ..., N$, evaluate the normalised weights $W_1^i = \frac{\pi_1(X_0^i)}{\nu_0(X_0^i)}$ for each particle such that: $\sum_{i=1}^N W_1^i = 1$;

4 Resampling Step:

5 Multiply/discard particles with respect to normalised weights to obtain N particles $\{X_1^i\}_{i=1,...,N}$;

6 **Iteration** n; $n \in \mathcal{N}\backslash\{1\}$

7 Sampling Step

8 For $i = 1, ..., N$, sample $Y_n^i \sim K_n(X_{n-1}^i, \cdot)$;

9 For $i = 1, ..., N$, evaluate weights and then normalised: $W_n^i, \sum_{i=1}^N W_n^{(i)} = 1$;

10 Resampling Step

11 Multiply/discard particles in $\{Y_n^i\}_{i=1,...,N}$ with respect to high/low weights $\{W_n^i\}_{i=1,...,N}$ to obtain N particles $\{X_n^i\}_{i=1,...,N}$.

The algorithm can be understood as an adaptive importance sampling technique, where the initial set of particles $\{X_0^i\}_{i=1,...,N}$ are sampled from a distribution ν_0. An importance weight is calculated for each particle using W_1^i and normalised so that $\sum_{i=1}^N W_1^i = 1$. Particles are subsequently resampled relative to their importance weights - particles with the highest importance weights are multiplied, while those with comparatively small weights are discarded.

At index n, the path of each particle is extended according to a Markov kernel $K_n(X_{n-1}^i, Y^i)$. For example, if K_n is a random walk kernel, their movement can be understood as a local exploration of the distribution at index n. For indices $n > 1$, the importance weight for each particle is given by:

$$W_n(x_{n-1}, x_n) = \frac{\pi_n(x_n)L_{n-1}(x_n, x_{n-1})}{\pi_{n-1}(x_{n-1})K_n(x_{n-1}, x_n)}, \tag{6.3}$$

where $\{L_n\}_{n=1,...,d}$ are 'backward in time' Markov kernels with density $L_{n-1}(x_n, x_{n-1})$ introduced to permit the calculation of particle weights when $n > 1$. In this formulation of the SMC, the authors of [21] note that while the choice of L_n and K_n are arbitrary (as long as (6.3) is well defined), L_n can be optimised for a given transi-

tion kernel K_n. We now have a set of weighted particles $\{X_n^i, W_n^i\}_{i=1,\dots,N}$ at index n. Particles can then be resampled according to their normalises importance weights, as was done in the initial step.

The SMC methodology formalised in Algorithm 8 is general, and allows for many potential choices for $\{\pi_n, K_n, L_n\}_{n \in \mathcal{N}}$, leading to a wide range of possible algorithms. As our aim is to generate an algorithm which combines the skipping sampler kernel with the SMC methodology to improve rare event sampling in higher dimensions, we direct our focus to SMC formulations which:

1. move a set of particles from an initial distribution ν_0 to a desired target distribution $\pi_d = \pi$ through a sequence of intermediate distributions π_n for $n = 1, \dots, d - 1$;

2. employ an MCMC transition kernel to move particles between intermediate distributions;

3. resamples particles at the end of each step, to ensure that, at index n, particles are distributed approximately according to π_n. The authors of [21] highlight various schemes which can be used to resample particles with the simplest method being to sample N new particles from the multinomial distribution of their weights $\{W_n^{(i)}\}$. This is the approach which will be taken in all subsequent SMC algorithms in this thesis.

4. has the following form for L_n:

$$L_n(x_n, x_{n-1}) = \frac{\pi_n(x_{n-1}) K_n(x_{n-1}, x_n)}{\pi_n(x_n)}, \tag{6.4}$$

According to [21], this is a common choice for $\{L_n\}$ when K_n is an MCMC kernel with unique stationary distribution π_n and $\pi_n \approx \pi_{n-1}$. This simplifies the un-normalised particle weight at index $n > 1$ to:

$$W_n(x_{n-1}, x_n) = \pi_n(x_{n-1})/\pi_{n-1}(x_{n-1}). \tag{6.5}$$

6.2 Sequential Monte Carlo with skipping (SMC-S)

In this section, we exploit the flexibility of the SMC framework to propose the *sequential Monte Carlo with skipping* (SMC-S) algorithm, which uses the skipping sampler, introduced in Chapter 1, as the transition kernel to move particles from π_{n-1} at index $n - 1$ to π_n at index n. As previously mentioned, SMC algorithms outperform MCMC algorithms when sampling in higher dimensions, while the skipping sampler is designed to improve rare event sampling. Therefore, the SMC-S

algorithm is intended to address the challenge of sampling rare events in higher dimensions.

Notation: *Given an underlying distribution ρ, we write the target density as*

$$\pi = \frac{\rho \mathbb{1}_C}{\rho(C)} \tag{6.6}$$

for connected support $C \subset \mathbb{R}^d$.

The SMC-S methodology can be understood as an SMC algorithm with the following choices for $\{\pi_n, K_n, L_n\}$: let $\{\pi_n\}_{n=1,\dots,d}$ be a sequence of distributions, each with form as given in Equation 6.6, characterised by a sequence of supports $\{C_n\}_{n=1,\dots,d}$ where $C_1 \supset C_2 \supset \dots \supset C_{d-1} \supset C_d$. The support C_1 is associated with $\pi_1 = \frac{\rho \mathbb{1}_{C_1}}{\rho(C_1)}$, which is designed to be simple to sample. The algorithm progresses sequentially through the intermediate distributions to the desired distribution $\pi_d := \pi$ with desired support $C_d = C$.

To improve the algorithm's ability to move particles from index n to $n+1$ when the support C_n is connected, the skipping sampler will be used as the transition kernel K_n (see step 10 of Algorithm 8). For each particle, the state Y_n^i is generated by a skipping sampler kernel with target density π_n. Starting at X_{n-1}^i, a skipping direction is first sampled, followed by independent jump increments to generate a proposal Z_n^i as described in skipping sampler algorithm (see Section 1.4). The state $Y_n^i = Z_n^i$ if the proposal is accepted, otherwise $Y_n^i = X_{n-1}^i$. Therefore, the skipping sampler transition kernel can be understood as a skipping sampler algorithm initiated at X_{n-1}^i with target π_n, using a single instance of the skipping proposal. This restriction in the number of proposals will be relaxed in Section 6.3.2.

Since the transition kernel of the SMC-S is a MCMC kernel with unique stationary distribution π_n, the reverse Markov kernel $L_n(x_n, x_{n-1})$ will be given by Equation (6.4) with associated simplified incremental weight function (6.5).

6.2.1 Numerical results for the SMC-S

This section presents two case studies of the SMC-S drawing samples, the first from a sample space $E \subset \mathbb{R}$ and the second from $E \subset \mathbb{R}^2$. These examples are intended to provide an intuition for the methodological and performance differences between the SMC-S and MCMC algorithms before we discuss problems in higher dimensions.

The SMC-S will be benchmarked against other sampling procedures introduced earlier in the thesis, including the skipping sampler, doubling sampler and a sequential Monte Carlo with a random-walk transition kernel (SMC-R). To compare the performance of these algorithms, the following metrics will be used:

1. The empirical mean ($\hat{\mu}$) of the samples generated by each algorithm ;

2. The Monte Carlo Standard Error (MCSE) using the *Batch Means Statistic* (see Section 1.3.1) and

3. The Mean Squared Jump Distance (MSJD) (see Section 1.3.2).

Case study in $\mathbb{R}$

The skipping sampler, doubling sampler, SMC-R and SMC-S algorithms were used to draw samples from the target density $\pi = \frac{\rho \mathbb{1}_C}{\rho(C)}$, where $\rho \sim \sum_{i=1}^{V} \gamma_i h_i(x)$ is a mixed Gaussian model for $h_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Parameter values for this model are provided in Table 6.1. The support $C := \{x \in \mathbb{R} : |x| \geq 27\}$ was designed so that C^c is convex, allowing the doubling sampler to be applied. Additionally, the parameters chosen were selected to ensure C and ρ are centred on $x = 0$, giving a known expectation of $\mathbb{E}_\pi(X) = 0$ as a benchmark for algorithmic performance. The target density π has four modes, symmetric about $x = 0$, located at $x = \mu_i$ for $i = 1, \ldots, 4$.

Table 6.1: Parameters associated with the Mixed Gaussian underlying distribution $\rho \sim \sum_{i=1}^{V} \gamma_i h_i(x)$.

Parameter	Value
V	4
γ	$\{0.15, 0.35, 0.35, 0.15\}$
μ	$\{-60, -30, 30, 60\}$
σ^2	$\{4, 12, 12, 4\}$

The following methodologies were used for each sampling algorithm:

1. Random-walk Metropolis (RW-M) sampler: a random-walk proposal density $X_{n+1} \sim \mathcal{N}(X_n, \sigma^2)$, with $\sigma^2 = 400$ chosen to generate an acceptance rate of approximately 25%. This algorithm was initialised at $X_0 = -52$, a state between the two modes of π on the negative x-axis. Such an initial state is potentially challenging for MCMC algorithms, which may become trapped in either mode on the negative x-axis, without sampling from the connected components of C located on the positive x-axis. This algorithm used $N = 900,000$ proposals, to equal the total number of proposals conducted by the SMC algorithms.

2. A skipping sampler: This was also initialised at $X_0 = -52$ with proposal density $X_{n+1} \sim \mathcal{N}(X_n, 400)$. A halting regime of $K = 6$ was chosen ensure the skipping chain can traverse the region of interest. This algorithm also conducted $N = 900,000$ proposals.

3. The doubling sampler (*D-Skip*): This was initialised at $X_0 = -52$, with $N = 900,000$ proposals. Skipping directions are sampled uniformly, with

an unconditional jump density $R_i \sim \exp\{10\}$.

4. SMC-R: $N = 20,000$ particles were sampled initially from $X_0^{(i)} \sim \mathcal{N}(-20, 400)$ for $i = \{1, \ldots, N\}$. The transition kernel K_n is a Gaussian random-walk proposal with density $Y_n^i \sim \mathcal{N}\left(X_{n-1}^{(i)}, 400\right)$ for $i \in 1, \ldots, N$. This algorithm moved particles through 45 distributions $\pi_{1:45}$ with associated support $C_{1:45} = \{x \in R : |x| > T_j\}$ for thresholds $T_j = [6, 7, ..., 50]$. The choice of 45 distributions was not required for the success of the algorithm, but was chosen to illustrate how the SMC can smoothly transition between π_{n-1} and π_n in a sequential manner. We explore how the number of intermediate distributions affect sampling performance in the next case study.

5. SMC-S: $N = 20,000$ particles were sampled initially from $X_0^{(i)} \sim \mathcal{N}(-20, 400)$ for $i = \{1, ...N\}$. The transition kernel K_n is a skipping sampler MCMC kernel with density π_n. Each skipping sampler MCMC algorithm used a single skipping chain proposal. The parameters for the skipping chain are identical to those detailed for the skipping sampler approach above. The SMC-S uses the same sequence of intermediate distributions $\{\pi_n\}_{n=1,\ldots,45}$ as the SMC-R.

Each algorithm was run for 200 independent simulations, allowing inferences about the distribution of the means, Monte Carlo standard errors and the mean squared jump distances. The results of the experiment are presented below in table 6.2.

Table 6.2: The performance metrics of each algorithm: M-RW: a random-walk Metropolis algorithm; Skip: skipping sampler; D-Skip: doubling sampler; SMC-R: a SMC algorithm with random-walk transition kernel; SMC-S: SMC algorithm with skipping transition kernel.

Metric	$M - RW$	$Skip$	$D - Skip$	$SMC - R$	$SMC - S$
Avg. $\hat{\mu}_\pi$	-59.9994	-59.6385	0.0462	4.2003	0.0557
Avg. MCSE	0.0129	0.0797	0.4087	0.6296	0.5748
Avg. MSJD	1.2945	1.2977	477.6756	103.7352	85.1286
Avg. CPU (s)	10.52	17.15	15.12	16.71	16.95

With reference to Table 6.2, following the 200 independent simulations, the mean of the sample averages was 0.0557 with average Monte Carlo standard error of 0.5748. Using the Monte Carlo standard error to form an interval analogous to

a confidence interval, we observe $\mathbb{E}_\pi(X) \in 0.0557 \pm 0.5748$, implying the sample means generated by the SMC-S algorithm are consistent with the expected value of the target distribution. While the algorithm benefits from the large number of particles distributed across the sample space, this feature alone does not improve sampling performance. For instance, the mean of the sample averages drawn with the SMC-R algorithm was 4.2003 with a MCSE of 0.6296, which implies the SMC-R generates samples which are inconsistent with π, as the $\mathbb{E}_\pi(X) \notin 4.2003 \pm 0.6286$.

The average of the sample means for the random-walk Metropolis and skipping sampler algorithms indicate they may have become trapped in the local mode centred at $x = -60$. The average MCSE of 0.0129 and 0.0797 reported respectively for these algorithms across the 200 experiments indicate that these chains consistently become stuck in this mode and are unable to sample the connected component of C on the positive x-axis. The average MSJD for both algorithms reflect this hypothesis as the average displacement provided by accepted proposals was approximately 1.3 units, while for context, the width of C^c is 52 units. Thus, proposals were dominated by local moves, and samples were restricted to the local modes nearest the point of initialization.

Of the MCMC methodologies, the results of the doubling sampler indicate superior performance over the random-walk Metropolis, skipping and the SMC-R algorithms. The average of sample means are 0.0462 with an average MCSE of 04087. As $\mathbb{E}_\pi(X) \in 0.0462 \pm 0.4087$, we can conclude the doubling sampler produces samples which are consistent with π and therefore include states from the components of C on the positive x-axis. This is derived from its ability to sample larger jump increments, allowing proposals to traverse C^c and explore the sample space. However, despite its superior performance over many of the other algorithms, this methodology is strictly limited to cases where C^c is convex, necessitating *a priori* knowledge of the geometry of C. The SMC-S is not constrained by this requirement, and can therefore be more generally applied.

Sampling a super-level set of the Egg-holder function

The Egg-holder function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x) = -(x_2 + 47) \sin\left(\sqrt{\left|x_2 + \frac{x_1}{2} + 47\right|}\right) - x_1 \sin(\sqrt{|x_1 - (x_2 + 47)|}), \quad (6.7)$$

discussed in Chapter 5, is a difficult to optimise function due to the large number of local minima. We construct a target density π on the Egg-holder function with form given by (6.6) where ρ is a mixed Gaussian distribution $\rho = \sum_{i=1}^{V} \gamma_i h_i(x)$, γ_i is a weight applied to each density function, $h_i(x) \sim \mathcal{N}(\mu_i, \Sigma_i)$ with mean μ_i and covariance matrix Σ_i. The parameters for ρ are given in Table 6.3.

Parameter	Value			
i	1	2	3	4
γ_i	0.25	0.25	0.25	0.25
μ_i	$[50, -30]$	$[-10, -21]$	$[61, 21]$	$[-100, 100]$
Σ_i	$\begin{bmatrix} 600 & 250 \\ 250 & 600 \end{bmatrix}$	$\begin{bmatrix} 600 & 420 \\ 420 & 600 \end{bmatrix}$	$\begin{bmatrix} 500 & 78 \\ 78 & 500 \end{bmatrix}$	$\begin{bmatrix} 400 & 81 \\ 81 & 400 \end{bmatrix}$

Table 6.3: *The parameters associated with the mixed Gaussian underlying distribution* $\rho = \sum_{i=1}^{V} \gamma_i h_i(x)$ *where* $h_i(x) \sim \mathcal{N}(\mu_i, \Sigma_i)$ *and* $V = 4$.

The support $C \subset \mathbb{R}^2$ is the superlevel set $C = \{x : f(x) \geq 50\}$ of the Egg-holder function, with parameters chosen such that C^c includes part of the high density region of ρ, leading to π having disjoint support in both the high and low density regions of ρ. Such sampling problems may be challenging as Monte Carlo techniques require a large number of iterations to draw random samples from the resulting low-density regions of π, while random-walk Metropolis algorithms are likely to become trapped in separated modal regions.

An SMC-S algorithm was used to draw samples from π using seven distributions π_n for $n = 1, \ldots, 7$, defined by a sequence of sublevel sets with decreasing area $C_n = \{x \in \mathbb{R}^2 : f(x) \geq 15 + 5n\}$. The SMC-S uses $N = 10,000$ particles sampled initially from $\nu \sim \mathcal{N}([10, 10], \sigma_\nu^2 \times \mathbb{I}_2)$ where $\sigma_\nu^2 = 100$ and $\mathbb{I}_2$ is the 2×2 identity matrix. The state Y_n^i is generated by the transition kernel K_n which is the skipping sampler kernel with target density π_n. The underlying proposal density of the skipping sampler is $q \sim \mathcal{N}(X_n^i, 100 \times \mathbb{I}_2)$, with a halting regime of $K = 6$.

For comparison, the following algorithms were also used to draw samples from π:

1. A skipping sampler algorithm initiated at $X_0 = [98, 31] \in C$ using $70,000$ proposals. This number of proposals was chosen so the skipping sampler conducts an equivalent number of proposals as the SMC algorithms ($7 \times 10,000$). A Gaussian proposal density is used with variance tuned to generate an acceptance rate of approximately 20% and a halting regime of $K = 6$.

2. An SMC algorithm with random-walk proposal kernel (SMC-R) with $N = 10,000$ particles also sampled initially from ν_0. The transition kernel K_n is a Gaussian random-walk kernel $K_n \sim \mathcal{N}(X_n^i, 100 \times \mathbb{I}_2)$, similar to the SMC-S' transition kernel. For consistency, the SMC-R employs the same intermediate distributions as the SMC-S algorithm.

Figure 6.1 illustrates the samples from π generated by each of the sampling methods investigated. The SMC-R algorithm was only able to draw samples from a single component of C closest to the high density region of the initial distribution of
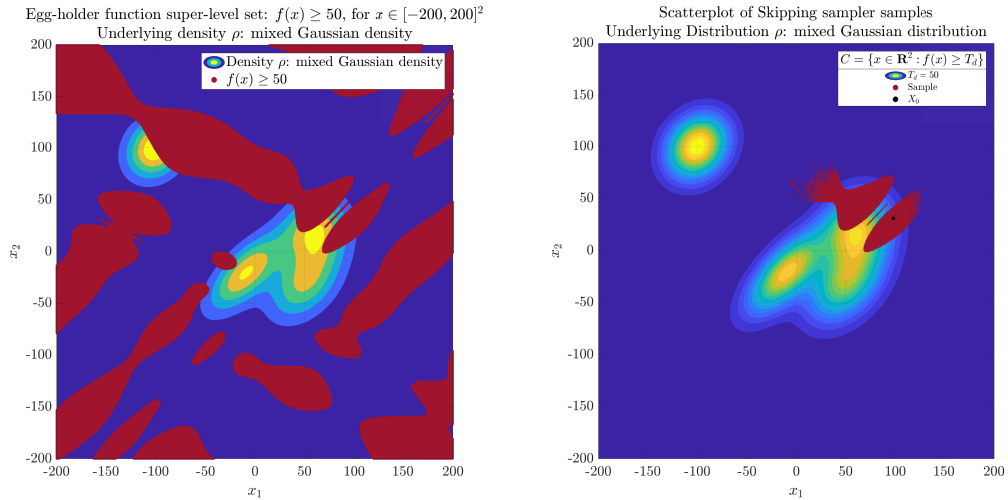
(a) Superlevel set of the Eggholder function $f(x) \geq 50$

(b) Skipping sampler
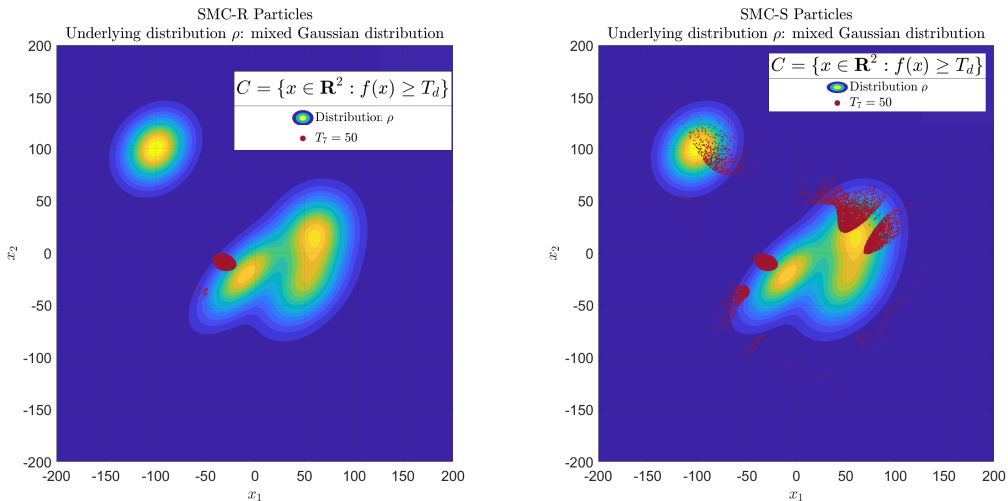
(c) SMC-R

(d) SMC-S

Figure 6.1: Illustrations of the superlevel set $f(x) \geq 50$ for $x \in [-200, 200]^2$ and samples of π generated by the skipping sampler, the SMC-R and SMC-S algorithms. The underlying mixed Gaussian density $\rho = \sum_{i=1}^{V} \gamma_i h_i(x)$ where $h_i(x) \sim \mathcal{N}(\mu_i, \Sigma_i)$ is shown as a heat map. The parameters for ρ are provided in Table 6.3.

particles ν_0. This correlation between the sample and initial distribution of particles is indicative of limited exploration of the sample space. This may be a consequence of the random-walk transition kernel favouring local moves- if transitions are unlikely to directly traverse C_n^c, particles may concentrate in a few components of C.

Referencing Figure 6.1b, the skipping sampler algorithm was able to draw samples from three connected components of C, as the skipping proposal allows the transition kernel to traverse C^c, a benefit discussed in Chapters 1 and 5. However, careful observation of Figure 6.1b reveals these components were the three proximate to the initial state of the skipping sampler algorithm X_0. Additionally, no samples were generated from the modal regions of π in the second or third quadrants of the sample space.

With reference to the computational time required, the skipping sampler required four seconds of computational time; the SMC-R took approximately five seconds to complete its evaluation; while the SMC-S required approximately seven seconds to complete. The advantage in exploration provided by the SMC-S however may compensate for the increased computational time required by the algorithm.
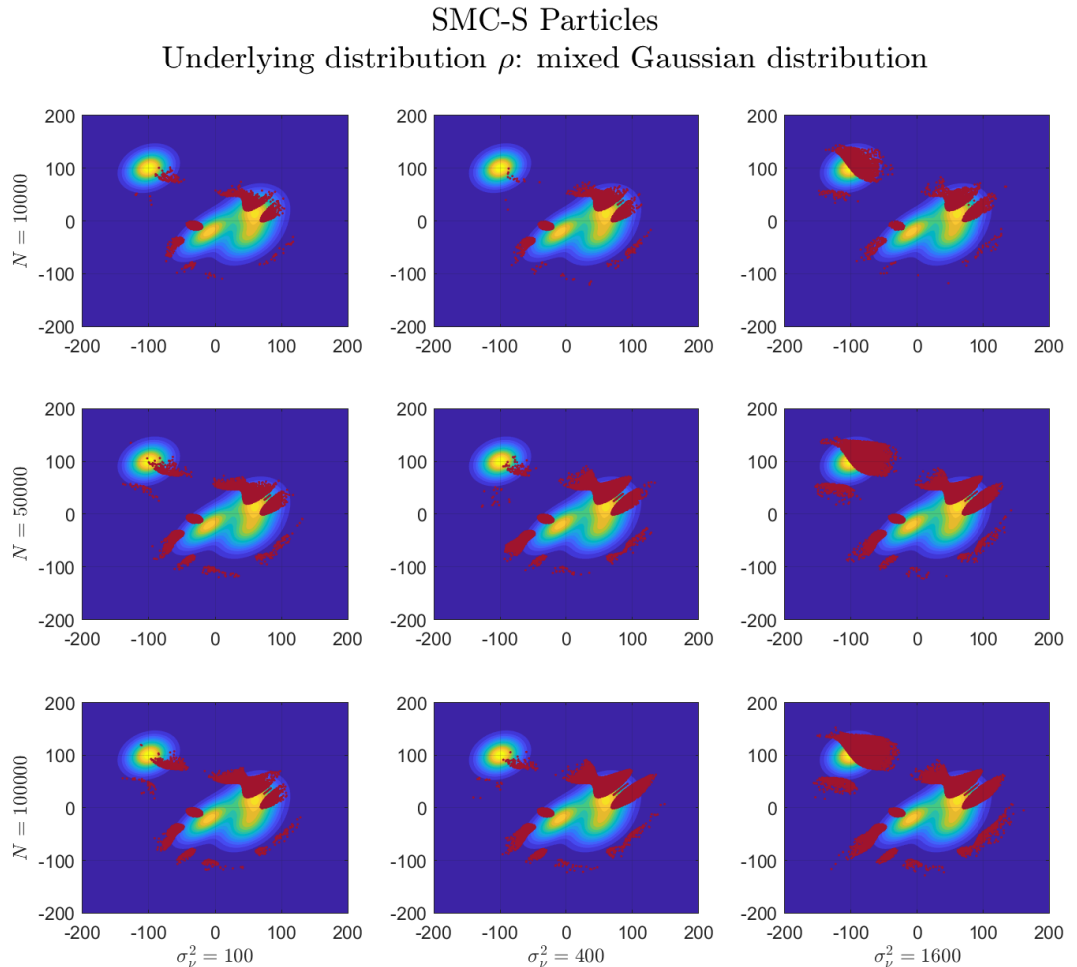
SMC-S Particles
Underlying distribution ρ: mixed Gaussian distribution

Figure 6.2: Scatter plots of samples drawn with the SMC-S from π with mixed Gaussian density as the underlying distribution with support $C = \{x \in [-200, 200]^2 : f(x) \geq 50\}$. Plots are produced for different combinations of N, the number of particles and σ_ν^2, the main diagonal component of the co-variance matrix of the initial distribution of particles.

Figure 6.2 illustrates how the dispersion of the initial distribution particles ν_0 and the number of particles N affect the performance of the SMC-S. First, note that as the main diagonal component of the covariance matrix of ν_0 (denoted by σ_ν^2) increases, so too does the exploration of the sample space. This is observed by comparing the dispersion of the final samples for the (N, σ_ν^2) parameter-pairs $(10,000, 1,600)$ and $(100,000, 100)$. In the former, more samples are drawn from the distant mode in quadrant two, producing a more representative sample of π, despite using fewer particles than the latter case. These results suggest non-local
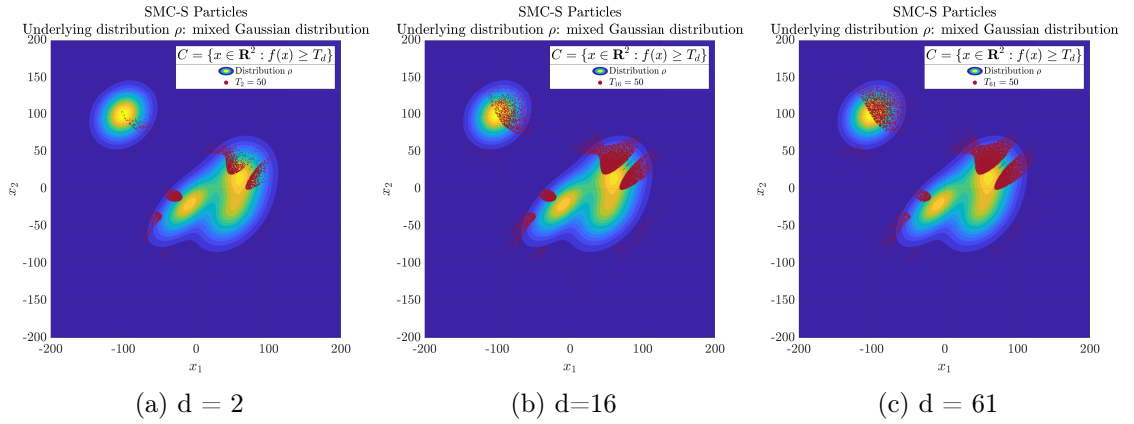
(a) d = 2   (b) d=16   (c) d = 61

Figure 6.3: Illustrations of the superlevel set $f(x) \geq 50$ for $x \in [-200, 200]^2$ and samples of π taken from the SMC-S when the number of distributions used d is 2, 16 and 61. Simulation parameters are: $N = 20,000$, $k = 20$, skipping proposal $\mathcal{N}(X_n, 625 \times \mathbb{I}_2)$ where $\mathbb{I}_2$ is the 2×2 identity matrix. The choice of k and proposal variance are designed to permit particle skipping proposal to traverse the width of the sample space.

exploration of the sample space may be better achieved by the SMC-S when the initial set of particles has a large dispersion relative to the size of the desired sample space.

Observing Figure 6.2 again, increasing the number of particles N allowed samples generated by the SMC-S to better simulate connected components of C with comparatively smaller areas. This is observed by the sampling of the narrow connected component in the first quadrant, with more samples drawn from this small strip for larger values of N. As sample space sizes and dimensions can vary, we instead guide our discussion to the particle density, calculated as the number of particles N divided by the volume of the sample space, when such spaces are bounded. Thus, for the Egg-holder function with sample space $x \in [-200, 200]^2$ and $N = 10,000$ particles, a particle density of 0.0625 particles per square unit was in effect, while for $N = 100,000$ particles, the particle density was ten times higher at $0.625/unit^2$. However, as N increased, the execution time required also increased, with times of 6s, 7s and 9s required for $N = 10,000$, $N = 50,000$ and $N = 100,000$ respectively. These results suggest that, for an SMC-S algorithm, increased particle densities may permit better sampling of small-area components, albeit at the expense of increased computational effort.

Finally, Figure 6.3 illustrates an investigation into the relationship between the number of intermediate distributions π_n used by the SMC-S and sampling outcomes. Fixing the distribution of the initial particles as $\nu \sim \mathcal{N}([10, 10], 100 \times \mathbb{I}_2)$ for all simulations, when only two distributions are used, fewer samples are drawn from the second quadrant of the sample space. However, as the number of intermediate distributions increases, more samples are drawn from this distant mode. Furthermore, Figure 6.3c shows the distribution of particles more closely follows π as the number

of intermediate distributions increases. However, as expected, increasing the number of intermediate distributions leads to a linear increase in time required to complete the simulation, with approximate times of 2s, 15s and 52s required when 2, 16 and 61 distributions were used respectively.

These results may have implications for the tuning of the SMC-S algorithm in higher dimensions. For example, to improve global exploration, the user can increase the dispersion of the initial particle set without significant additional computational effort. However, careful consideration must be given to the choice of N and the number of intermediate distributions, as these choices affect the sampling outcome but also the time taken to execute the simulation. The number of particles N should be chosen to provide an appropriate particle density, which will depend on the volume of the sample space. Finally, users should avoid choosing too few intermediate distributions when designing the SMC-S, as this was associated with poor algorithm performance.

6.3 Sequential Monte Carlo with skipping in higher dimensions

First, consider the rare-event sampling problem from Chapter 4: samples of power disturbances were desired from $\mathbb{R}^6$ using the skipping sampler, with initial proposal density $q \sim \mathcal{N}(X, \Sigma)$ for some covariance matrix $\Sigma \in \mathbb{R}^{6 \times 6}$. This implies the direction of the skipping chain is in fact sampled from a fixed distribution, independent of the previous state. While this sufficed when the dimension was $m = 6$, sampling skipping directions from a fixed distribution may be a sub-optimal strategy for moving particles between connected components of C_n in high dimensions.

Instead, this sub-section presents the *sequential Monte Carlo with non-fixed angular skipping* (SMC-AS), an amended SMC-S methodology intended to increase the proportion of particles which successfully transition between regions of interest in high dimensions. The SMC-AS algorithm employs a skipping sampler transition kernel K_n where the skipping direction for each particle is sampled from an empirical distribution, which uses the locations of all particles to bias the directions of the skipping chain. Such a skipping sampler formulation was proposed by the authors of [76] and is discussed in Section 1.4. As a reminder, if the angular distribution is non-constant, then the acceptance probability for the skipping sampler depends additionally on the ratio of the angular densities as follows: if $q_\varphi(X, \Phi)$ is the density of direction Φ from the state X, then for $\Phi = \frac{Y-X}{|Y-X|}$, the required acceptance probability is:

$$\alpha(X, Y) = \min\left(1, \frac{\pi(Y)q_\varphi(Y, -\Phi)}{\pi(X)q_\varphi(X, \Phi)}\right), \tag{6.8}$$

where Y is the proposed state from the skipping proposal.

6.3.1 Empirical angular distributions $q_\varphi(\cdot, \cdot)$

How to construct the empirical distributions $q_\varphi(\cdot, \cdot)$ for particle X_n^i will now be addressed. At the start of each index n, SMC algorithms have a set of particles $\{X_{n-1}^i\}_{i=1,\ldots,N} \in C_{n-1}$, distributed approximately according to π_{n-1} due to resampling. At index n, we sample the skipping direction for each particle X_{n-1}^i uniformly from the set of directions between X_{n-1}^i and the perturbed regions of the $N-1$ particles $\{X_{n-1}^j\}_{j \neq i, j=1,\ldots,N}$ given by:

$$\phi_n^{ij} = \left\{ \varphi_n^{ij} \in \mathbb{S}^{m-1} : \varphi_n^{ij} = \frac{X_{n-1}^j + \epsilon_n^{ij} - X_{n-1}^i}{\left| X_{n-1}^j + \epsilon_n^{ij} - X_{n-1}^i \right|} \right\} \quad \text{for} \quad \begin{array}{l} i \in \{1, \ldots, N\}, \\ j \in \{1, \ldots, N\}/\{i\} \end{array} , \quad (6.9)$$

where ϵ_n^{ij} is, for example, Gaussian noise. Once particles are sufficiently dispersed and $C_n \approx C_{n-1}$, the idea is that this procedure increases the likelihood that a particle's skipping trajectory φ_n^{ij} allows its skipping chain to intersect another component of C_n.

Since the SMC methodology given in Algorithm 8 can be specified by the choice of $\{\pi_n, K_n, L_n\}$, the SMC-AS algorithm can be understood as an SMC algorithm with: a sequence of distributions π_n with supports C_n for $n = 1, \ldots, d$, (see (6.6)); L_n is given by Equation (6.4); and the transition kernel K_n being the skipping sampler, where the angular density is the non-fixed, empirical distribution $q_\varphi^n(\cdot, \cdot)$ and the radial jump density is $R_k \sim \mathbb{U}[a, b]$ for parameters $a, b \in \mathbb{R}^+$ and $b > a$. Note that the acceptance probability (1.22) also requires the calculation of $q_\varphi^n(Y_n^i, \cdot)$ for each proposal Y_n^i.

6.3.2 Numerical Results for the SMC-AS

We evaluate the performance of an SMC-AS algorithm by drawing samples from a target density $\pi = \frac{\rho \mathbb{1}_C}{\rho(C)}$ where $\rho \sim \mathbb{U}[-400, 400]^m$ and support $C = \cup_{i=1}^4 B_i(A)$ where $B_i(A) = \{x \in \mathbb{R}^m : ||x - v_i|| \leq A\}$, disjoint hyper-spheres with identical radius A and centres v_i given by the ith row of the matrix V:

$$V \in \mathbb{R}^{4 \times m} = \begin{bmatrix} 200 & 200 & 200 & \ldots & 200 \\ -200 & -200 & -200 & \ldots & -200 \\ -200 & 200 & -200 & \ldots & 200 \\ 200 & -200 & 200 & \ldots & -200 \end{bmatrix}$$

Similar to the methodology used in Section 6.2.1, the target density and spacial distribution of the hyper-spheres were chosen to ensure that $\mathbb{E}_\pi(X) = [0, \ldots, 0] \in \mathbb{R}^m$. As the dimension of the sample space increases, so too does the volume and, by the construction of C, so does the Euclidean distance between its connected components. To compensate for this, we increase the number of particles N as m

increases. An SMC-AS algorithm is implemented to sample from π as follows: N particles are sampled from an initial distribution $\nu \sim \mathcal{N}([50, \ldots, 50] \in \mathbb{R}^m, 300^2 \times \mathbb{I}_m)$ where $\mathbb{I}_m$ is the $m \times m$ identity matrix. The distribution of ν_0 was intentionally designed to be different from ρ in terms of centrality and dispersion so to discern if the SMC-AS methodology truly transitions particles from ν_0 to π. The intermediate distributions π_n have supports $\cup_{i=1}^4 B_i(A_n)$ with hyper-sphere radii given by $A_n = 1.1 - 0.1n$ for $n = 1, \ldots, 7$. The initial radius A_1 is selected so the volume of C_1 is 80% of the volume of the domain $E := [-400, 400]^m$.

The transition kernels K_n for the SMC-AS are specified below:

1. A skipping sampler kernel with target π_n, where the skipping direction is sampled uniformly from an empirical angular distribution given in (6.9) with $\epsilon_{ij} \sim \mathcal{N}(\mathbf{0} \in \mathbb{R}^m, 0.001 \times \mathbb{I}_m)$ is intended to perturb the positions of X_j^{n-1} for $j \neq i$. The unconditional radial jump density is given by $\mathbb{U}[15, 20]$.

2. A skipping sampler kernel as described in (1) above but iterated 20 times (for distinction, this is denoted SMC-AS20).

As a benchmark, these results are compared to the following:

(i) A skipping sampler algorithm, with halting regime of $K = 100$, an underlying Gaussian proposal with covariance matrix given by $s \times \mathbb{I}_m$ where $s \in \mathbb{R}^+$ was adjusted to yield an acceptance rate of approximately 25%.

(ii) A random-walk Metropolis algorithm with Gaussian proposal function with covariance matrix given by $f \times \mathbb{I}_d$ where f was adjusted achieve an acceptance rate of approximately 25%.

(iii) An SMC-S algorithm with underlying Gaussian proposal distribution with symmetric covariance matrix tuned to generate an approximate 25% acceptance ratio and a fixed halting regime $K = 100$. The number of particles is presented in Table 6.4

Each algorithm was run for 30 independent simulations, allowing inferences about the distribution of the sample means. The results from these algorithms are detailed in Table 6.4. The table also records the *transition rate* which is the proportion of accepted moves which resulted in a transition between the four different connected components of C.

Table 6.4: Comparison of the performance of sampling algorithms as the dimension m of the sample space increases. Key: MH- Metropolis-Hastings algorithm; MH-S- skipping sampler; SMC-S: SMC algorithm with skipping transition kernel; SMC- AS: SMC algorithm with a skipping sampler transition kernel which samples the skipping direction from an empirical distribution; SMC-AS20: SMC-AS algorithm with 20 skipping proposals; AED: the average of the Euclidean distances between the sample means $\hat{\mu}$ and $E_\pi(X)$; the average distribution of samples reports the proportion of elements in the final samples in each component of C.

m	Statistic	RW-M	MH-S	SMC-S	SMC-AS	SMC- AS20
2	No. of Particles/Proposals	140,000	140,000	20,000	20,000	20,000
	AED	145.23	3.91	6.11	5.02	3.15
	Avg. distribution of samples	$[0.15, 0.14, 0.49, 0.22]$	$[0.25, 0.24, 0.24, 0.27]$	$[0.22, 0.26, 0.27, 0.26]$	$[0.25, 0.23, 0.25, 0.27]$	$[0.25, 0.24, 0.26, 0.25]$
	Avg. transition rate	0.01%	9.1%	23.1%	41.3%	66%
	Avg. time taken	116s	469s	381s	532s	2001s
5	No. of Particles	280,000	280,000	40,000	40,000	40,000
	AED	882.8	29.1	20.6	11.2	9.4
	Avg. distribution of samples	$[0.41, 0.59, 0, 0]$	$[0.25, 0.25, 0.26, 0.24]$	$[0.25, 0.24, 0.24, 0.26]$	$[0.21, 0.23, 0.28, 0.28]$	$[0.25, 0.25, 0.26, 0.24]$
	Avg. transition rate	0.02%	0.7%	26.3%	59.1%	65.4%
	Avg. time taken	201s	449s	785s	920s	9287s
10	No. of Particles	350,000	350,000	50,000	50,000	50,000
	AED	1,831.1	1,723.6	682.8	422.0	21.4
	Avg. distribution of samples	$[0.48, 0.52, 0, 0]$	$[0.51, 0.49, 0, 0]$	$[0.17, 0.21, 0.11, 0.51]$	$[0.19, 0.39, 0.20, 0.22]$	$[0.25, 0.25, 0.26, 0.24]$
	Avg. transition rate	0.1%	2.3%	27.3%	34.2%	63.7%
	Avg. time taken	291s	776s	909s	2,413s	32,148s

We evaluate the results summarised in Table 6.4 in a pairwise manner. Firstly, we briefly comment on the results of the random-walk Metropolis (RW-M) and skipping sampler (MH-S) algorithms: for all dimensions investigated, the MH-S algorithm produced samples with lower average Euclidean distances (AED) between the sample means $\hat{\mu}$ and $E_\pi(X)$, higher average transition rates and samples which are uniformly distributed between the four components of C when $m = 2$ and $m = 5$. As discussed in Section 6.2.1, the skipping sampler's ability to traverse the region C^c improves global exploration and permits higher transition rates between connected components. However, the MH-S was unable to generate samples from components three or four when $m = 10$. This suggests that increasing the displacement of each proposal via skipping may be insufficient to effectively sample connected components in higher dimensions, due to the additional challenge of generating a direction Φ which allows the skipping chain to intersect with a component of C. This limits the number of transitions and traps the chain in one or few connected components.

The potential contribution of the SMC methodology to sampling in higher dimensions can be inferred by comparing the results of the MH-S and sequential Monte Carlo with skipping (SMC-S) algorithms. While both algorithms move particles using a skipping sampler kernel where the skipping direction is sampled from a fixed distribution, the SMC-S algorithm moves multiple particles $\{X_n^i\}_{i=1,...,N}$ through a sequence of carefully constructed distributions π_n for $n = 1, \ldots, d$. With reference to Table 6.4, the SMC-S algorithm generated higher transition rates and samples with lower AED than the MH-S algorithm in all dimensions. Furthermore, skipping multiple particles via intermediate distributions was clearly advantageous when $m = 10$, as the SMC-S was able to generate samples in each component of C, unlike the MH-S. Under the SMC-S methodology, the initial intermediate distributions π_n and supports C_n are constructed to allow particles to easily transition between components. However, as n increases and C_n concentrates on C, the separation between components increases, potentially trapping particles in local modes. This phenomena may be exacerbated in higher dimensions when skipping directions are sampled from a fixed distribution, due to the aforementioned additional challenge of sampling an appropriate direction to intersect with other components of C. This is observed for the SMC-S when $m = 10$- while the SMC-S draws samples from all four components of C, it is clear most samples are localised in component four.

To overcome the challenge of directional sampling in higher dimensions, the SMC-AS algorithm instead samples each particle's skipping direction from a data-driven, empirical distribution $q_\varphi(\cdot, \cdot)$ described in Section 6.3.1. Comparing the results of the SMC-AS algorithm to those of the SMC-S algorithm, the results in Table 6.4 indicate the SMC-AS generated samples with smaller AEDs and higher transition rates than the SMC-S in each dimension investigated. Focusing on $m = 10$, the average distribution of samples generated by the SMC-AS algorithm was more

evenly distributed across the components of C than those generated by the SMC-S, indicative of improved sampling results in higher dimensions. The SMC-AS infers information about the distribution of the connected components of C_n from the relative positions of particles at index $n-1$. It then uses this information to construct the empirical distributions $q_\varphi(\cdot, \cdot)$ which biases sampled skipping directions to those which are more likely to intersect with other components of C_n. As the primary methodological difference between the SMC-S and SMC-AS algorithms is the latter's use of the empirical angular distributions, this comparison strongly suggests their use is directly associated with increased transition rates and provides an effective mechanism by which to tackle the challenge of directional sampling in higher dimensions.

Finally, to evaluate the effect of using an SMC-AS algorithm with multiple instances of the skipping sampler kernel, we compare the outcomes of the SMC-AS to those of the SMC-AS20, which uses twenty instances of the skipping sampler transition kernel. On average, the distribution of samples generated by the SMC-AS20 appears to be evenly distributed among all components of C in each dimension investigated. Additionally, when $m = 10$, the SMC-AS had an AED statistic of 422 and an average transition rate of 34.2%, while the SMC-AS20 had an AED of 21.4 and an average transition rate of 63.7%, indicating a significant improvement in sampling performance. To understand the higher transition rate of the SMC-AS20, first consider the SMC-AS; after a single instance of the skipping sampler transition kernel, if a particle's trajectory did not intersect with any component of C_n, then $Y_n^i = X_{n-1}^i$ is returned, potentially trapping that particle in the current component of C_n. However, when multiple instances of the skipping sampler kernel are employed, it is more likely that at least one instance has transitioned the particle to a new component of C_n, increasing the likelihood that the resulting Y_n^i is in a different component.

6.4   Discussion

The results of this chapter suggest that sampling the skipping direction from an empirical distribution derived from the distribution of particles may be a more effective mechanism to move particles between well-separated modes of π in high dimensions. In particular, the results suggest the distribution of particles provides useful information about the geometry of C_n at index n, which may be exploited by a transition kernel with data-driven sampling of skipping directions to increase the transition rate of particles. In the absence of *a priori* information about the geometry of C_n, this methodology may provide an advantage over algorithms which sample the proposal direction from a fixed distribution. However, implementation of the SMC-AS20 is constrained by the significant increase in computation time

required to construct distributions $q_\varphi(X_n^i, \cdot)$ and $q_\varphi(Y, \cdot)$ for each particle and each proposal, which increases quadratically with the number of particles. While this increase in execution time can be partially mitigated as the SMC-AS is easily parallelised in implementation, this additional computational effort must be considered when choosing the parameters for the algorithm, such as the number of particles N, the number of proposals for the skipping sampler transition kernel and the number of intermediate distributions.

Collectively, the family of SMC-S algorithms investigated in this chapter provides an intriguing methodology to tackle challenging rare-event sampling problems, allowing the user to benefit from the advantages of sequential sampling of intermediate distributions and non-local exploration of the sample space by the skipping sampler. In high dimension problems, using the current distribution of particles to inform the transition kernel has been demonstrated in this chapter to improve sampling outcomes when compared to previously mentioned Metropolis-Hastings and also to other SMC routines. As such, there is scope for further investigation and development of this methodology.

Chapter 7

Conclusion

This thesis explored various applications of the Metropolis-class *skipping sampler* algorithm, which was designed to draw samples from a desired set $C \subset \mathbb{R}^d$ which may be disjoint and well-separated. Such problems are commonly encountered in risk and reliability studies in power systems, where other random-walk Metropolis-Hastings algorithms are likely to become trapped in local modes, leading to unrepresentative samples. Instead, we demonstrated the skipping sampler was able to transition between connected components of C by updating the initial proposal in a linear fashion until C is entered or updates are halted.

With the relative advantages of the skipping sampler established, it was used to sample adverse rare-events of critical importance in power systems with high penetration of low-inertial, renewable energy sources. To achieve this, we developed a novel power system model based on the *third order model*, which included dynamics for network frequency, voltage, generator governor action and automatic voltage regulation (AVR). Critically, the power system model also modelled the activation of frequency-related protection systems, including the disconnection of transmission lines, generators and loads due to large frequency deviations. This novel power system model was initially used to map the activation of frequency-related protection systems to the magnitude and spatial distribution of random power disturbances within general, stylised network topologies. Power disturbances were modelled as correlated power injections, intended to represent forecast errors in the scheduled generation from renewable energy sources, induced by, for example, exogenous weather phenomena. As many weather events can be non-local in nature, they can induce correlated forecast errors in renewable power generation across large portions of the power grid.

Chapter 3 paired the novel power system model with a Monte Carlo methodology to draw samples of power disturbances which led to a network failure- the disconnection of lines, loads or generators, while recording the number of such activations - the *cascade size*. It was revealed that in highly connected networks, the distribution of cascade sizes was bi-modal- replicating similar findings in the lit-

erature and validating the novel power systems model. With respect to network connectivity, highly connected networks were shown to be more resilient, requiring disturbances with larger average magnitudes to generate cascades of similar size to those in sparsely connected networks. However, large disturbances at a single node in highly connected networks can have global effects, leading to large cascade sizes across the network. Conversely, the effect of large power disturbances at a single node in sparsely-connected were primarily local, constrained by the limited connectivity of the network.

A subsequent comprehensive analysis was conducted to investigate the effects of correlated power disturbances on a more realistic network, and to understand how battery energy storage systems (BESS) can reduce the size of cascades. This investigation summarised in Chapter 4 employed the novel power system model described in Chapter 2, along with models for a fast-response BESS and network automatic generation control (AGC) for frequency regulation. This power system model was applied to the Kundur Two-Area network, an IEEE test network based on the Australian power grid. Again, we aimed to sample power disturbances which, when applied to the power system network, led to an adverse frequency event in the network. As a BESS was expected to improve the resilience of the network against power disturbances, adverse frequency events were expected to be rare, rendering Monte Carlo sampling inefficient. To overcome this challenge, we instead employed the skipping sampler to draw samples of relevant power disturbances. The results of this study highlighted the benefits of battery systems with maximum power output of 50MW - 400MW to reduce the severity of adverse frequency events following power disturbances. However, for storage systems with higher power outputs, the inter-area transmission lines became increasingly vulnerable to excess power flows induced by the battery's response to large power disturbances. This triggered the lines' disconnection at a critical point of power instability, leading to a cascade of multiple failures in both areas. Additionally, when BESS power outputs exceed 400MW, the frequency regulatory role of battery units as governed by an AGC unit can exacerbate adverse frequency events when random power disturbances correlate with the AGC commanded state of the battery.

The results of these studies highlight the high-impact consequences of forecast errors in renewable energy generation schedules in modern power systems which are increasingly reliant on intermittent wind and solar energy sources. This motivates the urgent need for additional research into the effects of such renewables-induced power disturbances. Direct extensions of the current work can include evaluating the impact of power disturbances and the mitigating benefits of battery storage on larger, more complex networks, from which inferences about modern power systems can be easily drawn. Other extensions also include incorporating the results of recent empirical studies which revealed frequency deviations induced by renewable

power injections have non-Gaussian distributions [97]. Additional research is also needed to investigate the optimal spatial distribution BESS units in a network to mitigate the risk of frequency excursions and large contingencies in networks with high penetration of distributed renewable generation.

Beyond risk-assessment analyses, this thesis explored how the skipping sampler can be applied to the area of global stochastic optimisation, which is routinely used to aid decision-making in the energy industry, as they account for the uncertainties inherent to energy balancing, economic planning and dispatch decisions. Specifically, Chapter 5 explored *basin hopping with skipping* (BH-S), an algorithm for global non-convex optimisation which replaced the random-walk perturbation step of the BH routine with the skipping sampler.

We compared the effectiveness and efficiency of the BH-S algorithm to those of the basin-hopping (BH) algorithm when applied to locate the global minima of various energy landscapes from the global optimisation literature. This investigation revealed two, general classes of energy landscapes based on the relative performance of the BH-S and BH algorithms. First, energy landscapes where performance favoured the BH-S algorithm were those where the basin containing the global minimum was well-separated from other local minima. In such landscapes, improved global exploration promoted by the skipping perturbation allowed the algorithm to locate the global minimum basin in less time, while the BH algorithm consistently became trapped in local minima as the separation of basins made direct transitions between modes by the random-walk perturbation unlikely. Secondly, landscapes where performance favoured the BH were those with topologies where the skipping perturbation was not required or unsuited to locate the global minimum basin, including bowl-like topologies and those where the basin of the global minima was almost point-like. On such landscapes, the skipping perturbation did not contribute to locating the global minimum, leading to lower effectiveness and efficiency.

Chapter 5 concluded by discussing approaches to tuning the BH-S algorithm, as well as areas for future research, including exploring the use of multiple, interacting skipping chains for the perturbation step, inspired by a particle swarm approach. Additionally, given the BH-S improved ability to explore an energy-landscape to locate well-separated modes, a promising future application of the BH-S algorithm is to the field of molecular chemistry and the problem of metastability, where much emphasis is placed on efficiently escaping local, metastable regions of an energy function by transitioning across large energy barriers to locate the true minimum state of a system.

Finally, as intermittent renewable energy sources increase in prevalence in modern power systems, evaluating high impact, low probability events necessitate modelling a large number of stochastic variables. Therefore, rare-event analyses of such systems will be high-dimension problems, which, as noted by the literature, poses a

significant challenge for MCMC algorithms, including the skipping sampler. When C is connected and well-separated in higher dimensions, it becomes increasingly unlikely sampled skipping trajectories will intersect with any component of C. With the aim of improving sampling results in $\mathbb{R}^d$, we conducted an empirical study of the sequential Monte Carlo with skipping (SMC-S) algorithms, which pair the skipping sampler with a sequential Monte Carlo (SMC) algorithm, hence evolving a population of particles rather than just one. This methodology exploits the fact that at each stage of an SMC algorithm, the current population of particles contains information about the geometry of C.

In a similar approach taken to develop the BH-S algorithm, the SMC-S algorithms employ the skipping sampler as the transition kernel for an SMC algorithm to facilitate global exploration and draw samples from connected components of interest in $\mathbb{R}^d$. Specifically, using the current spatial distribution of particles to infer the geometry of C, the skipping sampler transition kernel samples the skipping trajectory for each particle from an empirical distribution which biases each particle's trajectory to those which are more likely to intersect with disjoint components of C. Simulations reveal SMC-S algorithms significantly improve the rate at which particles transition between regions of interest in $\mathbb{R}^d$, leading to improved sampling outcomes when compared to MCMC and other SMC routines.

However, while SMC-S algorithms exhibit improved exploration of $\mathbb{R}^d$, this is at the cost of increased computational time required for evaluation. Additional research in this area should focus on developing heuristic guidelines to improve algorithmic efficiency and reduce computational efforts. Another potential extension is to utilise a multi-step approach where the SMC-S methodology is initially used to infer the geometry of regions of interest in $\mathbb{R}^d$ from which an empirical distributions for skipping trajectories can be constructed. The second step involves applying only the skipping sampler algorithm, where the skipping trajectory is sampled from an empirical distribution. The objective is to use the SMC-S gather information about the geometry of the components of C from the distribution of particles, then use this information to construct the empirical distribution from which trajectories for a skipping sampler algorithm will be sampled. This provides the skipping sampler with information about the geometry of C, and may reduce the number of operations when compared to a full SMC-S operation.

Unlike traditional power grids which were centralised and easily controllable, 21st century power systems will be characterised by a high proportion of renewable energy sources (RES) which generate power intermittently, smart devices which regulate power consumption and distributed energy storage which balance the system in real-time. However, the design and operation of modern power systems must account for greater uncertainty and complexity, necessitating specialised algorithms to conduct risk assessments and support decision-making under uncertainty. In this thesis, we

provided a number of tools which can be employed to model high-RES networks, and demonstrated how coupling power system simulators with specialised rare-event samplers can yield insights about the risks posed by high RES penetration, and how battery storage systems can be effectively applied to mitigate these effects. We can therefore view the methodologies discussed in this thesis as another set of tools to be used in the transition to carbon neutral energy systems, enabling stakeholders to effectively design power systems which take advantage of the increasing capabilities of clean, sustainable energy sources such as wind and solar based technologies.

Bibliography

[1] ENA engineering recommendation G59: Recommendations for the connection of generating plant to the distribution systems of licensed network operators. Retrieved April 10, 2020, from `https://www.ofgem.gov.uk/system/files/docs/2018/07/riio-2_july_decision_document_final_300718.pdf`, 2015.

[2] Over Frequency Generation Shedding Schedule Table of Contents. Retrieved April 10, 2020, from `https://www.eirgridgroup.com/site-files/library/EirGrid/OPI_INN_Over_Frequency_Generation_Shedding_Schedule_Summary_Report.pdf`, 2018.

[3] Battery Storage in the United States : An Update on Market Trends. Retrieved on March 14 2020 from `https://www.eia.gov/analysis/studies/electricity/batterystorage/pdf/battery_storage.pdf`, 2020.

[4] Paul M. Anderson and A. A. Fouad. *Power system control and stability, second edition*. 2002.

[5] Göran Andersson, Peter Donalek, Richard Farmer, Nikos Hatziargyriou, Innocent Kamwa, Prabhashankar Kundur, Nelson Martins, John Paserba, Pouyan Pourbeik, Juan Sanchez-Gasca, et al. Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance. *IEEE transactions on Power Systems*, 20(4):1922–1928, 2005.

[6] I. Andricioaei, J.E. Straub, and A.F. Voter. Smart darting Monte Carlo. *The Journal of Chemical Physics*, 114(16):6994–7000, 2001.

[7] Yves F. Atchade and Jeffrey S. Rosenthal. On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[8] Mohammed Benidris, Salem Elsaiah, Samer Sulaeman, and Joydeep Mitra. Transient stability of distributed generators in the presence of energy storage devices. In *2012 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2012.

[9] Alexandros Beskos, Natesh Pillai, Gareth Roberts, Jesus-Maria Sanz-Serna, and andrew Stuart. Optimal tuning of the hybrid monte carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[10] Enric Ribera Borrell, Jannes Quer, Lorenz Richter, and Christof Schütte. Improving control based importance sampling strategies for metastable diffusions via adapted metadynamics. Technical Report 21-40, ZIB, Takustr. 7, 14195 Berlin, 2021.

[11] Freddy Bouchet, Joran Rolland, and Jeroen Wouters. Rare event sampling methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8):080402, 2019.

[12] Freddy Bouchet, Joran Rolland, and Jeroen Wouters. Rare event sampling methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8):080402, 2019.

[13] S. Brooks, A. Gelman, G. Jones, and X. Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[14] Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. Handbook of Markov Chain Monte Carlo. *Handbook of Markov Chain Monte Carlo*, 20116022:113–162, 2011.

[15] Rebekka Burkholz, Hans J Herrmann, and Frank Schweitzer. Explicit size distributions of failure cascades redefine systemic risk on finite networks. *Scientific reports*, 8(1):1–8, 2018.

[16] Francesco Cadini, Gian Luca Agliardi, and Enrico Zio. Estimation of rare event probabilities in power transmission networks subject to cascading failures. *Reliability Engineering & System Safety*, 158:9–20, 2017. Special Sections : Reliability and Safety Certification of Software-Intensive Systems.

[17] E. J. Chen and W. D. Kelton. A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality. *Simulation*, 83:683 – 694, 2007.

[18] Michael Chertkov, Feng Pan, and Mikhail G Stepanov. Predicting failures in power grids: The case of static overloads. *IEEE Transactions on Smart Grid*, 2(1):162–172, 2010.

[19] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[20] Mihai Cucuringu, Andrea Pizzoferrato, and Yves van Gennip. An mbo scheme for clustering and semi-supervised clustering of signed networks. 2019.

[21] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68:411–436, 2006.

[22] J. D. Doll, Nuria Plattner, David L. Freeman, Yufei Liu, and Paul Dupuis. Rare-event sampling: Occupation-based performance measures for parallel tempering and infinite swapping monte carlo methods. *The Journal of Chemical Physics*, 137(20):204112, 2012.

[23] Florian Dorfler and Francesco Bullo. Synchronization of power networks: Network reduction and effective resistance. *IFAC Proceedings Volumes*, 43(19):197–202, 2010.

[24] Florian Dorfler and Francesco Bullo. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(1):150–163, 2012.

[25] Arnaud Doucet, Nando Freitas, and Neil Gordon. An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice. Springer, Berlin*, 01 2001.

[26] Mehdi Ghazavi Dozein and Pierluigi Mancarella. Application of utility-connected battery energy storage system for integrated dynamic services. In *2019 IEEE Milan PowerTech*, pages 1–6. IEEE, 2019.

[27] Mehdi Ghazavi Dozein and Pierluigi Mancarella. Possible negative interactions between fast frequency response from utility-scale battery storage and interconnector protection schemes. In *2019 29th Australasian Universities Power Engineering Conference (AUPEC)*, pages 1–6. IEEE, 2019.

[28] Q. Duan and D. Kroese. Splitting for optimization. *Computers & Operations Research*, 73:119–131, 2016.

[29] Adam Dysko, Dimitrios Tzelepis, and Campbell Booth. Assessment of risks resulting from the adjustment of ROCOF based loss of mains protection settings phase II. Technical report, Institute for Energy and Environment, Glasgow, 2015.

[30] K Eswaramma and G Surya Kalyan. An Automatic Voltage Regulator (AVR) system control using a P-I-DD controller. *International Journal of Advance Engineering and Research Development*, 4(6), 2017.

[31] Dani Gamerman and Hedibert Lopes. *Markov Chain monte Carlo, Stochastic Simulatoin for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, 2nd editio edition, 2006.

[32] A. Gavana. Infinity global optimization benchmarks and *AMPGO*, 2013.

[33] Ralf Gehrke. *First-principles basin-hopping for the structure determination of atomic clusters.* PhD thesis, Freie Universität Berlin, 2009.

[34] Charles Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

[35] Dolf Gielen, Francisco Boshell, Deger Saygin, Morgan D Bazilian, Nicholas Wagner, and Ricardo Gorini. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, 24:38–50, 2019.

[36] Dolf Gielen et al. Global energy transformation: A roadmap to 2050. 2018.

[37] M. P. Goodridge, J. Moriarty, and A. Pizzoferrato. Distributions of cascade sizes in power system emergency response. In *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6, 2020.

[38] Maldon Patrice Goodridge, John Moriarty, and Andrea Pizzoferrato. Distributions of cascade sizes in power system emergency response. pages 1–6, 2020.

[39] H. Haehne, J. Schottler, M. Waechter, J. Peinke, and O. Kamps. The footprint of atmospheric turbulence in power grid frequency measurements. *EPL (Europhysics Letters)*, 121(3):30001, feb 2018.

[40] K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[41] Paul Hines, Karthikeyan Balasubramaniam, and Eduardo Cotilla Sanchez. Cascading failures in power grids. *Ieee Potentials*, 28(5):24–30, 2009.

[42] Samuel Homan and Solomon Brown. An analysis of frequency events in great britain. *Energy Reports*, 6:63–69, 2020. 4th Annual CDT Conference in Energy Storage &Â Its Applications.

[43] J. Hossain and H.R. Pota. *Robust Control for Grid Voltage Stability: High Penetration of Renewable Energy: Interfacing Conventional and Renewable Power Generation Resources.* Power Systems. Springer Singapore, 2014.

[44] Rao Huang, Jian Xiang Bi, Lei Li, and Yu Hua Wen. Basin Hopping Genetic Algorithm for Global Optimization of PtCo Clusters. *Journal of chemical information and modeling*, 2020.

[45] Modu M. Ibrahim, Jibril D. Jiya, and Idakwo O. Harrison. Modelling and simulation of automatic voltage regulator system. *International Journal of Computer Applications*, 178(1):24–28, Nov 2017.

[46] Boldea Ion. *Prime Movers, Synchronous Generators*. CRC Press, 2015.

[47] P. Jain and A. M. Agogino. Global optimization using the multistart method. *Journal of Mechanical Design*, 115(4):770–775, December 1993.

[48] Ahvand Jalali, Mehdi Ghazavi Dozein, and Pierluigi Mancarella. Frequency stability provision from battery energy storage system considering cascading failure s with applications to separation events in australia. In *2019 IEEE Milan PowerTech*, pages 1–6. IEEE, 2019.

[49] M. Jamil and X. S. Yang. A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2):150, 2013.

[50] Galin L Jones, Murali Haran, Brian S Caffo, and Ronald Neath. Fixed-width output analysis for markov chain monte carlo. *Journal of the American Statistical Association*, 101(476):1537–1547, 2006.

[51] Wenyun Ju, Kai Sun, and Rui Yao. Simulation of cascading outages using a power-flow model considering frequency. *IEEE Access*, 6:37784–37795, 2018.

[52] Nemica Kadel, Wei Sun, and Qun Zhou. On battery storage system for load pickup in power system restoration. In *2014 IEEE PES General Meeting/ Conference & Exposition*, pages 1–5. IEEE, 2014.

[53] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995.

[54] E.W. Kimbark. *Power System Stability*. Number v. 1-3 in Power System Stability. John Wiley & Sons, 1995.

[55] S Kirkpatrick, C D Gelatt Jr., and M P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[56] Prabha Kundur. Power System Stability And Control by Prabha Kundur.pdf, 1994.

[57] Prabha Kundur, Neal J Balu, and Mark G Lauby. *Power system stability and control*, volume 7. McGraw-hill New York, 1994.

[58] Prabha Kundur, John Paserba, Venkat Ajjarapu, Göran Andersson, Anjan Bose, Claudio Canizares, Nikos Hatziargyriou, David Hill, Alex Stankovic, Carson Taylor, et al. Definition and classification of power system stability. *IEEE transactions on Power Systems*, 19(2):1387–1401, 2004.

[59] Prabha S. Kundur. Power system stability. In *Power System Stability and Control, Third Edition*, pages 8–1–8–11. 2017.

[60] S. Lan, J. Streets, and B. Shahbaba. Wormhole Hamiltonian Monte Carlo. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pages 1953–1959, 2014.

[61] R. H. Leary. Global optimization on funneling landscapes. *Journal of Global Optimization*, 18(4):367–383, 2000.

[62] Dongchan Lee, Liviu Aolaritei, Thanh Long Vu, and Konstantin Turitsyn. Robustness against disturbances in power systems under frequency constraints. *IEEE Transactions on Control of Network Systems*, 6(3):971–979, 2019.

[63] Robert Lee, Michael Assante, and Tim Conway. Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016.

[64] Viktor A. Levi, Jovan M. Nahman, and Dusko P. Nedic. Security modeling for power system reliability evaluation. *IEEE Transactions on Power Systems*, 2001.

[65] Viktor A Levi, Jovan M Nahman, and Dusko P Nedic. Security modeling for power system reliability evaluation. *IEEE Transactions on Power Systems*, 16(1):29–37, 2001.

[66] D. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.

[67] J.S. Liu, J.S. Liu, and S.L. Jun. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer New York, 2001.

[68] M. Lu, W. A.W. ZainalAbidin, T. Masri, D. H.A. Lee, and S. Chen. Underfrequency load shedding (UFLS) schemes, A survey. *International Journal of Applied Engineering Research*, 11(1):456–472, 2016.

[69] R. Martí. Multi-start methods. In *Handbook of Metaheuristics*, pages 355–368. Kluwer Academic Publishers, 2003.

[70] Pascal Mercier, Rachid Cherkaoui, and Alexandre Oudalov. Optimizing a battery energy storage system for frequency control application in an isolated power system. *IEEE Transactions on Power Systems*, 24(3):1469–1477, 2009.

[71] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[72] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[73] Rodrigo Moreno, Alexandre Street, José M Arroyo, and Pierluigi Mancarella. Planning low-carbon electricity systems under uncertainty considering operational flexibility and smart grid technologies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2100):20160305, 2017.

[74] John Moriarty, Jure Vogrinc, and Alessandro Zocca. Frequency violations from random disturbances: an MCMC approach. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1598–1603. IEEE, 2018.

[75] John Moriarty, Jure Vogrinc, and Alessandro Zocca. The Skipping Sampler: A new approach to sample from complex conditional densities. *arXiv preprint arXiv:1905.09964*, 2019.

[76] John Moriarty, Jure Vogrinc, and Alessandro Zocca. A Metropolis-class sampler for targets with non-convex support. 2020.

[77] A Muir and J Lopatto. Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations, 2004.

[78] Dusko P Nedic. *Simulation Of Large System Disturbances*. Master's thesis, Electrical engineering and electronics, 2003.

[79] Tommaso Nesti, Alessandro Zocca, and Bert Zwart. Emergent failures and cascades in power grids: a statistical physics perspective. *Physical review letters*, 120(25):258301, 2018.

[80] Thi Le Thu Nguyen, FranÃ§ois Septier, Gareth W. Peters, and Yves Delignon. Efficient sequential monte-carlo samplers for bayesian inference. *IEEE Transactions on Signal Processing*, 64(5):1305–1319, 2016.

[81] A. G. Nogales. Conditional expectation of a markov kernel given another with some applications in statistical inference and disease diagnosis. *Statistics*, 54(2):239–256, 2020.

[82] Brian Olson, Irina Hashmi, Kevin Molloy, and Amarda Shehu. Basin Hopping as a General and Versatile Optimization Framework for the Characterization of Biological Macromolecules. *Advances in Artificial Intelligence*, 2012.

[83] Alexdandre Oudalov, Daniel Chartouni, Christian Ohler, and G_ Linhofer. Value analysis of battery energy storage applications in power systems. In *2006 IEEE PES Power Systems Conference and Exposition*, pages 2206–2211. IEEE, 2006.

[84] Art B Owen, Yury Maximov, Michael Chertkov, et al. Importance sampling the union of rare events with an application to power systems analysis. *Electronic Journal of Statistics*, 13(1):231–254, 2019.

[85] Jose Pablo Figueroa Zelaya. *Electric Power Systems Security Analysis in Smart GRids with Renewable Energy*. Masters thesis, Universidad de Oviedo, 2017.

[86] Martín Leandro Paleico and Jörg Behler. A flexible and adaptive grid algorithm for global optimization utilizing basin hopping Monte Carlo. *Journal of Chemical Physics*, 2020.

[87] Mathaios Panteli and Pierluigi Mancarella. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. *Electric Power Systems Research*, 127:259–270, 2015.

[88] Brendan Patch and Bert Zwart. Analyzing large frequency disruptions in power systems using large deviations theory. In *Probabilistic Methods Applied to Power Systems (PMAPS) 2020*, 2020.

[89] E. Pompe, C. Holmes, and K. Łatuszyński. A framework for adaptive MCMC targeting multimodal distributions. *The Annals of Statistics*, 48(5), October 2020.

[90] Sebastián Püschel-Løvengreen, Mehdi Ghazavi Dozein, Steven Low, and Pierluigi Mancarella. Separation event-constrained optimal power flow to enhance resilience in low-inertia power systems. *Electric Power Systems Research*, 189:106678, 2020.

[91] Elyas Rakhshani and Pedro Rodriguez. Inertia emulation in ac/dc interconnected power systems using derivative technique considering frequency measurement effects. *IEEE Transactions on Power Systems*, 32(5):3338–3351, 2016.

[92] Vaishali Rampurkar, Polgani Pentayya, Harivittal A Mangalvedekar, and Faruk Kazi. Cascading failure analysis for indian power grid. *IEEE Transactions on Smart Grid*, 7(4):1951–1960, 2016.

[93] G. O Roberts, J. Rosenthal, et al. General state space markov chains and mcmc algorithms. *Probability surveys*, 1:20–71, 2004.

[94] G.O. Roberts and A.F.M. Smith. Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms. *Stochastic Processes and their Applications*, 49(2):207–216, 1994.

[95] Gustavo G. Rondina and Juarez L.F. Da Silva. Revised basin-hopping monte carlo algorithm for structure optimization of clusters and nanoparticles. *Journal of Chemical Information and Modeling*, 2013.

[96] R. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and computing in applied probability*, 1(2):127–190, 1999.

[97] Leonardo Rydin Gorjão, Richard Jumar, Heiko Maass, Veit Hagenmeyer, G Cigdem Yalcin, Johannes Kruse, Marc Timme, Christian Beck, Dirk Witthaut, and Benjamin Schäfer. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nature communications*, 11(1):1–11, 2020.

[98] Abdelhay A. Sallam and Om P. Malik. *Power System Stability: Modelling, Analysis and Control.* 2015.

[99] Benjamin Schäfer, Dirk Witthaut, Marc Timme, and Vito Latora. Dynamically induced cascading failures in power grids. *Nature communications*, 9(1):1–13, 2018.

[100] Katrin Schmietendorf, Joachim Peinke, Rudolf Friedrich, and Oliver Kamps. Self-organized synchronization and voltage stability in networks of synchronous machines. *European Physical Journal: Special Topics*, 223(12):2577–2592, 2014.

[101] Katrin Schmietendorf, Joachim Peinke, Rudolf Friedrich, and Oliver Kamps. Self-organized synchronization and voltage stability in networks of synchronous machines. *The European Physical Journal Special Topics*, 223(12):2577–2592, 2014.

[102] M. Schumer and K. Steiglitz. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 13(3):270–276, June 1968.

[103] Babak Shahbaba, Luis Martinez Lomeli, Tian Chen, and Shiwei Lan. Deep markov chain monte carlo, 2019.

[104] C. Sminchisescu and M. Welling. Generalized Darting Monte Carlo. In Marina Meila and Xiaotong Shen, editors, *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, volume 2, pages 516–523. PMLR, 21–24 Mar 2007.

143

[105] C. Sminchisescu, M. Welling, and G. Hinton. A mode-hopping MCMC sampler. Technical report, CSRG-478, University of Toronto, 2003.

[106] S. Surjanovic and D. Bingham. Virtual library of simulation experiments: test functions and datasets. Retrieved March 9, 2021, from http://www.sfu.ca/~ssurjano/egg.html.

[107] Luke Tierney. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22(4):1701–1762, 1994.

[108] H. Tjelmeland and B.K. Hegstad. Mode jumping proposals in MCMC. *Scandinavian Journal of Statistics*, 28(1):205–223, 2001.

[109] Dimiter Tsvetkov, Lyubomir Hristov, and Ralitsa Angelova-Slavova. On the convergence of the metropolis-hastings markov chains, 2020.

[110] Andreas Ulbig, Theodor S. Borsche, and GÃ¶ran Andersson. Impact of low rotational inertia on power system stability and operation. *IFAC Proceedings Volumes*, 47(3):7290–7297, 2014. 19th IFAC World Congress.

[111] David J. Wales and Jonathan P.K. Doye. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *Journal of Physical Chemistry A*, 1997.

[112] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[113] Ruifeng Yan, Tapan Kumar Saha, Feifei Bai, Huajie Gu, et al. The anatomy of the 2016 south australia blackout: A catastrophic event in a high renewable network. *IEEE Transactions on Power Systems*, 33(5):5374–5388, 2018.

[114] Rui Yao, Shaowei Huang, Kai Sun, Feng Liu, Xuemin Zhang, and Shengwei Mei. A multi-timescale quasi-dynamic model for simulation of cascading outages. *IEEE Transactions on Power Systems*, 31(4):3189–3201, 2015.

[115] Yutian Zhou, Mathaios Panteli, Rodrigo Moreno, and Pierluigi Mancarella. System-level assessment of reliability and resilience provision from microgrids. *Applied Energy*, 230:374–392, 2018.

[116] Gert Zöller, Matthias Holschneider, and Yehuda Ben-Zion. Quasi-static and quasi-dynamic modeling of earthquake failure at intermediate scales. In *Computational Earthquake Science Part I*, pages 2103–2118. Springer, 2004.

Chapter 8

Appendix

8.1 Appendix 1: Kundur two-area network case study parameters

The following susceptance matrices describes the Kron-reduced Kundur Two Area Network (KTAS) network before and after a line-tripping event. The susceptance matrix depends on the indicator function Ω which changes from 1 to 0 when a line trip occurs. The difference occurs in line 5-6, which is the inter-connector line of the reduced network.

$$
B(\Omega) = \begin{bmatrix}
-18.9278 & 7.8461 & 0 & 0 & 12.9499 & 0 \\
7.8461 & -38.0413 & 0 & 0 & 32.5581 & 0 \\
0 & 0 & -19.0178 & 7.8461 & 0 & 12.9499 \\
0 & 0 & 7.8461 & -38.3713 & 0 & 32.5581 \\
12.9499 & 32.5581 & 0 & 0 & -52.1891 & 9.0982 \cdot \Omega \\
0 & 0 & 12.9499 & 32.5581 & 9.0982 \cdot \Omega & -50.6891
\end{bmatrix}.
$$

Table 8.1: Table of parameters used for Kundur's two-area network in simulations for Chapter 4

Sym.	Meaning	Value	Units
$\delta(0)$	Electrical Phase Angle (nodal; t $= 0$)	$\{0.4, 0.2, -0.1, -0.3, -0.1, -0.5\}$	[rad]
$\dot{\delta}(0)$	Rotor Angular Velocity (nodal; t $= 0$)	$\{0, 0, 0, 0, 0, 0\}$	$[s^{-1}]$
$\ddot{\delta}(0)$	Rate of Change of Frequency (nodal; t= 0)	$\{0, 0, 0, 0, 0, 0, \}$	$[s^{-2}]$
$v(0)$	Voltages (nodal; t= 0)	230	kV
$\rho(0)$	Initial Governor Contribution to Power	$\{0, 0, 0, 0, 0, 0\}$	$[MW]$
P^G	Initial Generator Power (nodal; t= 0)	$\{735, 735, 735, 735, 0, 0\}$	$[MW]$
P^L	Initial Loads (nodal; t= 0)	$\{100, 100, 50, 50, 967, 1, 767\}$	$[MW]$
D	Load Damping Factor	2	[%]
T_d	Transient Time constant	8	[s]
E_0	Rotor Field Voltage	20	[kV]
L	Equivalent Machine Reactance $X_d - X_d'$	$\{1.07, 1.07, 1.07, 1.07, 1.07, 1.07\}$	ohms

Sym.	Meaning	Value	Units
$\mathcal{W}$	Governor Deadband Frequency Range	[59.95, 60.05]	[Hz]
A	Governor Droop Response	2	MW/Δf
Battery parameters			
B^m	Maximum Battery Power	$B^m \in \{0, 100, 200, ..., 1000\}$	MW
B^r	Maximum Battery Power for Regulation FCAS	$\frac{1}{2}B^m$	[MW]
b^+	Initial Battery State Parameter	1	%
T^b	AGC Signal Interval	4	s
F^d	Battery Deadband Deviation	0.05	[Hz]
F^n	Emergency FCAS Frequency Deviation	0.15	[Hz]
F^m	Frequency deviation associated with Maximum Battery Power	1	[Hz]
AVR parameters			
T^s	Sensor Time Constant	0.05	[s]
K^s	Sensor Gain Constant	1	
T^a	Amplifier Time Constant	0.1	[s]
K^a	Amplifier Gain Constant	10	
T^e	Exciter Time Constant	1	[s]
K_e	Exciter Gain Constant	10	
Protection system parameters			
F^+	OFGS Threshold	62	[Hz]
G	RoCoF Trip Threshold	3	$[Hzs^{-2}]$
F^-	Progressive UFLS Thresholds	$\{59.5, 59, 58.5, 58, 57.5\}$	[Hz]
P^ϕ	Line Trip Power Flow Deviation Threshold	510	[MW]
T^ϕ	Line Trip Relay Delay	4	[s]
Generator and line data			
P^m	Maximum Generator Power	900	[MW]
M	Generator Angular Momentum	$\{0.3104, 0.3104, 0.2948, 0.2948\}$	[Ws^2]
X_d	Transient Reactance	4.05	[ohms]
X_d'	Sub-transient Reactance	2.93	[ohms]
B	line susceptances	see Matrix above	[siemens]

Table 8.2: Generator Data for Kundur's two-area network: 900 MVA, 20kV base

Generator	Rating (MVA)	X_d (pu)	X_d' (pu)	T_{d0}' (s)	H (s)
G1	900	1.8	0.3	8	6.5
G1	900	1.8	0.3	8	6.5
G1	900	1.8	0.3	8	6.175
G1	900	1.8	0.3	8	6.175

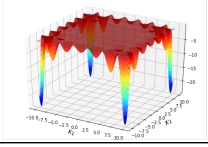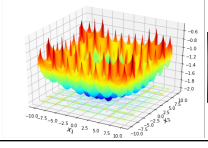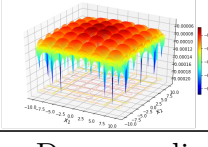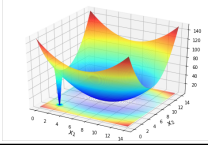Table 8.3: Transmission Network Data for Kundur's two-area network: 100 MVA, 230kV base [57]
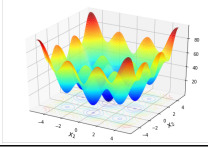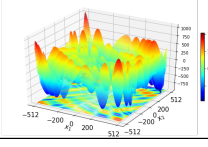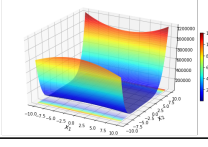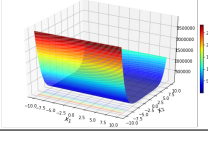
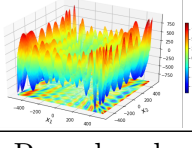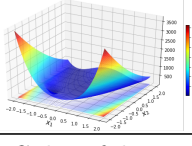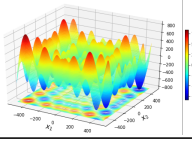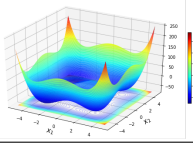| From Bus	To Bus	R (pu)	X (pu)	B (pu)
1	5	0	0.15/9	0
2	6	0	0.15/9	0
3	11	0	0.15/9	0
4	10	0	0.15/9	0
5	6	25×0.0001	25×0.001	25×0.00175
10	11	25×0.0001	25×0.001	25×0.00175
6	7	10×0.0001	10×0.001	10×0.00175
9	10	10×0.0001	10×0.001	10×0.00175
7	8	110×0.0001	110×0.001	110×0.00175
7	8	110×0.0001	110×0.001	110×0.00175
8	9	110×0.0001	110×0.001	110×0.00175
8	9	110×0.0001	110×0.001	110×0.00175
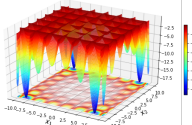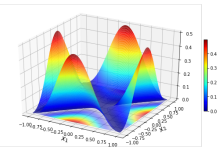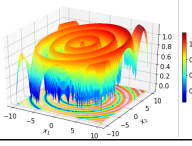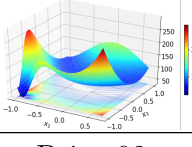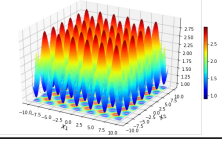
8.2 Appendix II: Performance comparison of basin-hopping and basin-hopping with skipping algorithms
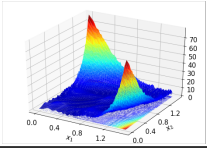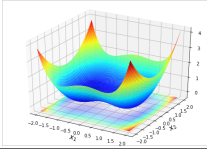
Table 8.4 records the results for all landscapes in Figure 5.1. For each landscape, both BH and BH-S were hand tuned in order to maximise their efficiency as defined in Section 5.3.1. The following notation is used:

- ρ_c and ρ_s are the effectiveness of BH and BH-S respectively;

- ϵ_c and ϵ_s are the efficiency of BH and BH-S respectively;

- υ_1 is the expected mean jump distance among random walk steps;

- υ_s is the expected mean jump distance among skipping transitions, i.e., when $k > 1$;

- $\mathbb{P}_s$ is the probability that, conditional on the BH-S perturbation being accepted, skipping had occurred ($k > 1$);

- ν_1 and ν_s are the expected mean jump distances among random walk steps ($k = 1$) and skipping steps ($k > 1$), respectively.

Table 8.4: Performance metrics for the BH and BH-S algorithms on all landscapes in Figure 5.1.

	BH				BH-S						
Function	σ	ρ_c	ϵ_c	ν_1	σ	K	ρ_s	ϵ_s	ν_1	ν_s	$\mathbb{P}_s$
Carrom Table	2	99.8	556	1.5	$\sqrt{2}$	10	100	925	2.9	9.4	86.8
Cross in Tray	2	96.5	446	1.7	$\sqrt{2}$	10	100	676	2.7	9.3	85.1
Cross Leg Table	0.4	15.5	48	0.8	$\sqrt{2}$	10	12.7	45	1.8	6.6	45.7
Damavandi	0.1	0.2	3	0.5	0.3	150	32.9	28	N/A	34.9	100
Eggcrate	1	99.7	377	1	1	10	99.7	647	1.9	7.6	94.8
Egg-holder	100	2.2	13	12.5	10	25	38.7	116	7.1	178.1	98.6
El Attar Vidyasagar Dutta EAVD	8	99.6	231	3.2	5	10	63.4	393	4.8	5.8	39.3
Freudenstein-Roth	2	82.6	176	1.6	$\sqrt{2}$	10	97.2	551	4.9	11.3	99.5

Function	BH				BH-S						
	σ	ρ_c	ϵ_c	ν_1	σ	K	ρ_s	ϵ_s	ν_1	ν_s	$\mathbb{P}_s$
Holder Table	2	99.8	453	1.4	$\sqrt{2}$	10	100	695	2.4	9.1	82.4
Keane	2	47	272	1.8	0.9	25	53.6	301	1.6	12.4	96.6
Mishra-03	2	65.9	56	1.8	$\sqrt{2}$	10	5.4	17	1.6	12.1	96.7
Modified Rosenbrock	0.4	5.3	1	0.8	0.4	25	83.8	31	N/A	7.7	100
Price 02	2	44.4	234	1.8	0.9	25	60.6	208	N/A	17.1	100
Rana	200	5.5	13	17.6	5	75	20.5	18	1.8	224.5	98.5
Rosenbrock	0.2	99.3	149	0.6	2	10	100	695	N/A	N/A	0
Schwefel-07	10	4.1	38	4	7	25	61.9	275	N/A	181.6	100
Styblinski Tang	1	99.6	537	1.2	1	10	100	840	N/A	8.7	99.8

	BH				BH-S						
Function	σ	ρ_c	ϵ_c	ν_1	σ	K	ρ_s	ϵ_s	ν_1	ν_s	$\mathbb{P}_s$
Whitley											
	0.4	86.4	121	0.7	0.7	50	31.8	21	0.6	17	37.7
Zirilli											
	0.2	99.9	707	0.6	$\sqrt{2}$	10	97.9	987	1.9	5.3	49.4