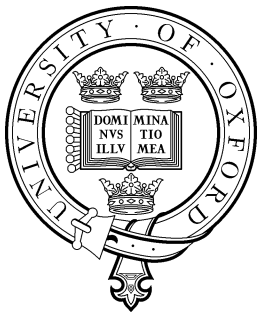
Theory of Killing and Regeneration in Continuous-time Monte Carlo Sampling



Andi Q. Wang
St Peter's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

January 2020

For my family

Acknowledgements

It is said that a thesis is never completed, merely submitted. Nevertheless, I feel a strong sense of closure with the submission of this thesis.

Firstly, I would like to thank my two supervisors, Prof. David Steinsaltz and Prof. Gareth Roberts. Thank you for your encouragement, your tutelage and your support. I have learnt a lot from both of you during the past three years, and not merely things related to Monte Carlo!

Within the Department of Statistics, Oxford, I am grateful to Prof. Paul Chleboun (now at the University of Warwick), Prof. Arnaud Doucet and Prof. Matthias Winkel for having served as my transfer and/or confirmation examiners. Thanks also to Prof. Sam Cohen in the Department of Mathematics for numerous helpful conversations. Within the Department of Statistics, Warwick, I am grateful to Divakar Kumar and Dr. Murray Pollock. It has been a joy to work together, and I look forward to future projects! I am grateful also to Prof. Martin Kolb (University of Paderborn) for our collaboration during the early work of this thesis. I am very grateful to Prof. Jose Blanchet and his group in Management Science and Engineering, Stanford, for their kind hospitality in hosting me for two months, March – May 2018. I am grateful to Prof. George Deligiannidis and Prof. Gersende Fort for serving as my thesis examiners.

I am grateful to my colleagues on the OxWaSP CDT, and the third floor probability group, for a fun and stimulating environment. In particular, it has been a pleasure sharing with my officemates Alex, Bohao, Ian this past year as I have been writing up this thesis. I will miss our fun and interesting distractions!

Financial support from the EPSRC and the OxWaSP CDT (through grant EP/L016710/1) is gratefully acknowledged; without it I would never have embarked upon this DPhil!

I am particularly thankful to the postgraduate Thesis community within St. Ebbe's church, for their support and encouragement over the past

four years. Far too many people have encouraged me, challenged me, supported me to name in person. Let me just mention Tables 5, 7 and 7 over the years, the Snek Squad, the Hot Sauce Choir, and everyone who has been held captive by my various passions; maths, board games, theology, or otherwise: you know who you are! Thanks in particular to Abe Ng, a friend who sticks closer than a brother.

Finally, thanks to my parents and Lindi – your prayers and support have been a real pillar for me, during my whole course of life, not just for these past 4 years. Thank you.

I thank the Lord for my time in Oxford – for bringing this Thesis and life chapter to a close, for all the people I have met, all the things I have done, all the things I have learned.

S.D.G.

Andi Wang

First submitted: 27 September 2019

Corrections made: 15 January 2020

Abstract

We consider the theory of killing and regeneration for continuous-time Monte Carlo samplers. After a brief introduction in Chapter 1, we begin in Chapter 2 by reviewing some background material relevant to this thesis, including *quasi-stationary Monte Carlo methods*. These methods are designed to sample from the quasi-stationary distribution of a killed Markov process, and were recently developed to perform scalable Bayesian inference.

In Chapter 3 we prove natural sufficient conditions for the quasi-limiting distribution of a killed diffusion to coincide with a target density of interest. We also quantify the rate of convergence to quasi-stationarity by relating the killed diffusion to an appropriate Langevin diffusion. As an example, we consider a killed Ornstein–Uhlenbeck process with Gaussian quasi-stationary distribution.

In Chapter 4 we prove convergence of a specific quasi-stationary Monte Carlo method known as 'ReScaLE'. We consider the asymptotic behavior of the normalized weighted empirical occupation measures of a diffusion process on a compact manifold which is also killed at a given rate and regenerated at a random location, distributed according to the weighted empirical occupation measure. We show that the weighted occupation measures almost surely comprise an asymptotic pseudo-trajectory for a certain deterministic measure-valued semiflow, after suitably rescaling the time, and that with probability one they converge to the quasi-stationary distribution of the killed diffusion.

In Chapter 5 we introduce the Restore sampler. This is a continuous-time sampler, which combines general local dynamics with rebirths from a fixed global rebirth distribution, which occur at a state-dependent rate. In certain settings this rate can be chosen to enforce stationarity of a given target density. The resulting sampler has several desirable properties: simplicity, lack of rejections, regenerations and a potential coupling from

the past implementation. The Restore sampler can also be used as a recipe for introducing rejection-free moves into existing MCMC samplers in continuous time. Some simple examples are given to illustrate the potential of Restore.

We conclude the thesis in Chapter 6 with some concluding comments and open questions for future work.

Contents

1 Introduction **1**

1.1 Thesis outline 3

2 Background **5**

2.1 Introduction 5

2.2 Monte Carlo inference 5

2.2.1 Markov chain Monte Carlo 6

2.2.2 Regeneration 9

2.2.3 Continuous-time Monte Carlo 11

2.3 Quasi-stationarity 13

2.3.1 Definitions and basic properties 14

2.3.2 Characterisation of quasi-stationary distributions 18

2.3.3 Quasi-stationary distributions of killed diffusions 20

2.4 Quasi-stationary Monte Carlo methods 21

2.4.1 Simulation of quasi-stationary distributions 23

2.5 Regenerative processes and quasi-stationarity 26

3 Theoretical properties of quasi-stationary Monte Carlo methods **28**

3.1 Introduction 28

3.1.1 Summary of main results 28

3.1.2 Convergence to quasi-stationarity 31

3.1.3 Rate of convergence 32

3.1.4 Examples of κ 34

3.2 Example: Ornstein–Uhlenbeck process targeting a Gaussian density . 35

3.3 Mathematical preliminaries 38

3.3.1 Definitions 38

3.3.2 The killed Markov semigroup 39

3.3.3 Operator theory 41

3.4 Proof of Theorem 3.1.1 . . . 44
3.5 Rates of Convergence . . . 46
3.5.1 Proof of Theorem 3.1.2 . . . 46
3.6 Discussion . . . 50

4 An approximation scheme for quasi-stationary distributions of killed diffusions 53
4.1 Introduction . . . 53
4.2 Introduction . . . 53
4.2.1 Main results . . . 54
4.3 Preliminaries . . . 58
4.3.1 General background . . . 58
4.3.2 Fixed Rebirth processes . . . 62
4.4 Deterministic Flow . . . 68
4.4.1 Basic properties . . . 68
4.4.2 Stability of π . . . 69
4.5 Asymptotic Pseudo-Trajectories . . . 72
4.5.1 Basic properties and definitions . . . 72
4.5.2 Poisson Equation . . . 75
4.5.3 Bounding the discrepancy . . . 77
4.5.3.1 $\epsilon_t^{(1)}(s)f$ and $\epsilon_t^{(2)}(s)f$ . . . 78
4.5.3.2 $\epsilon_t^{(3)}(s)f$ . . . 79
4.5.3.3 $\epsilon_t^{(4)}(s)f$ . . . 79
4.5.3.4 $\epsilon_t^{(5)}(s)f$ . . . 81
4.5.3.5 Concluding the proof . . . 82
4.5.4 Proof of Corollary 4.2.3 . . . 82
4.6 Subsequent work . . . 82

5 The Restore sampler 84
5.1 Introduction . . . 84
5.1.1 Previous work . . . 86
5.2 The Restore process . . . 88
5.2.1 Chapter overview . . . 90
5.3 Invariance . . . 92
5.3.1 Countable state space setting . . . 92
5.3.1.1 Example . . . 93
5.3.2 Symmetric diffusions . . . 94

5.3.2.1 Examples . . . 98
5.3.3 Jump processes . . . 99
5.3.3.1 Example . . . 104
5.4 Limiting properties . . . 104
5.4.1 Coupling from the past . . . 104
5.4.1.1 Example: classical rejection sampler . . . 107
5.4.2 Central Limit Theorem . . . 109
5.5 Simulation . . . 111
5.5.1 Minimal rebirth distribution . . . 111
5.5.2 Tuning the parameter C . . . 113
5.5.3 Truncated killing . . . 114
5.6 Examples . . . 121
5.6.1 Multivariate Gaussian . . . 121
5.6.1.1 Minimal rebirth distribution . . . 121
5.6.2 Cauchy posterior . . . 124
5.6.2.1 Minimal rebirth distribution . . . 128
5.6.3 Jump process: correcting a poor approximation . . . 129

6 Conclusions and further work **134**

A Symmetric diffusions **137**
A.1 Basic Assumptions . . . 137
A.2 Theorem 2.5 of Demuth and van Casteren [2000] . . . 138
A.3 Proof of Lemma 5.3.6 . . . 139
A.4 Proof of Theorem 5.3.7 . . . 141

Bibliography **145**

Chapter 1

Introduction

Monte Carlo methods emerged in the middle of the twentieth century, originally in the context of simulating neutron trajectories at Los Alamos, and have since become a fundamental tool in scientific computing. Such methods reduce the problem of computing intractable integrals to the simulation of random draws from probability distributions. It turns out that even though the desired quantities are deterministic, such stochastic algorithms can massively outperform their deterministic counterparts, especially as the dimension of the problem increases.

The broad goal of Monte Carlo methods is to draw samples from a given probability distribution π. For instance, π might be a Bayesian posterior distribution arising from some scientific inquiry. A great breakthrough for Bayesian inference, and Monte Carlo methods more broadly, was the development of Markov chain Monte Carlo (MCMC), beginning in the 1950s. MCMC significantly expanded the class of scientific models which were amenable to Bayesian analysis, providing simple and elegant algorithms to obtain draws from a given target.

Today, over 60 years on from the original Monte Carlo revolution, a new revolution is taking place. With the monumental advances in data gathering and data storage technology, the size of modern data sets are now of a scale inconceivable to the original developers of Monte Carlo methods. Often referred to nebulously as the problem of 'Big Data', the result is that many of the traditional, established methods are becoming increasingly unsuitable. For example, the seemingly innocuous operation of evaluating a posterior density pointwise is no longer feasible when the data set is extremely large, as such an operation involves iterating over the entire data set.

As such, many new algorithms have emerged over the last twenty years to address these challenged posed by modern data sets. One particular class of methods – which is the focus of this thesis – is the class of *continuous-time Monte Carlo methods*. The methods discussed in this thesis are all *exact* methods, meaning that the algorithms

all target the exact form of the target distribution, albeit potentially asymptotically. Numerous approximate methods exist to perform inference for large data sets; two representative examples are stochastic gradient methods, as described in Welling and Teh [2011], or deterministic approximation methods such as INLA Rue et al. [2009]. We will not discuss such methods in this thesis.

Despite the fact that computers are fundamentally discrete machines, continuous-time methods have nevertheless shown considerable promise in tackling the challenges associated with large data sets. As will be reviewed in Chapter 2, continuous-time processes have long had application in Monte Carlo, but it is only in very recent years that we have seen the development of genuinely continuous-time algorithms. Such algorithms include piecewise-deterministic Markov chain Monte Carlo (PDMCMC), Fearnhead et al. [2018], Bouchard-Côté et al. [2018], Bierkens et al. [2019] and quasi-stationary Monte Carlo (QSMC), Pollock et al. [2016], Kumar [2019].

Broadly speaking, the motivation behind continuous-time Monte Carlo is twofold. First, it is often more straightforward to construct nonreversible processes in continuous time, typically by augmenting the target with a momentum variable. Nonreversible processes often converge to equilibrium faster than reversible processes; see, for instance, Bierkens [2016]. Secondly, in terms of large data sets, continuous-time methods often only require access to a small number of data points per iteration, without compromising the exactness of the method. This can provide massive computational advantages, see Pollock et al. [2016], Bierkens et al. [2019].

With both PDMCMC and QSMC methods, an object of fundamental importance is an inhomogeneous Poisson process defined by a rate. Suppose we have a continuous-time Markov process $X = (X_t)$ evolving on a state space E, and a measurable function $\kappa : E \to [0, \infty)$. We can define a random time τ_∂ as follows:

$$\tau_\partial := \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s) \, \mathrm{d}s \geq \xi \right\},$$

where the random threshold $\xi \sim \mathrm{Exp}(1)$ is independent of X. For both PDMCMC and QSMC methods, inhomogeneous Poisson processes with such arrival times are crucial for ensuring convergence to stationarity of the process. Indeed, these times function as the continuous-time analogue to the acceptance–rejection step in traditional Markov chain Monte Carlo samplers.

In this thesis, we will begin by studying the class of quasi-stationary Monte Carlo methods. These methods make use of *killing* in continuous time in such a way that

the resulting quasi-stationary distribution coincides with the target π. Namely, when τ_∂ is defined as above, π is such that

$$\mathbb{P}_\pi(X_t \in \cdot \,|\tau_\partial > t) = \pi, \quad t \geq 0.$$

The goal of QSMC methods is to then sample from such a quasi-stationary distribution.

QSMC methods debuted with the ScaLE algorithm of Pollock et al. [2016], where it was shown that these methods allowed for the principled use of *subsampling*, one technique which can be used to address the issue of large data sets. The ScaLE algorithm is one of the few algorithms which can make use of subsampling without incurring error. We will consider the fundamental convergence theory related to QSMC in Chapter 3, and consider an alternative implementation of QSMC based on stochastic approximation and prove its validity on compact spaces in Chapter 4.

After studying the theory of QSMC methods in Chapters 3 and 4, we will then discuss a novel continuous-time sampler, which we have called the Restore sampler in Chapter 5. This is a new continuous-time Monte Carlo method, which naturally exhibits *regeneration times*. These are times at which the underlying process 'starts again', and the presence of such times provides numerous theoretical and practical benefits.

1.1 Thesis outline

Broadly speaking, the aim of this thesis is to expand the existing theory for continuous-time Monte Carlo methods, focusing on the mechanisms of killing and regeneration based on inhomogeneous Poisson processes. We will give theoretical results which are immediately applicable for QSMC methods in Chapters 3 and 4, and in Chapter 5 we contribute a novel continuous-time sampler.

In Chapter 2 we will discuss the background needed for the rest of the thesis. We will briefly survey the development of Markov chain Monte Carlo, and then focus on two main points: regeneration, and continuous-time methods. I will then discuss the basic theory of quasi-stationarity, especially for diffusions, and then discuss quasi-stationary Monte Carlo methods. Finally, I will discuss regenerative processes, which will play a pivotal role later in the thesis.

The focus of Chapters 3 and 4 will be to provide fundamental theory to further our understanding of QSMC algorithms. In Chapter 3 we give some fundamental theoretical results which extend the somewhat rudimentary theory provided in the

original QSMC article Pollock et al. [2016]. Namely, we formulate sufficient conditions under which the quasi-stationary distribution of a killed diffusion process coincides with a density function of interest. We also prove a result concerning the spectral gap, which provides useful heuristic guidance regarding rates of convergence.

In Chapter 4 we address the practical issue of implementing QSMC, and prove that an alternative sample path approach is also valid in the context of QSMC. The resulting QSMC method is known as ReScaLE (Regenerating ScaLE), whose computational properties have been studied in the work of Kumar [2019]. In Chapter 4 we prove convergence of the ReScaLE algorithm to the quasi-stationary distribution, in the compact state space setting. To do this we will make use of the 'ODE method', following the path laid out in Benaïm et al. [2002] and Benaïm et al. [2018].

In Chapter 5, we will discuss the Restore sampler. This is a novel continuous-time sampler, constructed by introducing regeneration times to an underlying fixed Markov process. The novelty of the Restore sampler is that it gives a recipe for aligning the local dynamics defined by a given Markov process and global dynamics given by a fixed rebirth distribution, in such a way as to cause π-stationarity. This can be done in settings where neither the local nor global dynamics are themselves π-invariant. In this chapter we construct the Restore sampler, and then prove π-invariance in several settings. We complement this with some theoretical results, including a *coupling from the past* implementation, and a result on a truncated version of the algorithm. Finally, we demonstrate the potential of the sampler by giving some simple examples.

Finally, in Chapter 6 I will sum up the work of this thesis and present some discussion and limitations of this work, and give some open problems for future research.

Chapter 2

Background

2.1 Introduction

In this chapter I will survey some fundamental background material relevant for the work of this thesis.

I will first discuss Monte Carlo methods, in particular mentioning regenerative methods and continuous-time methods.

I will then present the basic theory of quasi-stationarity. I will then discuss the recent class of quasi-stationary Monte Carlo (QSMC) methods, which are the subjects of Chapters 3 and 4.

Finally, I discuss a class of continuous-time Markov processes known as 'regenerative processes', which appear in Chapter 4 as auxiliary processes but are central to Chapter 5.

2.2 Monte Carlo inference

Many practical problems can be reduced to the computation of an integral. Direct approaches to evaluating such integrals analytically or otherwise can be cumbersome and time-consuming. It turns out that with modern computing power, a *stochastic* algorithm, based loosely on repetition and averaging, can provide very good estimates at a fraction of the cost.

For example, suppose I simultaneously threw ten (fair) 6-sided dice. What is the probability that the sum equals precisely 42? Of course, this can be calculated exactly. In this case, the most straightforward approach – using probability generating functions – would involve finding the coefficient of x^{42} in $(x+x^2+x^3+x^4+x^5+x^6)^{10}$. On the other hand, it is very straightforward to simulate virtual realisations of dice

rolls, and simply use, say, the proportion of 42s out of several thousand games as a reasonable estimate.

Within Bayesian statistics, the goal is to understand a *posterior* distribution of the form

$$\pi(x) = C\,\pi_0(x)\prod_{i=1}^{N} f_i(x), \tag{2.1}$$

for some functions $\pi_0, f_1, \ldots, f_N$. Here π_0 represents the *prior* distribution, encoding our prior beliefs about the state of the system. The f_i are the *likelihood* functions corresponding to the statistical model under scrutiny, representing the data we have gathered. C is the normalising constant, and the posterior reflects our updated beliefs having conducted the experiment.

Historically, priors and likelihoods were chosen so that the resulting posteriors π were *conjugate*, that is, falling within the same family of 'nice' distributions as the prior π_0. Thus, since the posteriors were guaranteed to be well-understood, most quantities of interest such as quantiles or moments were immediately available.

It was not until the advent of general purpose Monte Carlo algorithms that this constraint of conjugacy was removed. If we are able to draw samples $X_1, \ldots, X_n$ from π , then we can unbiasedly estimate any integral $\pi(f)$ (for reasonable functions f) via the sample average

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i). \tag{2.2}$$

The original development of Monte Carlo methods is typically attributed to Stanislaw Ulam and John von Neumann in the late 1940s, as described in Robert and Casella [2011]. The name 'Monte Carlo' was suggested by Nicholas Metropolis, in reference to the casino in Monaco that Ulam's uncle would frequent, as reported in Metropolis [1987].

2.2.1 Markov chain Monte Carlo

The great breakthrough for Bayesian inference was the development of Markov chain Monte Carlo (MCMC). π is the target density, assumed to be positive. The idea behind MCMC is to *construct* a Markov chain whose invariant distribution coincides with the target distribution π. Then, assuming we can simulate the chain, and under sufficient regularity conditions, we can use runs of the Markov chain as approximate draws from π.

Mathematically, the development of MCMC brought about a paradigm shift in the theory of Markov chains. In traditional probability theory, the particular problem

or model under scrutiny presents the mathematician with a Markov chain. The goal is then to understand the chain: its recurrence structure, its limiting behaviour, its stationary distribution (if one exists). MCMC turned this on its head: we *begin* with a desired distribution π, and *build* a chain with π as its stationary distribution. Then, we prove that the chain indeed converges to π, and ideally understand its rate of convergence to π.

The most famous MCMC algorithm is the celebrated Metropolis–Hastings (MH) algorithm, Metropolis et al. [1953], Hastings [1970]. At each step of the algorithm a new location is proposed, given the current location x, according to a density $q(x, \cdot)$, the *proposal kernel*. This proposal is then accepted with a certain probability, or rejected. Define the acceptance probability $\alpha(x, y)$ to be

$$\alpha(x, y) := \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \wedge 1, \tag{2.3}$$

for each x, y. The general Metropolis–Hastings algorithm is given in Algorithm 1.

Algorithm 1 Metropolis–Hastings

initialise: $X_0 = x_0, i = 0$
while $i < n$ **do**
    $i \leftarrow i + 1$
    simulate $Y_i \sim q(X_{i-1}, \cdot)$
    with probability $\alpha(x, y)$
        $X_i \leftarrow Y_i$
    else
        $X_i \leftarrow X_{i-1}$
return $X_1, \ldots, X_n$.

It can be proven under mild conditions (see Robert and Casella [2004]) that the output $X_1, \ldots X_n$ forms a π-reversible Markov chain, that is irreducible and aperiodic, so by the ergodic theorem, the sample average as in (2.2) can be used to approximate $\pi(f)$.

The MH algorithm is an incredibly elegant algorithm for sampling approximately from π. It is very flexible; applicable for both discrete and continuous state spaces, and requires only pointwise evaluations of the target density (which need not even be normalised). Despite its simplicity, it is surprisingly effective. As such, over half a century later the Metropolis–Hastings algorithm — even in vanilla form, as presented in Algorithm 1 — is still enduringly popular and widely used by practitioners.

Popular choices of the kernel $q(x, y)$ include a mean-zero distribution centered around the current location x, leading to *Random Walk Metropolis* (RWM), or al-

ternatively choosing $q(x, y)$ to be independent of x – so at each iteration we propose from a fixed distribution – leading to what is known as the *Independence Sampler*.

In practice, however, there are several issues which need to be overcome in order for MH to be effective.

1. One particularly prominent issue is *burn-in*. Early samples from the Markov chain will not be distributed according to the stationary distribution; standard results concerning exactness of MCMC algorithms are asymptotic results, which are valid only as the number of steps $n \to \infty$. Practically, this means it is not clear how long to run the chain to ensure convergence to equilibrium. Related to this is the mathematical issue of reversibility. The MH chain, by construction, is always reversible. It has been shown, however, that non-reversible processes can converge to equilibrium much faster than reversible processes; see for instance Bierkens [2016], Diaconis et al. [2000].

2. The output of an MCMC algorithm is a realisation of a Markov chain, and the observations are not (in general) independent. This poses an issue when attempting to describe the variance of the estimators obtained.

3. One increasingly prominent issue is the fact that MCMC algorithms are inherently sequential, which means that they are not amenable to straightforward parallelisation, as would be suitable for modern computing architecture.

4. For complicated target densities the MH chain can have a tendency to get 'stuck' in regions of the state space: acceptance probabilities can become prohibitively low, rendering the sampler highly inefficient. The chain exhibits many rejections, where it simply remains stationary for long periods of time and fails to adequately explore the target.

5. In today's digital age it is not uncommon to encounter vast data sets, with the value of N in (2.1) taking values in the million, billions, or even more. In this setting, pointwise evaluation of π – which is required multiple times in every iteration of the MH algorithm – is expensive, or in some cases impossible.

The work of this thesis is generally motivated by a desire to overcome (some of) these hurdles. Quasi-stationary Monte Carlo methods, for instance, the topics of Chapters 3, 4, are designed to tackle issues 3 and 5. The Restore sampler of Chapter 5, is a simple and straightforward sampler that, we hope, can contribute to the overcoming problems 1 through 4.

I will now go on to discuss one line of research which has aimed to overcome the first three of the issues mentioned above.

2.2.2 Regeneration

Consider a discrete-time Markov chain $X = (X_n)_{n \geq 0}$. Intuitively, a *regeneration time* for X is a stopping time at which the chain 'starts afresh', and goes on to evolve independently from the past with an identical law. For a discrete state-space Markov chain, the hitting times of any fixed state form a sequence of regeneration times.

Processes exhibiting regenerations have been studied extensively in the operations research literature; see, for instance, the book of Asmussen [2003]. Regenerations are important in many areas, such as queuing and renewal theory. We focus our discussion now on regenerations in the context of Monte Carlo inference, which goes back to Mykland et al. [1995].

Suppose we are in a setting where we know there exists a sequence of almost surely finite regeneration times $T_1 < T_2 < \ldots$. It follows then that the *tours* of the chain between successive regeneration times are independent and identically distributed. Provided the chain has the correct invariant distribution, we can look at the variables

$$Z_i = \sum_{j=T_{i-1}+1}^{T_i} f(X_j),$$

along with the tour lengths $N_i = T_i - T_{i-1}$ and then consider

$$\hat{f}_n = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n N_i}.$$

It can be shown using the strong law of large numbers that $\hat{f}_n$ will converge almost surely to $\pi(f)$. If the chain is observed for a fixed number of tours, burn-in is not necessary, since the average along each tour has the correct expectation, namely $\pi(f)$. Moreover since the pairs (Z_i, N_i) are independent and identically distributed (i.i.d.), the discrepancies

$$\sqrt{n}(\hat{f}_n - \pi(f))$$

converge to a normal $N(0, \sigma_f^2)$ in distribution (provided the variances are finite). Unlike for traditional MCMC, this asymptotic variance σ_f^2 is much easier to estimate in this regenerative setting. Finally, since the tours are genuinely i.i.d., they can be simulated in parallel.

The study of regeneration in the context of MCMC originates with the work of Mykland et al. [1995]. In this work a method for introducing regeneration times using

the technique of *Nummelin splitting*, Nummelin [1978], is described. A substantive advance came later in Hobert et al. [2002], where new conditions were formulated under which the central limit theorem of Mykland et al. [1995] holds, which are considerably easier to verify in practice. As pointed out in Mykland et al. [1995], one general issue with regenerative methods is that simply identifying regeneration times within a sampler is really a means of analysis, and does not in itself speed up mixing. Brockwell and Kadane [2005] took a slightly different approach. By adding an artificial atom outside of the space itself, and constructing an augmented target density, they proposed a new sampler which possesses regenerations. Further work has been done developing new methods to utilise regenerations, such as Minh et al. [2012], Nguyen [2015] and Lee et al. [2014].

The use of regeneration times is also integral to the general approach of Meyn and Tweedie [1993, 1994], and also Rosenthal [1995], to the study of stability and convergence to equilibrium of Markov chains. See also the doctoral thesis of Latuszynski [2009], where regenerative methods are used to prove various results, including central limit theorems for uniformly ergodic Markov chains under a second moment condition. We will return to the topic of regenerations when discussing the closely-related notion of regenerative processes in Section 2.5.

I will now briefly discuss the related technique of *coupling from the past* (CFTP), which goes back to the seminal work of Propp and Wilson [1996]. CFTP is an elegant technique also proposed to overcome issues related to burn-in. CFTP is an MCMC technique typically applicable in situations when the chain in question has a strong sense of stochastic monotonicity, as pointed out in Lee et al. [2014], such as random walks on finite intervals. Intuitively, the idea is to run the chain from the distant past ($t = -\infty$) until the present ($t = 0$), using a sequence of generating random variables whose distribution is known. If the present value at $t = 0$ only depends on a finite number of these generating random variables, since the chain must have reached its equilibrium state, we have obtained an exact draw from its invariant distribution. This was debuted for discrete spaces in Propp and Wilson [1996], but has been extended to more general spaces, such as the multigamma coupler of Murdoch and Green [1998].

In Chapter 5, we will introduce the Restore sampler, which is a continuous-time process on a general state space which innately possesses regeneration times, and will also be amenable to CFTP in certain situations.

2.2.3 Continuous-time Monte Carlo

MCMC was a major breakthrough for obtaining samples from a given density functions. The Metropolis–Hastings algorithm was and is the primary workhorse for such methods. Mathematically, it makes use of the (now mature) theory of discrete-time Markov chains on general state spaces, as described in the book of Meyn and Tweedie [1993].

However, it was realised that *continuous-time* processes can be immensely valuable in the context of MCMC. A very natural context in which they can play a role is in the choice of proposal kernel $q(x, y)$ in Algorithm 1.

Suppose π is a smooth target density on $\mathbb{R}^d$. Ideally, we would like to choose a proposal kernel $q(x, y)$ which is well-adapted to the target density π, in that it naturally drifts towards regions of higher π-mass. One continuous-time process which does this is the Langevin diffusion, defined as the solution of the following stochastic differential equation (SDE)

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \tag{2.4}$$

where B_t is a standard Brownian motion on $\mathbb{R}^d$. Under certain conditions on the target π (see Roberts and Tweedie [1996]), the Langevin diffusion has a unique weak solution and, most importantly, has invariant density given by π, with the law of the diffusion at time t converging weakly to π as $t \to \infty$.

Thus the law of the Langevin diffusion at a fixed time t, started from a given position x, would make an excellent proposal kernel $q(x, \cdot)$. However, except in very special cases (such as when the target is a Gaussian), we are unable to obtain exact draws from this kernel, and so a discretised form of (2.4) is used instead. The time-discretised form of (2.4) can be directly used to approximately sample from π, giving rise to the Unadjusted Langevin Algorithm (ULA), or it can be nested within an MH accept-reject step, giving rise to the Metropolis-Adjusted Langevin Algorithm (MALA), see Roberts and Rosenthal [1998].

Another reason why continuous-time processes are valuable for Monte Carlo inference is that such processes more naturally exhibit *momentum*, and as such are often *nonreversible*. Since nonreversible processes can possess a sense of directionality, they can avoid the problem of 'backtracking' which plagues reversible methods such as the vanilla MH algorithm, see Bierkens [2016], Diaconis et al. [2000].

One particularly influential MCMC method which makes use of continuous-time dynamics is Hamiltonian Monte Carlo (HMC), Duane et al. [1987], Neal [2011], Betancourt [2017]. The idea here is to augment our particle exploring the state space

with a momentum variable. The joint variables of space and momentum then evolve together according to *Hamiltonian dynamics*, a system of ordinary differential equations (ODEs). Unfortunately, as was the case with the Langevin diffusion, exact simulation of the Hamiltonian flow is not possible, and so a time discretisation is used, and an MH accept-reject step is used to correct for the error incurred. HMC has been demonstrated to be very effective at exploring complex distributions in high dimensions, and it is precisely the addition of this momentum which enables it to do so.

The above methods, ULA/MALA and HMC, made use of continuous-time processes, but ultimately are still embedded within discrete-time algorithms. In recent years a new class of continuous-time processes have found application in Monte Carlo inference, leading to genuinely continuous-time algorithms: piecewise-deterministic Markov processes (PDMPs).

These processes, first introduced by Davis [1984], follow a deterministic trajectory (governed by an ODE) for a random amount of time. After this random amount of time, the process jumps to a new location, and then evolves according to a (possibly new) ODE, for another random period of time, and so on. Such processes were first discussed in the context of general Monte Carlo inference in Peters and de With [2012], and they have been used extensively within the physics literature; see, for instance Michel et al. [2014, 2015].

Currently, popular PDMP-based methods include the Bouncy Particle Sampler (BPS) of Bouchard-Côté et al. [2018] and the Zig-Zag sampler of Bierkens et al. [2019]. In both cases, the particle x_t exploring $\mathbb{R}^d$ is augmented with a *velocity* v_t. The particle $z_t = (x_t, v_t)$ moves in straight lines, according to its velocity vector which is piecewise-constant. For the BPS, the velocity is either confined to the surface of the unit sphere in $\mathbb{R}^d$, or can take any value in $\mathbb{R}^d$. For the Zig-Zag, the velocity is limited to the finite set $\{-1, 1\}^d$, hence the name.

For each of the two methods, there is a (particular) event rate $z \mapsto \lambda(z) \in [0, \infty)$, which defines the rate at which switching events occur. Mathematically, the times are given by $T_0 = 0$, and

$$T_i = \inf \left\{ t \geq T_{i-1} : \int_{T_{i-1}}^{t} \lambda(z_t) \, \mathrm{d}t \geq \xi_i \right\}, \quad i = 1, 2, \ldots, \tag{2.5}$$

where $\xi_1, \xi_2, \ldots$ are an i.i.d. sequence of Exp(1) random variables.

At the event times, say at T_i, the velocity vector v_{T_i} jumps. For the BPS, it is reflected (in the standard Euclidean sense) in the (hyper-)plane normal to the gradient

of the log-density at the point x_{T_i}. For the Zig-Zag, one of the velocity components is flipped $-1 \leftrightarrow +1$. The BPS, in addition, has *refreshment* events, which occur at a constant rate, at which the velocity is resampled from some fixed distribution, such as an isotropic Gaussian on $\mathbb{R}^d$ or a uniform on the sphere. This is to ensure ergodicity of the resulting process.

For each process, the event rate λ is defined in such a way that, given the particular bounce mechanism, the marginal x-component of the resulting invariant distribution of the PDMP equals π.

PDMPs have attracted a lot of attention within Monte Carlo research for several reasons. Firstly, the resulting samplers are nonreversible. The additional momentum variable precisely allows the particle to move in the same direction for 'as long as possible' before bouncing. In addition, the resulting samplers do not exhibit rejections; the particle is constantly moving. The equivalent phenomenon would be a very short inter-bounce time, but post-bounce, the dynamics are such that another rapid bounce is very unlikely, if not impossible. The Zig-Zag, in addition, is amenable to *subsampling*, a computational technique which allows the sampler to handle posterior distributions comprising a large number of observations efficiently. We will return to subsampling in Section 2.4.

It is this lack of rejections which showcases the utility of moving into continuous-time. In discrete time, since there is an inherent lower bound on the amount of time spent in a given state (namely, 1), certain transitions must be prohibited to avoid spending too much time in states of low probability. On the other hand, in continuous time the particle is able to quickly leave poor states, and so no rejection mechanism is required.

The Restore sampler, which will be described in Chapter 5, is a new continuous-time sampler which shares many characteristics with the methods described above. It features event times as in (2.5), which will turn out to be regeneration times as described above in Section 2.2.2.

We now turn to discuss the probabilistic notion of *quasi-stationarity*. We return to continuous-time Monte Carlo in Section 2.4, when we discuss quasi-stationary Monte Carlo methods.

2.3 Quasi-stationarity

In this section I will introduce the notion of quasi-stationarity and related concepts, giving basic definitions and some foundational results. This material is largely syn-

thesised from the book of Collet et al. [2013] and the review paper Méléard and Villemonais [2012]. For a list of related publications (up to 2015), see the bibliography of Pollett [2015].

2.3.1 Definitions and basic properties

The theory of quasi-stationarity concerns the limiting properties of absorbed Markov processes conditioned on non-absorption. In full generality let us consider a continuous-time right-continuous Markov process $X = (X_t)_{t\geq 0}$ on a measurable state space $\mathcal{X} \cup \{\partial\}$, where $\partial \notin \mathcal{X}$ is some state affixed to $\mathcal{X}$, representing the absorbing trap. Of course, the theory of quasi-stationarity can also be considered in discrete-time, mutatis mutandis. Since our subsequent applications are all in continuous time, I will not discuss discrete-time theory here.

Following the approach of Collet et al. [2013], we take the probability space Ω to be the space of right-continuous paths in $\mathcal{X} \cup \{\partial\}$, equipped with the cylinder σ-algebra $\mathcal{F}$ which is generated by the coordinate mappings $X_t(\omega) = \omega(t)$, for each $t \geq 0$. We take the filtration $(\mathcal{F}_t)_{t\geq 0}$ to be a filtration for which X is adapted. We assume that we have a family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}\cup\{\partial\}}$ satisfying the following properties:

- $\mathbb{P}_x(X_0 = x) = 1$, for all $x \in \mathcal{X} \cup \{\partial\}$.
- For any measurable set A, the map $x \mapsto \mathbb{P}_x(A)$ is measurable.
- We have the Markov property: for the shift maps $(\Theta_t)_{t\geq 0}$, $\Theta_t : \Omega \to \Omega, \Theta_t(\omega) = \omega(t + \cdot)$,
$$\mathbb{P}_x(\Theta_t(X) \in A|\mathcal{F}_t) = \mathbb{P}_{X_t}(A), \tag{2.6}$$
$\mathbb{P}_x$-almost surely, for each $x \in \mathcal{X} \cup \{\partial\}$ and measurable $A \in \mathcal{F}$.

For a probability measure μ on $\mathcal{X} \cup \{\partial\}$ we write $\mathbb{P}_\mu = \int \mu(\mathrm{d}x)\mathbb{P}_x$. Thus for delta masses δ_x, $\mathbb{P}_x$ and $\mathbb{P}_{\delta_x}$ denote the same object, which we will just write $\mathbb{P}_x$.

Let τ_∂ denote the *absorption time*, or *killing time*, namely

$$\tau_\partial := \inf\{t \geq 0 : X_t = \partial\}.$$

By convention we put $\inf \varnothing = \infty$. Since we will be concerned only with behaviour up until τ_∂, without loss of generality we take it to be absorbing, so $X_t = \partial$ for all $t \geq \tau_\partial$. However in Chapters 4 and 5 we will consider processes which upon being killed instantaneously return to $\mathcal{X}$. These processes are reviewed in Section 2.5 below.

In our applications we will typically take τ_∂ to be defined through a *killing rate*, $\kappa : \mathcal{X} \to [0, \infty)$, where κ is a locally bounded measurable function. This situation arises when the trap ∂ is only accessible from $\mathcal{X}$ via jumps at the state-dependent rate κ. In this case τ_∂ can be represented as

$$\tau_\partial = \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s) \, \mathrm{d}s \geq \xi \right\}, \tag{2.7}$$

where $\xi \sim \mathrm{Exp}(1)$ is independent of X. This is entirely analogous to (2.5).

In this case we typically construct the killed process by defining the laws $(\tilde{\mathbb{P}}_x)_{x \in \mathcal{X}}$ of an unkilled Markov process evolving on $\mathcal{X}$, then using (2.7) to *define* τ_∂, where ξ is an exogenous independent $\mathrm{Exp}(1)$ random variable. This setting is often referred to as *soft killing* (in contrast to *hard killing*, which typically refers to instantaneous killing at a boundary).

Throughout we will often work with the *sub*-Markovian semigroup $\{P_t : t \geq 0\}$ on $\mathcal{X}$, defined for appropriate functions $f : \mathcal{X} \to \mathbb{R}$ as

$$P_t f(x) = \mathbb{E}_x[f(X_t) 1\{\tau_\partial > t\}].$$

With this definition, we note that the semigroup property still holds, by the Markov property (2.6); namely, for bounded measurable functions f, $x \in \mathcal{X}$ and any $s, t \geq 0$,

$$P_{t+s} f = P_s(P_t f).$$

In the case when τ_∂ arises as in (2.7) for a killing rate κ, we may write this semigroup as $\{P_t^\kappa : t \geq 0\}$, and we have the following Feynman–Kac representation:

$$P_t^\kappa f(x) = \mathbb{E}_x[f(X_t) 1\{\tau_\partial > t\}] = \tilde{\mathbb{E}}_x \left[f(X_t) \, \mathrm{e}^{-\int_0^t \kappa(X_s) \, \mathrm{d}s} \right],$$

where $\tilde{\mathbb{E}}_x$ is the law of the underlying *unkilled* process. The latter equality follows since, conditional on $(X_s : s \in [0, t])$, τ_∂ is the first arrival time of an inhomogeneous Poisson process with rate function $s \mapsto \kappa(X_s)$.

We now define quasi-stationary distributions.

Definition 2.3.1. *A probability measure π on $\mathcal{X}$ is a quasi-stationary distribution if for each $t \geq 0$ and measurable set $B \subset \mathcal{X}$,*

$$\mathbb{P}_\pi(X_t \in B | \tau_\partial > t) = \pi(B).$$

A closely related concept is that of a *quasi-limiting distribution*.

Definition 2.3.2. *A probability measure π on $\mathcal{X}$ is a quasi-limiting distribution if for each $x \in \mathcal{X}$, for each measurable set $B \subset \mathcal{X}$,*

$$\mathbb{P}_x(X_t \in B | \tau_\partial > t) \to \pi(B), \qquad \text{as } t \to \infty.$$

Remark 2.3.3. There is no unified terminology in the literature regarding the definition of a quasi-limiting distribution. This definition is consistent with Kolb and Steinsaltz [2012] and equation (3.6) of Collet et al. [2013], but it differs from that of Méléard and Villemonais [2012] and Champagnat and Villemonais [2017], where a quasi-limiting distribution π is one where there *exists* some initial distribution ν such that

$$\mathbb{P}_\nu(X_t \in B | \tau_\partial > t) \to \pi(B)$$

for each measurable subset B. This alternative definition is in fact equivalent to π being quasi-stationary (Proposition 1, Méléard and Villemonais [2012]). Definition 2.3.2 given above is often referred to in the literature as a *Yaglom limit*.

Proposition 2.3.4. *Let π be a quasi-limiting distribution. Then π is a quasi-stationary distribution.*

Proof. This is analogous to the proof of Proposition 1 of Méléard and Villemonais [2012], which we include for completeness. Take any $x \in \mathcal{X}$. Since π is a quasi-limiting distribution, we know that for any bounded measurable function $f : \mathcal{X} \to \mathbb{R}$,

$$\pi(f) = \lim_{t\to\infty} \mathbb{E}_x[f(X_t)|\tau_\partial > t] = \lim_{t\to\infty} \frac{\mathbb{E}_x[f(X_t)1\{\tau_\partial > t\}]}{\mathbb{P}_x(\tau_\partial > t)} = \lim_{t\to\infty} \frac{P_t f(x)}{P_t(1_\mathcal{X})(x)}.$$

For fixed $s \geq 0$, we apply this with $f(z) = \mathbb{P}_z(\tau_\partial > s) = P_s(1_\mathcal{X})(z)$ for $z \in \mathcal{X}$. By the definition of a quasi-limiting distribution and the Markov property,

$$\mathbb{P}_\pi(\tau_\partial > s) = \lim_{t\to\infty} \frac{P_{t+s}(1_\mathcal{X})(x)}{\mathbb{P}_x(\tau_\partial > t)} = \lim_{t\to\infty} \frac{\mathbb{P}_x(\tau_\partial > t+s)}{\mathbb{P}_x(\tau_\partial > t)}.$$

Similarly, now taking $f(z) = \mathbb{P}_z(X_s \in A, \tau_\partial > s)$ for a fixed measurable subset $A \subset \mathcal{X}$,

$$\begin{aligned}\mathbb{P}_\pi(X_s \in A, \tau_\partial > s) &= \lim_{t\to\infty} \frac{\mathbb{P}_x(X_{t+s} \in A, \tau_\partial > t+s)}{\mathbb{P}_x(\tau_\partial > t+s)} \cdot \frac{\mathbb{P}_x(\tau_\partial > t+s)}{\mathbb{P}_x(\tau_\partial > t)} \\ &= \pi(A)\mathbb{P}_\pi(\tau_\partial > s).\end{aligned}$$

From this we can conclude that $\mathbb{P}_\pi(X_s \in A | \tau_\partial > s) = \pi(A)$. So π is quasi-stationary. □

When the process X is initialised from a quasi-stationary distribution π, by definition this means that the distribution of X_t – conditional on survival – is also equal to π, for all $t \geq 0$. Thus we might hypothesise that the rate of killing – i.e. the rate of transfer to ∂ – is constant over time. Indeed, this is exactly the case, and so when started from quasi-stationarity, the killing time τ_∂ has an exponential distribution.

Theorem 2.3.5. *Let π be a quasi-stationary distribution. Then there exists a constant $\lambda \geq 0$, which depends on π, such that for each $t \geq 0$,*

$$\mathbb{P}_\pi(\tau_\partial > t) = \mathrm{e}^{-\lambda t}.$$

Remark 2.3.6. We may refer to this value of λ corresponding to π as the killing rate or extinction rate associated with π.

Proof. This can be found, say, in [Collet et al., 2013, Section 2.3] and we include it here for completeness.

Similarly to above, consider the function $f(z) = \mathbb{P}_z(\tau_\partial > t)$ for $z \in \mathcal{X}$, where $s \geq 0$ is fixed. Analogous calculations to the above proof show that for fixed $t \geq 0$,

$$\mathbb{P}_\pi(\tau_\partial > t + s) = \mathbb{P}_\pi(\tau_\partial > s)\mathbb{P}_\pi(\tau_\partial > t).$$

Thus there exists some $\lambda \geq 0$ such that $\mathbb{P}_\pi(\tau_\partial > t) = \mathrm{e}^{-\lambda t}$. □

Remark 2.3.7. When τ_∂ is defined through a killing rate κ as in (2.7), if π is a quasi-stationary distribution its corresponding rate of killing λ can be easily computed as

$$\lambda = \int_{\mathcal{X}} \kappa(x)\pi(\mathrm{d}x). \tag{2.8}$$

In general, quasi-stationary distributions are not unique; an absorbed Markov process can possess an infinity of quasi-stationary distributions, even when the process is irreducible prior to killing. This is one aspect where the study of quasi-stationarity differs to the study of stationarity for unkilled Markov processes. For example, in Martinez and San Martin [1994] it is shown that for a one-dimensional Brownian motion on $[0, \infty)$ killed at 0, with drift, for $\alpha > 0$,

$$X_t = B_t - \alpha t, \tag{2.9}$$

there is a continuum of quasi-stationary distributions, indexed by $c \in [-\alpha^2/2, 0)$. The values $c < 0$ correspond to the asymptotic rate of killing precisely as in Theorem 2.3.5; $\lambda = -c$. However by irreducibility (prior to killing) only one of these can

be the quasi-limiting distribution, and it is the extremal quasi-stationary distribution corresponding to $c = -\alpha^2/2$. Of the quasi-stationary distributions, the extremal quasi-limiting distribution has the fastest decay at infinity, as pointed out in Martinez et al. [1998] and similarly by Mandl [1961].

Another regard in which the study of quasi-stationarity differs from stationarity is that the set of conditional measures $\{\mathbb{P}(X_t \in \cdot\,|\tau_\partial > t) : t \geq 0\}$ are not consistent, in the sense that they do not arise as the time marginals of a single stochastic process. This prevents the use of standard probabilistic techniques such as conditioning and the Markov property. Instead – as we will see in Chapter 3 – we will make use of R. Tweedie's R-theory, Tweedie [1974], and use the semigroup approach pioneered by Mandl [1961].

2.3.2 Characterisation of quasi-stationary distributions

We now turn to the issue of characterising quasi-stationary distributions. For simplicity of exposition we begin with the case of a finite state space S. In this setting a Markov process X can be described entirely by its rate matrix. Recall the following classical result.

Proposition 2.3.8. *For an unkilled Markov process with rate matrix Q with entries $\{q(i,j) : i,j \in S\}$, a probability measure μ on S is stationary if and only if*

$$\mu Q = 0.$$

For a killed Markov process, its corresponding rate matrix will possess rows whose sum is *strictly negative*, which reflects how probability mass is escaping. Let us write $\kappa(i)$, $i \in S$, for the defect of row i: $\kappa(i) := -\sum_{j\in S} q(i,j) \geq 0$. The analogous condition for quasi-stationarity is as follows.

Proposition 2.3.9. *For a killed Markov process, a probability measure π on S is quasi-stationary if and only if it satisfies*

$$\pi Q = -\lambda\pi,$$

where $\lambda = \sum_{i\in S} \pi(i)\kappa(i)$, and $\lambda \leq -q(i,i)$ for each $i \in S$.

This is proven for countable state spaces as Theorem 4.4 of Collet et al. [2013]. This λ is of course the same as the exponential killing rate in Theorem 2.3.5.

In terms of the sub-Markovian semigroup $\{P^\kappa_t : t \geq 0\}$, what this entails is that for functions f on S,

$$\pi P^\kappa_t f = \mathrm{e}^{-\lambda t}\pi(f).$$

On finite state spaces, the existence of quasi-stationary probability vectors π follows for certain sub-Markovian generators Q from the Perron–Frobenius Theorem. The version we state here is from [Méléard and Villemonais, 2012, Section 3].

Theorem 2.3.10 (Perron–Frobenius Theorem)**.** *Let $\{P_t^\kappa : t \geq 0\}$ be a sub-Markovian semigroup on a finite state space S such that the entries of $P_{t_0}^\kappa$ are strictly positive for some $t_0 > 0$. Then there exists a unique positive eigenvalue ρ of $P_{t_0}^\kappa$, which is the maximum of the moduli of the eigenvalues. ρ possesses left- and right- eigenvectors π and ψ respectively, that can be chosen to be strictly positive with the following normalisations: $\sum_{i\in S}\pi(i) = 1$, $\sum_{i\in S}\pi(i)\psi(i) = 1$. Setting $\lambda := -\log\rho$, the entries of P_t^κ can then be written*

$$p_t^\kappa(i,j) = \mathrm{e}^{-\lambda t}\psi(i)\pi(j) + O(\mathrm{e}^{-\gamma t}), \quad i,j \in S,$$

for some $\gamma > \lambda$.

Thus it follows that in the case when $P_{t_0}^\kappa$ has strictly positive entries for some $t_0 > 0$, we have the existence of vectors π, ψ and scalar λ as in Theorem 2.3.10 satisfying

$$\lim_{t\to\infty} \mathrm{e}^{\lambda t}\,\mathbb{P}_i(X_t = j) = \psi(i)\pi(j), \quad i,j \in S.$$

It follows that π is the (unique) quasi-limiting distribution, and hence π is quasi-stationary.

Although we will not make use of it in this work, the vector ψ is intimately related to the Q-process, which is, heuristically, the Markov chain on S corresponding to trajectories which are *never killed*. More precisely, for a sequence $0 < s_1 < \cdots < s_k < t$ and points $i_0, \ldots, i_k \in S$, it is possible to define an unkilled Markov process Z started from $i \in S$ with law

$$\tilde{\mathbb{P}}_i(Z_{s_1} = i_1, \ldots, Z_{s_k} = i_k) := \lim_{t\to\infty} \mathbb{P}_i(Y_{s_1} = i_1, \ldots, Y_{s_k} = i_k | \tau_\partial > t).$$

Z is an unkilled process with transition rates given by

$$\tilde{q}(i,j) = q(i,j)\frac{\psi(j)}{\psi(i)}, \quad i \neq j$$
$$\tilde{q}(i,i) = \lambda + q(i,i).$$

The details can be found in [Collet et al., 2013, Chapter 3.2]. Under irreducibility of Q, the resulting Q-process Z is positive recurrent and has stationary distribution given by $(\pi(i)\psi(i) : i \in S)$.

2.3.3 Quasi-stationary distributions of killed diffusions

For this thesis we will focus particularly on the quasi-stationary distributions of killed diffusions. In particular, we focus on the situation when the diffusion process is killed at a time τ_∂ defined via (2.7). Recall that this is often referred to as the *soft killing* setting; *hard* killing typically refers to the situation where the diffusion is killed immediately upon hitting the boundary of some enveloping set. Some key contributions in this field are Mandl [1961], Pinsky [1985], Steinsaltz and Evans [2007], Kolb and Steinsaltz [2012], Champagnat and Villemonais [2017].

Throughout we will be considering diffusions on $\mathbb{R}^d$ of the form

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + dB_t, \tag{2.10}$$

where $A : \mathbb{R}^d \to \mathbb{R}$ is a smooth function. This is of course similar to the Langevin diffusion (2.4), except here we do not assume that $\exp(2A)$ is necessarily integrable. The assumption of gradient-form drifts in (2.10) allows us to employ Hilbert space theory.

In particular, we can define the measure Γ given by

$$\Gamma(\mathrm{d}y) = \gamma(y)\,\mathrm{d}y, \quad \gamma(y) = \exp(2A(y)), \quad y \in \mathbb{R}^d.$$

This is the 'reversing' measure of this diffusion; on the Hilbert space $\mathcal{L}^2(\Gamma)$ the generator can be realised as a self-adjoint operator.

The (formal) infinitesimal generator of (2.10) killed at rate κ is

$$-L^\kappa f(x) = \frac{1}{2}\Delta f(x) + \nabla A(x) \cdot \nabla f(x) - \kappa(x) f(x),$$

where Δ denotes the Laplacian and ∇ the gradient operator on $\mathbb{R}^d$. We will make this precise in Chapter 3. We note that here L^κ is *minus* the generator, following the notation of Kolb and Steinsaltz [2012].

In analogue with Proposition 2.3.9, in order to obtain quasi-stationary distributions for the diffusion (2.10) killed at rate κ, we will be interested in solutions of

$$L^\kappa f = \lambda f,$$

for $\lambda > 0$. In particular we seek positive integrable solutions f, whose eigenvalue λ correspond to the base of the $\mathcal{L}^2$ spectrum, λ_0. This λ_0 is the asymptotic rate of killing associated with the quasi-stationary distribution (Theorem 2.3.5).

A careful analysis in the one-dimensional setting was conducted in Kolb and Steinsaltz [2012], and some extensions to multi-dimensional settings are given in in [Champagnat and Villemonais, 2017, Section 4].

When dealing with diffusions killed at rate κ, one particularly useful object is the transition sub-density $p^\kappa(t, x, y)$. This has the property that

$$P_t^\kappa f(x) = \mathbb{E}_x[f(X_t)1\{\tau_\partial > t\}] = \int p^\kappa(t, x, y) f(y) \, \mathrm{d}\Gamma(y), \quad x \in \mathbb{R}^d,$$

for all reasonable f. Its existence follows from the derivation of Demuth and van Casteren [2000]. We will make use of this throughout this work, and so the main assumptions (known as 'BASSA') are included in Appendix A for convenience.

An important situation is the *high killing* case, first considered in Theorem 4.3 of Kolb and Steinsaltz [2012] and extended in Theorem 4.5 of Champagnat and Villemonais [2017]. This is the situation when there exists a compact set $D_0 \subset \mathbb{R}^d$ such that

$$\inf_{x \in \mathbb{R}^d \setminus D_0} \kappa(x) > \lambda_0.$$

In this setting, Theorem 4.5 of Champagnat and Villemonais [2017] shows that there is a unique quasi-stationary distribution π, and in fact convergence of the conditioned laws to π as in Definition 2.3.2 is exponential.

Intuitively, since the asymptotic rate of killing is λ_0, the conditioned process cannot spend extended periods of time outside of the compact set D_0 — hence the conditioned laws converge to the quasi-stationary distribution.

In the one-dimensional case on $[0, \infty)$, Kolb and Steinsaltz [2012] also consider the low-killing case, when $\lim_{x\to\infty} \kappa(x)$ exists and $\lim_{x\to\infty} \kappa(x) < \lambda_0$. In this case the important condition is whether or not the unkilled diffusion is *recurrent*. When the diffusion is recurrent, there is convergence to quasi-stationarity; otherwise, in the transient case, the process escapes to infinity.

Outside of these results, to the best of my knowledge, there have been few results concerning quasi-stationary distributions of diffusions tailored to the *soft killing* setting with potentially unbounded killing; most authors focus on hard killing at boundaries or restrict to only bounded killing rates.

2.4 Quasi-stationary Monte Carlo methods

A significant part of this thesis is concerned with the application of quasi-stationarity to Monte Carlo problems. In Section 2.2.1 I described the celebrated Metropolis–Hastings algorithm for obtaining samples from π; recall that the MH algorithm requires the pointwise evaluation of π at each iteration, in the calculation of the acceptance probability (2.3).

This calculation, while seemingly innocuous, can become computationally infeasible in modern 'Big Data' settings where the number of observations N is large; pointwise evaluation of the posterior (2.1) is a prohibitive $O(N)$ calculation which involves looping over the entire data set.

One computational technique developed to get around this bottleneck is *subsampling*. As the name suggests, the idea of subsampling is to use a subsample of the data when evaluating the posterior. Instead of using all N data points, a subset of, say, $O(1)$ is used when computing π. This is the approach of the Stochastic Gradient Langevin Dynamics (SGLD) method of Welling and Teh [2011]. SGLD is a very influential and powerful method; however, in practice it is still inexact unless the step sizes shrink, in the sense that the time discretisation employed introduces some bias; the algorithm does not target the exact posterior π.

Direct approaches based around subsampling a random subset of the N terms in (2.1) to obtain an estimate of the product have also been proposed, although this results in unacceptably large errors in the target distribution unless the subset itself is $O(N)$; see, for instance, the discussions in Bardenet et al. [2017].

On the other hand, in recent work, Pollock et al. [2016], a new paradigm was proposed, *quasi-stationary Monte Carlo* (QSMC). The idea is to construct a killed Brownian motion whose *quasi-stationary* distribution coincides with π. The article then proposes algorithms to obtain draws from the quasi-stationary distribution. This killed diffusion framework allowed for the use of subsampling *without* incurring bias: the output of the algorithm is (asymptotically) a set of draws from the exact posterior π. This algorithm was called ScaLE, for the Scalable Langevin Exact algorithm, drawing attention to the fact that it is scalable – able to utilise subsampling – while retaining the exact posterior as its target.

Heuristically, the reason why QSMC methods can employ subsampling without incurring error is because in the killed diffusion framework, access to the target density π is only needed through unbiased estimators of $\nabla \log \pi$. Since π, as in (2.1), is a (large) product of the form

$$\pi \propto \prod_{i=0}^{N} f_i,$$

$\nabla \log \pi$ will be a (large) sum,

$$\nabla \log \pi = \sum_{i=0}^{N} \nabla \log f_i.$$

However, large sums can be easily and cheaply estimated in an unbiased fashion; take I to be a discrete uniform random variable on the set $\{0, 1, \ldots, N\}$ and consider the estimator

$$(N+1)\nabla \log f_I.$$

It is easily seen that this is unbiased for $\nabla \log \pi$, and it is $O(1)$ in cost to simulate. For the full details, the interested reader is referred to Pollock et al. [2016].

Just as MCMC inverted the study of Markov chains by starting from a desired invariant distribution, so QSMC inverts the study of quasi-stationarity. Set

$$\kappa(x) = \frac{\Delta\pi}{2\pi}(x) + K,$$

where $K > 0$ is a constant chosen to ensure that κ is nonnegative (such a K is assumed to exist). Given a sufficiently regular target density π on $\mathbb{R}^d$, in Pollock et al. [2016] it is shown that a Brownian motion killed at rate κ, as in (4.2), possesses π as its (unique) quasi-stationary distribution. In Chapter 3 we will extend this to a more general class of diffusions, and remove some of the unnecessary regularity conditions imposed in Pollock et al. [2016].

Given a stochastic process with a quasi-stationary distribution π, we now turn to the practical question of how to approximately sample from π.

2.4.1 Simulation of quasi-stationary distributions

There are two broad approaches to obtaining draws from a quasi-stationary distribution: particle-based methods and sample-path methods; see Groisman and Jonckheere [2013] for a brief review.

The first approach, using particles, was the approach of the aforementioned ScaLE algorithm of Pollock et al. [2016]. Intuitively, the idea is to simulate a cloud of N particles each evolving independently according to some specified underlying dynamics. On top of this, each particle also experiences killing at a given rate. When a particle is killed, one of the other particles is selected according to some 'fitness' criteria, and duplicated, so the number of particles remains constant. They then continue to evolve independently.

These methods are based on the pioneering work of Burdzy et al. [2000], and are often referred to as Fleming–Viot systems. Such particle methods (in discrete time) are explored in great detail in the book of Del Moral [2004], and continuous-time methods in Del Moral and Miclo [2003]. In the context of simulation of quasi-stationarity distributions, these methods are explored in Villemonais [2011], Groisman and Jonckheere [2013]; see the references therein.

This particle approach is also analogous to the sequential Monte Carlo (SMC) approach, Doucet et al. [2000], Del Moral et al. [2006], to computational Bayesian inference. In the context of ScaLE, the authors were able to leverage the well-established theory of SMC to establish convergence of their particle system to the quasi-stationary distribution.

While the ScaLE algorithm has some desirable theoretical and computational properties, it also suffers several drawbacks. One particularly prominent drawback is the high complexity of implementing the algorithm. A cursory glance through Pollock et al. [2016] can convince the reader that the algorithm is (necessarily) involved. This motivated the search for alternative QSMC methods which would inherit the desirable properties of exactness and scalability, while avoiding the algorithmically complex particle approximation approach.

This brings us to the alternative approach to the simulation of quasi-stationary distributions: sample-path methods, also known as stochastic approximation approaches. This approach to QSMC takes advantage of the fact that quasi-stationary distributions can be written as solutions to fixed point equations in measure spaces (see the later equation (4.16) in Proposition 4.3.12).

These sample-path methods are heuristically as follows. A single trajectory of the killed process is run. When it is killed, the particle is instantaneously reborn at a new random location, which is drawn independently from the particle's empirical occupation measure up to that point. It then continues to evolve from this new starting position until it is killed again, at which point it is again reborn, and so on. The empirical occupation measure of the particle can be seen as an *urn process*, a generalisation of the classical Pólya urn of Eggenberger and Pólya [1923]. Indeed, it exhibits a natural 'self-reinforcement'; the particle is more likely to revisit states which are already popular, further increasing their popularity.

The study of this approach to simulate from quasi-stationary distributions has been studied by various people and remains an area of active research. Its validity was first demonstrated in Aldous et al. [1988], for finite state spaces in discrete time. Here the method was studied precisely in the context of urn processes. This was extended independently by Benaïm and Cloez [2015] and Blanchet et al. [2016], where rates of convergence were also established, both still in finite state spaces and discrete time. This was extended substantially in Benaïm et al. [2018], where general compact state space settings in discrete time were considered.

In the context of QSMC, this approach to the simulation of quasi-stationary distributions yields the QSMC method dubbed Regenerating ScaLE (ReScaLE). The

idea behind ReScaLE is that we simulate a *single* killed diffusion path, whose quasi-stationary distribution coincides with the Bayesian posterior π. Here we simulate a single killed diffusion trajectory, which is killed at a certain rate κ and then regenerated according to its empirical occupation measure. We repeat this for a large number of lifetimes, and then output the overall normalised empirical occupation measure as a proxy for π.

ReScaLE inherits the key properties that motivate the ScaLE algorithm, and algorithmically it is significantly more transparent and straightforward, but this is worthwhile only if the method provably leads to correct results.

Our contribution to this area is Wang et al. [2019b], presented in this thesis as Chapter 4. We prove that this stochastic approximation method converges to the quasi-stationary distribution when considering diffusions on compact Riemannian manifolds. Theoretical analysis of stochastic approximation approaches to numerically solving fixed point equations have been well-studied in finite-dimensional contexts, *cf.* Kushner and Yin [2003], Benaïm [1999], and the references above. At the time when the work of Chapter 4 was done, there was no theory appropriate for the measure-valued, continuous-time context of ReScaLE. The purpose of Chapter 4 is to demonstrate this fundamental convergence property in the compact manifold setting, leaving aside the practical and computational properties of the algorithm. For a thorough investigation of the computational properties, including application to *tall data*, see Kumar [2019].

Since the work of Chapter 4 was first done, several pieces of related work have since been performed. Mailler and Villemonais [2018] demonstrated convergence of a class of *measure-valued Pólya processes*, potentially on noncompact state spaces. When applied to diffusions, their results extend to noncompact state spaces under a few key assumptions. These will be discussed in Section 4.6. Even more recently, Benaïm et al. [2019] proved convergence to the quasi-stationary distribution of the analogous algorithm for a diffusion killed at the boundary of a bounded domain.

The study of such stochastic approximation methods to simulating from quasi-stationary distributions remains an active research area, and many open questions remain. Notably, a proof of convergence on noncompact state spaces of general uniformly elliptic diffusions with potentially unbounded (soft) killing has yet to be shown.

Outside of QSMC, these methods of approximating quasi-stationary distributions have also been applied in ecology, see Schreiber et al. [2018] and to the simulation of the contact process, de Oliveira and Dickman [2005], Dickman and de Oliveira [2005].

2.5 Regenerative processes and quasi-stationarity

One class of processes which is utilised in the work of the previous analyses, Benaïm and Cloez [2015], Benaïm et al. [2018], are *resurrected* or *returned* processes, special cases of *regenerative processes*. Such processes will play an auxiliary role in Chapter 4 – the 'Fixed Rebirth processes' – but will take center stage in Chapter 5, so we take the time now to briefly review them.

Consider a continuous-time Markov X process on a general state space $E \cup \{\partial\}$, where ∂ is some distinguished state. Now fix a probability measure μ, supported on E. We now heuristically define the μ-regenerative process X^μ: X^μ evolves according to the law of X, until the first time τ_∂ when it hits ∂. At that point it is instantaneously reborn; its new location is drawn independently from the distribution μ, and it continues to evolve from there. The times at which the process is reborn will precisely be regeneration times, as discussed in Section 2.2.2.

Regenerative processes have had a long history within the Markov process literature, going back to the very foundations of the field itself in the work of Doob [1945]. In this work, Doob offers the first explicit and general construction of a continuous-time Markov chain on $\mathbb{N} = \{1, 2, \dots\}$ from a given set of transition rates. In order to account for the possibility of explosion, a fixed probability vector $\mu = (p_1, p_2, \dots)$ is chosen, which dictates the instantaneous rebirth distribution after an explosion. This is referenced in Rogers and Williams [2000, III.26] as the *immediate return procedure*.

This is picked up by Feller [1954] to study one-dimensional diffusions. Feller considers a slight generalisation, the elementary return process, which can spend an exponential amount of time on the boundary. The general abstract notion of regenerative processes was then formalised by Smith [1955]. Regenerative processes have gone on to form the bedrock of many fields of study within probability, such as the study of renewal processes and queues, [Grimmett and Stirzaker, 2001, Chapter 10], Asmussen [2003].

In the context of this thesis there is a deep connection between regenerative processes and quasi-stationary distributions. This was first pointed out by Bartlett [1960], in the context of birth–death chains on $\mathbb{N}$. Here the stationary distribution of a regenerative process is used to approximate the quasi-stationary distribution of the birth–death chain. In the context of diffusions on $[0, 1]$ with absorbing boundaries, Ewens [1963, 1964] made use of a regenerative processes to study the behaviour of the diffusion prior to absorption in a population genetics setting. Regenerative processes are also used (in discrete time, for finite state spaces) in Darroch and Seneta [1965]

to study quasi-stationarity. There the authors noted that the resulting invariant distribution 'depends on [the rebirth distribution] to such an extent that it can be made into almost any distribution'. This is a problem we pick up for general state spaces in continuous time in Chapter 5.

More recently, Ferrari et al. [1995], later repeated in [Collet et al., 2013, Section 4.4], used regenerative processes in a very powerful way: to prove the existence of quasi-stationary distributions on countable state spaces. The authors realised that the mapping $\mu \mapsto \Pi(\mu)$ which sends μ to the invariant distribution of the μ-regenerative process (when it exists) is intimately related to quasi-stationary distributions: quasi-stationary distributions are fixed points of the mapping $\mu \mapsto \Phi(\mu)$; see in this work Proposition 4.3.12.

This realisation lead on to further work on the approximation of quasi-stationary distributions, such as Barbour and Pollett [2010, 2012] and the previously mentioned work on stochastic approximation approaches, Blanchet et al. [2016], Benaïm and Cloez [2015], Benaïm et al. [2018]. We, too, will make use of this in Chapter 4.

Finally, we mention the connection between the regenerative processes and the influential PageRank algorithm of Page et al. [1999], famously implemented for the search engine *Google*. The PageRank vector is the invariant distribution of the discrete-time random PageRank surf on a finite graph G. It can be described as follows. At each time step, with fixed probability $d \in [0, 1]$, the surfer will move to a random neighbour of their current vertex, chosen uniformly at random. With probability $1 - d$, however, the surfer 'gets bored' and teleports to a random vertex drawn from some fixed distribution μ on G. Thus PageRank can be seen as a discrete-time regenerative process, where the probability of being killed and teleporting is constant across the graph.

In Chapter 5, we will utilise a resurrected process in a novel way: as the basis of a new Monte Carlo algorithm for drawing samples from π. Namely, we will formulate conditions under which we can design a μ-regenerative process in continuous time which possesses a given target distribution π as its invariant distribution. This we call the *Restore sampler*. It is a genuinely continuous-time method, following the methods discussed in Section 2.2.3, and also naturally possessing regenerations, as discussed in Section 2.2.2.

We are now ready to begin the main work of this thesis in earnest!

Chapter 3

Theoretical properties of quasi-stationary Monte Carlo methods

3.1 Introduction

In this chapter we establish conditions under which the quasi-limiting distribution of a killed diffusion process coincides with a density function of interest. We will also investigate the rate of convergence to quasi-stationarity by comparing to an appropriate Langevin diffusion. These results are of independent interest and useful for the results of Chapter 4. This chapter has been published in *The Annals of Applied Probability* as Wang et al. [2019a]. The text here is very similar to the published article, except some arguments are fleshed out in a little more detail and some typographic errors have been corrected. For example, in Section 3.3.3 some additional details concerning the generator have been added.

This work was done in collaboration with my supervisors Prof. Gareth Roberts and Prof. David Steinsaltz, and with Prof. Martin Kolb (University of Paderborn).

3.1.1 Summary of main results

We summarise here the main results of this chapter, leaving the exact mathematical setting to be explicated in Section 3.3. We will be assuming throughout

Assumption 1. *π is a positive, smooth, and integrable function on $\mathbb{R}^d$.*

Consider the d-dimensional diffusion $X = (X_t)_{t\geq 0}$, defined as the (weak) solution of the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^d, \tag{3.1}$$

where W is a standard d-dimensional Brownian motion and ∇ denotes the gradient operator. We require

Assumption 2. *$A : \mathbb{R}^d \to \mathbb{R}$ is a smooth function such that the SDE (3.1) has a unique non-explosive weak solution.*

Suppose we wish to sample from a distribution π on $\mathbb{R}^d$ with a Lebesgue density, which we will also denote by π — the *target density* — satisfying Assumption 1. We are typically thinking of applications in which we have a statistical model and observed data for which π is the Bayesian posterior distribution. We would like to construct a *killing rate* $\kappa : \mathbb{R}^d \to [0, \infty)$ that makes π into the quasi-limiting distribution of the diffusion X. That is, we define the killing time as in (2.7) via

$$\tau_\partial := \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s) \, \mathrm{d}s \geq \xi \right\}, \tag{3.2}$$

where ξ is an exponential random variable with parameter 1 independent of X. This killing time τ_∂, when the cumulative hazard function $t \mapsto \int_0^t \kappa(X_s) \, \mathrm{d}s$ exceeds the (independent) threshold ξ, is equivalent to the first arrival time of a (doubly stochastic) Poisson process with rate function $t \mapsto \kappa(X_t)$.

We show that

$$\begin{aligned} &\mathbb{P}_x(X_t \in E \,|\tau_\partial > t) \to \pi(E) \text{ as } t \to \infty \\ &\qquad \text{for all } x \in \mathbb{R}^d \text{ and Borel-measurable } E \subset \mathbb{R}^d. \end{aligned} \tag{3.3}$$

To have confidence that this convergence is practically meaningful for a sampling algorithm, we need in addition to have some control over the rate of the convergence.

Our first result gives natural conditions under which the convergence (3.3) holds.

To begin with, we require the following compatibility condition between the tails of π and the underlying diffusion:

Assumption 3.

$$\int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))} \, \mathrm{d}y < \infty.$$

Assumption 3 is natural from a statistical point of view. Recall that without killing, the diffusion X has invariant density proportional to $\exp(2A)$ if this quantity is integrable (and certain regularity conditions hold; see Roberts and Tweedie [1996], Theorem 2.1). Assumption 3 can then be interpreted as requiring that the likelihood ratio $\pi(Y)/e^{2A(Y)}$ has finite variance when $Y \sim \exp(2A)$. This is what we would need to assume were we to target π by importance sampling from $\exp(2A)$.

In particular, Assumption 3 holds when the stronger 'rejection sampling' condition holds: that there exists some $M < \infty$ such that

$$\frac{\pi(y)}{\exp(2A(y))} < M \quad \forall y \in \mathbb{R}^d. \tag{3.4}$$

If $\exp(2A)$ is integrable, then this is precisely the condition that would allow us to sample from π using a rejection sampler with proposal density proportional to $\exp(2A)$. Informally, this demands that the asymptotic tail behavior of the diffusion be heavier than the tails of the target distribution. In particular, if the diffusion X is a Brownian motion on $\mathbb{R}^d$ ($A \equiv 0$ in (3.1)), Assumption 3 holds whenever the target density π is bounded.

We now define the appropriate killing rate κ, to be used to construct the killing time τ_∂ in (3.2). Define $\tilde{\kappa} : \mathbb{R}^d \to \mathbb{R}$ by

$$\tilde{\kappa}(y) := \frac{1}{2}\left(\frac{\Delta\pi}{\pi} - \frac{2\nabla A \cdot \nabla\pi}{\pi} - 2\Delta A\right)(y), \quad y \in \mathbb{R}^d \tag{3.5}$$

where Δ denotes the Laplacian operator. We require:

Assumption 4. *$\tilde{\kappa}$ is bounded below, and not identically zero.*

We will see that the correct killing rate is

$$\kappa = \tilde{\kappa} + K, \tag{3.6}$$

where $K := -\inf_{y\in\mathbb{R}^d} \tilde{\kappa}(y)$, chosen so that κ is non-negative everywhere. If $\tilde{\kappa}$ is identically zero, then there is no killing and we are in the familiar realm of stationary convergence of (unkilled) Markov processes; in fact, X will be a Langevin diffusion targeting π; see Roberts and Tweedie [1996]. To facilitate the development of intuition, some examples of κ in the case of $A \equiv 0$ are given in Section 3.1.4. Heuristically, this form for the killing rate makes π an eigenfunction for the generator of the killed diffusion, which corresponds to quasi-stationarity; see Section 3.3 for the mathematical details and further explanation.

The form of the untranslated killing rate in (3.5) also has the natural following interpretation. Writing $U := \log\pi$, which we can do since we are assuming π is positive, and as above thinking of $\exp(2A)$ as describing the asymptotic unkilled dynamics, we can rewrite (3.5) as

$$\tilde{\kappa}(y) = \frac{1}{2}\big(\Delta(U - 2A) + \nabla U \cdot \nabla(U - 2A)\big). \tag{3.7}$$

Written this way, we see $\tilde{\kappa}$ is a measure of the discrepancy between the derivatives of $\log\pi$ and $2A$, and Assumption 4 states that this discrepancy cannot be arbitrarily negative.

3.1.2 Convergence to quasi-stationarity

Theorem 3.1.1. *Suppose Assumptions 1, 2, 3 and 4 hold. Then X has quasi-limiting distribution π. That is, the convergence in* (3.3) *holds.*

Remarks

1. This significantly improves on Theorem 1 of Pollock et al. [2016]: their result only applied to killed Brownian motions, and their complicated condition on the tails of the target density has been removed. While Brownian motion — $A \equiv 0$ in (3.1) — is a natural choice of a 'proposal' diffusion, with developments in the exact simulation of diffusions, such as Beskos et al. [2006], there is potential to consider other diffusions as candidates. In Section 3.2 we consider an Ornstein–Uhlenbeck process targeting a Gaussian distribution.

2. We are not able to use the recent convergence results of Champagnat and Villemonais [2016]. Their approach is via minorisation-type conditions, which do not hold in our particular non-compact state space setting, and so we cannot apply their theorem on uniform exponential convergence.

3. Assumption 3 is in fact not a necessary condition. For example in Section 4.6 of Kolb and Steinsaltz [2012] the authors consider cases of low killing on $[0, \infty)$, where λ_0^κ, the bottom of the spectrum (in our case K; see Section 3.3.3), is not an eigenvalue in the $\mathcal{L}^2$ sense, but convergence to quasi-stationarity still occurs. Instead, the requirement is that the unkilled process be recurrent. In the context of quasi-stationary Monte Carlo methods, where we are free to choose the diffusion, Assumption 3 is a natural condition, since the excluded cases have zero spectral gap, hence inevitably poor convergence properties.

4. Theorem 3.1.1 also extends the results of Kolb and Steinsaltz [2012]: there the authors considered only (one-dimensional) cases where $\lim_{y\to\infty} \kappa(y) \neq \lambda_0^\kappa$. For example, our result gives convergence of killed Brownian motions with polynomially-tailed quasi-stationary distributions: in such cases $\lim_{\|y\|\to\infty} \kappa(y) = \lambda_0^\kappa$, but the conditions of Theorem 3.1.1 still hold, so we obtain convergence to quasi-stationarity.

5. We also obtain convergence of the conditional measures $\mathbb{P}_x(X_t \in \cdot\,|\tau_\partial > t)$ to π in total variation distance as $t \to \infty$, as shown in the proof of Theorem 7 of Tuominen and Tweedie [1979].

3.1.3 Rate of convergence

Our second result helps us to understand the rate of convergence to quasi-stationarity. Let $Z = (Z_t)_{t \geq 0}$ be the weak solution of the related SDE

$$\mathrm{d}Z_t = \frac{1}{2} \nabla \log \left(\frac{\pi^2}{\exp(2A)} \right) (Z_t) \, \mathrm{d}t + \mathrm{d}W_t, \tag{3.8}$$

with $Z_0 = x$. This is an example of a Langevin diffusion. Under suitable regularity conditions (see Theorem 2.1 of Roberts and Tweedie [1996]) the law of the diffusion Z_t converges to the distribution on $\mathbb{R}^d$ with Lebesgue density proportional to $\pi^2 / \exp(2A)$ as $t \to \infty$. (Assumption 3 guarantees that this is integrable.) Let $-L^Z$ denote the infinitesimal generator of this process and let $-L^\kappa$ denote the infinitesimal generator of the process (3.1) killed at rate κ. These operators will be constructed explicitly in Section 3.3.3 as self-adjoint operators on the appropriate $\mathcal{L}^2$ Hilbert spaces.

Writing $\gamma := \exp(2A)$, $\Gamma(\mathrm{d}x) := \gamma(x) \, \mathrm{d}x$ for the corresponding Borel measure on $\mathbb{R}^d$, which is the reversing measure of the diffusion X, and $\varphi := \pi / \gamma$, we have the following result.

Theorem 3.1.2. *Under the same conditions as Theorem 3.1.1 the $\mathcal{L}^2$ spectra of L^Z and L^κ agree, up to an additive constant. In particular, when L^Z has a spectral gap, the transition kernel of the killed process $p^\kappa(t, x, y)$ satisfies*

$$\left| e^{tK} p^\kappa(t, x, y) - \varphi(x) \varphi(y) \right| \leq C e^{-t(\lambda_1^Z - \lambda_0^Z)},$$

where $\lambda_1^Z > \lambda_0^Z = 0$ are the bottom two eigenvalues of the Langevin diffusion, and the constant C may depend on x and y. If the drift in (3.1) is bounded then C may be chosen independent of x and y.

If the measure Γ is such that $\Gamma(\mathbb{R}^d) < \infty$ then for an initial Γ-density $\psi \in \mathcal{L}^1(\Gamma) \cap \mathcal{L}^2(\Gamma)$,

$$\left| \mathbb{P}_\psi(X \in E \,|\, \tau_\partial > t) - \pi(E) \right| \leq C' e^{-t(\lambda_1^Z - \lambda_0^Z)},$$

for any measurable $E \subset \mathbb{R}^d$, where

$$C' = \frac{2 \left(\int \psi(x)^2 \, \mathrm{d}\Gamma(x) \right)^{1/2} \Gamma\left(\mathbb{R}^d\right)^{1/2}}{\int \psi(x) \pi(x) \, \mathrm{d}x \cdot \int \pi(x) \, \mathrm{d}x}.$$

The additive constant in Theorem 3.1.2 is K; that is, the spectrum of L^Z is the translation of the spectrum of L^κ by $+K$.

Theorem 3.1.2 tells us that the stationary convergence of the Langevin diffusion (3.8) and the quasi-stationary convergence of our killed diffusion occur at the same exponential rate, given by the equal spectral gaps. Since Langevin dynamics have been applied widely in computational statistics and the applied sciences, their rates of convergence have been studied extensively; see, for instance, the recent results of Dalalyan [2017] and Durmus and Moulines [2017]. Thus for many cases of π we will be able to accurately describe the rate of convergence in (3.3).

Theorem 3.1.2 also suggests that quasi-stationary Monte Carlo methods relying on (3.3) may converge relatively slowly for densities which are *multimodal*. In the case of $A \equiv 0$ (killed Brownian motion), if π is multimodal, then π^2 will typically be even more irregular, and the Langevin diffusion targeting π^2 will converge only gradually. On the other hand, quasi-stationary Monte Carlo methods should have good success targeting densities which are unimodal, such as logconcave densities. If π is unimodal, π^2 will be even more regular and have faster tail decay, leading to faster convergence of the Langevin diffusion. Such densities appear naturally in the context of Big-Data Bayesian inference. The Bernstein–von Mises theorem ([van der Vaart, 1998, Section 10.2]) tells us, for instance, that for large datasets the posterior distributions are approximately Gaussian.

Remarks

1. A sufficient condition for the existence of a spectral gap ($\lambda_1^\kappa > \lambda_0^\kappa$) is that

$$\liminf_{\|x\|\to\infty} \tilde{\kappa}(x) > 0. \tag{3.9}$$

   See for instance the proof of Lemma 3.3(v) of Kolb and Steinsaltz [2012], which carries over into our setting. Furthermore, if $\liminf_{\|x\|\to\infty} \tilde{\kappa}(x) = +\infty$ then this implies that the spectrum is purely discrete (the *essential spectrum* is empty). In the case of killed Brownian motion this holds for all exponentially-tailed densities of the form $\exp(-\beta\|x\|^\alpha)$ for some $\beta > 0, \alpha \geq 1$.

2. The Langevin diffusion in (3.8) is precisely the Q-process (the diffusion conditioned never to be killed) defined by the diffusion X and the killing time τ_∂. It is defined as the limit

$$\mathbb{Q}_x(A) := \lim_{T\to\infty} \mathbb{P}_x(A|T < \tau_\partial)$$

   for $A \in \sigma(X_s : s \leq t)$ for some $t \geq 0$.

3. Theorem 3.1.2 is a continuous state-space generalisation of Theorem 1 of Diaconis and Miclo [2014]: there the authors showed that in a finite state-space, rates of convergence to quasi-stationarity in total variation distance can be bounded above and below by constant multiples of the rates of convergence to stationarity in total variation of an appropriate unkilled process.

3.1.4 Examples of κ

In the simple and computationally important case of a killed Brownian motion ($A \equiv 0$ in (3.1)), $\tilde{\kappa}$ as defined in (3.5) simplifies down to

$$\tilde{\kappa}(y) = \frac{\Delta\pi}{2\pi}(y), \quad y \in \mathbb{R}^d.$$

In the following examples it can be easily checked that the conditions of Theorem 3.1.1 are satisfied.

- *Gaussian on $\mathbb{R}^d$.* Let $\sigma^2 > 0$ and $\pi(y) \propto \exp(-\|y\|^2/(2\sigma^2))$ for $y \in \mathbb{R}^d$, where throughout $\|\cdot\|$ denotes the Euclidean norm. Then straightforward calculation gives us that $\tilde{\kappa}(y) = \frac{1}{2}(\sigma^{-4}\|y\|^2 - \sigma^{-2}d)$ for $y \in \mathbb{R}^d$ and hence

  $$\kappa(y) = \frac{1}{2\sigma^4}\|y\|^2, \quad y \in \mathbb{R}^d.$$

  Since $\liminf_{\|y\|\to\infty} \tilde{\kappa}(y) > 0$ (in fact it's infinite), we expect exponential rates of convergence to quasi-stationarity, from condition (3.9). This example is considered in some detail in the case $d = 1$ in Section 3.2. This example also gives the independently interesting result that a Brownian motion on $\mathbb{R}^d$ killed at a quadratic rate will have a Gaussian quasi-limiting distribution.

- *Univariate exponential decay.* Consider a positive, smooth one-dimensional target density π with tail decay $\pi \propto \exp(-\beta|y|)$ for all y outside of a compact set $E \subset \mathbb{R}$, for some $\beta > 0$. We find that for all $y \in \mathbb{R}\backslash E$, $\tilde{\kappa}(y) = \beta^2$, that is, a positive constant. The killing rate κ will then also be constant asymptotically. By condition (3.9), we expect exponential convergence to quasi-stationarity.

- *Heavy-tailed case.* Consider a univariate Cauchy target, $\pi(y) \propto 1/(1+y^2)$ for $y \in \mathbb{R}$. Then simple calculation gives $\tilde{\kappa}(y) = \frac{3y^2-1}{(1+y^2)^2}$, for $y \in \mathbb{R}$ and then

  $$\kappa(y) = \frac{3y^2 - 1}{(1+y^2)^2} + 1, \quad y \in \mathbb{R}^d.$$

  We see here an example where $\liminf_{|y|\to\infty} \tilde{\kappa}(y) = 0$; the sufficient condition for a spectral gap (3.9) fails and we expect slower convergence.

3.2 Example: Ornstein–Uhlenbeck process targeting a Gaussian density

Before turning to the mathematical technicalities, we offer a mathematically tractable example that can be readily simulated: a killed Ornstein–Uhlenbeck process targeting a Gaussian distribution. For simplicity of presentation we discuss the univariate case $d = 1$. Analogous results hold in the multivariate case, but the notation is more cumbersome, and the calculations more involved.

Throughout this section, we write $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}, \sigma^2 > 0$ to denote the univariate Gaussian distribution with mean μ and variance σ^2.

In (3.1) we let $A(y) = -(\nu - y)^2/(4\tau^2)$ for each $y \in \mathbb{R}$, where $\nu \in \mathbb{R}, \tau^2 > 0$ are fixed. This defines a diffusion X as the weak solution of

$$\mathrm{d}X_t = \frac{1}{2\tau^2}(\nu - X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x. \tag{3.10}$$

The Ornstein–Uhlenbeck process X has a $\mathcal{N}(\nu, \tau^2)$ stationary distribution; the corresponding density function is proportional to $\exp(2A)$.

Fix $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, and let the target density be

$$\pi(y) = \exp\left\{ -\frac{1}{2\sigma^2}(y - \mu)^2 \right\}$$

for each $y \in \mathbb{R}$, the (unnormalised) density of a $\mathcal{N}(\mu, \sigma^2)$ random variable. We note that the regularity conditions — Assumptions 1 and 2 — hold.

The untranslated killing rate computed from (3.5) is for each $y \in \mathbb{R}$ given by

$$\tilde{\kappa}(y) = \frac{1}{2}\left(\frac{(y-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} + \frac{(\nu - y)(y - \mu)}{\tau^2\sigma^2} + \frac{1}{\tau^2} \right). \tag{3.11}$$

We now assume

$$\tau^2 > \sigma^2; \tag{3.12}$$

that is, the invariant distribution of the underlying diffusion has tails that are heavier than those of the target distribution. This makes the leading coefficient in the quadratic (3.11) positive, so that $\tilde{\kappa}$ is bounded below, meaning that Assumption 4 holds. In this case we will have a spectral gap (since the limit of the killing at infinity is $+\infty$; see (3.9), so we expect quasi-stationary convergence to occur at an exponential rate. Completing the square in (3.11) gives the minimum value

$$K := -\inf_{y \in \mathbb{R}} \tilde{\kappa}(y) = \frac{(\mu - \nu)^2}{8\tau^2(\tau^2 - \sigma^2)} + \frac{\tau^2 - \sigma^2}{2\tau^2\sigma^2}. \tag{3.13}$$

In Section 3.3.3 we will identify K with λ_0^κ, the bottom of the $\mathcal{L}^2$-spectrum of the generator of the killed diffusion, and so K is also the asymptotic rate of killing (see Lemma 4.2 of Kolb and Steinsaltz [2012]). We see from (3.13) that K is strictly positive, as general spectral theory predicts. Adding K to $\tilde{\kappa}$ and rearranging, we obtain the killing rate

$$\kappa(y) = \frac{\tau^2 - \sigma^2}{2\tau^2\sigma^4}\left(y - \left\{\frac{\mu+\nu}{2} + \frac{\tau^2}{\tau^2-\sigma^2}\left(\mu - \frac{\mu+\nu}{2}\right)\right\}\right)^2 \tag{3.14}$$

for $y \in \mathbb{R}$.

It remains to check Assumption 3. By direct calculation,

$$\frac{\pi^2}{\exp(2A)}(y) \propto \exp\left\{-\frac{1}{2}\frac{2\tau^2-\sigma^2}{\sigma^2\tau^2}\left(y - \frac{2\mu\tau^2 - \nu\sigma^2}{2\tau^2-\sigma^2}\right)^2\right\}.$$

Our assumption (3.12) guarantees this will be integrable, and in fact proportional to the density of the Gaussian distribution

$$\mathcal{N}\left(\frac{2\mu\tau^2-\nu\sigma^2}{2\tau^2-\sigma^2}, \frac{\sigma^2\tau^2}{2\tau^2-\sigma^2}\right). \tag{3.15}$$

So Theorem 3.1.1 allows us to conclude that π is the quasi-limiting distribution of our Ornstein–Uhlenbeck process (3.10) killed at rate (3.14), as long as (3.12) holds.

Since $\pi^2/\exp(2A)$ is the density of a Gaussian distribution, it follows that the corresponding Langevin diffusion (3.8) is another Ornstein–Uhlenbeck process, albeit with stationary distribution given by (3.15). In Metafune et al. [2002], the authors explicitly computed the $\mathcal{L}^p$ spectra of Ornstein–Uhlenbeck operators, and by applying their Theorem 3.1 we find that the $\mathcal{L}^2$ spectrum of L^Z is given by

$$\Sigma(L^Z) = \left\{\lambda_n^Z = \frac{n(2\tau^2-\sigma^2)}{2\sigma^2\tau^2} : n = 0, 1, 2, \dots\right\}.$$

By Theorem 3.1.2, this coincides (up to an additive constant) with the spectrum of our killed process (3.10). In particular, the spectral gap of our killed process is

$$\lambda_1^Z - \lambda_0^Z = \frac{2\tau^2-\sigma^2}{2\sigma^2\tau^2} = \frac{1}{\sigma^2} - \frac{1}{2\tau^2}\,.$$

For this example there are two mechanisms influencing the convergence to quasi-stationarity: the drift of the underlying diffusion (3.10), along with the killing (3.14) and subsequent conditioning on survival. It is interesting to note that the spectral gap is maximised when $\tau^2 \to \infty$, in which case the drift is 0. When in addition $\mu = \nu$, we see that the killing is also maximal, as measured by, say, the asymptotic killing rate

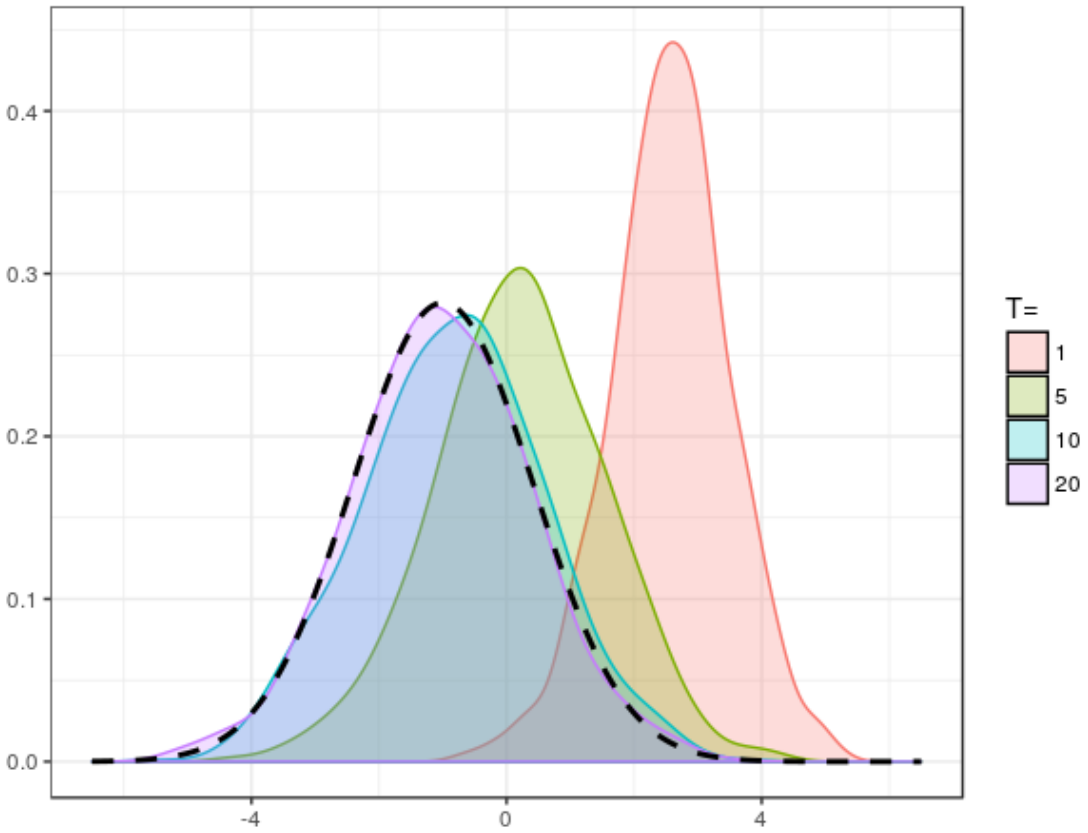


Figure 3.1: Estimates of the conditioned laws $\mathbb{P}_x(X_T \in \cdot | \tau_\partial > T)$ for the process (3.10) with parameters $\nu = 2, \tau^2 = 4$ started at $x = 3$ for various T. The dashed black line shows the quasi-stationary density, a Gaussian density with mean and variance $\mu = -1, \sigma^2 = 2$ respectively.

(3.13). This limit case $\tau^2 \to \infty$ corresponds to the case of killed Brownian motion ($A \equiv 0$ in (3.1)). This suggests that the rate of convergence to quasi-stationarity is determined more by the killing mechanism than by the underlying drift. However, depending on the method of implementation, a greater rate of killing could lead to reduced computational efficiency.

This simple example is amenable to simulation, as shown in Figure 3.1. The figure shows estimates of the conditional distributions $\mathbb{P}_x(X_T \in \cdot | \tau_\partial > T)$ for $T = 1, 5, 10, 20$ for the choices $\nu = 2, \tau^2 = 4, \mu = -1, \sigma^2 = 2$, and initial value $X_0 = x = 3$.

3.3 Mathematical preliminaries

3.3.1 Definitions

Fix $d \in \mathbb{N}$. Let ∇ denote the gradient operator; the d-dimensional vector with components $\nabla_i = \partial/\partial x_i$, $i = 1, \ldots, d$. We will denote the Laplacian operator by $\Delta := \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$. We are given functions $A : \mathbb{R}^d \to \mathbb{R}$ and $\pi : \mathbb{R}^d \to [0, \infty)$ that satisfy Assumptions 1, 2 and 3. For brevity we write

$$\gamma := \exp(2A).$$

π is our target density, which need not be normalised. In a slight abuse of notation we will also write π for the Borel probability measure on $\mathbb{R}^d$ with Lebesgue density proportional to π.

Let $C \equiv C([0, \infty), \mathbb{R}^d)$ denote the space of continuous functions mapping $[0, \infty) \to \mathbb{R}^d$, and let ω be a typical element. For each $t \geq 0$ let $X_t : C \to \mathbb{R}^d$ be the coordinate mapping $X_t(\omega) = \omega(t)$, and let $\mathcal{C} := \sigma(\{X_t : t \geq 0\})$ be the cylinder σ-algebra. For any $x \in \mathbb{R}^d$ let $\tilde{\mathbb{P}}_x$ be the measure on $(C, \mathcal{C})$ such that under $\tilde{\mathbb{P}}_x$, $X = (X_t)_{t \geq 0}$ is the weak solution to (3.1).

Define $\kappa : \mathbb{R}^d \to [0, \infty)$ by

$$\kappa(y) := \tilde{\kappa}(y) + K, \quad \forall y \in \mathbb{R}^d,$$

where $\tilde{\kappa}$, defined in (3.5), is required to satisfy Assumption 4, so that $K := -\inf_{y \in \mathbb{R}^d} \tilde{\kappa}(y)$ is finite. We augment our probability space to include an independent unit exponential random variable ξ, and define killing at rate κ as in (3.2), denoting this augmented space by $(\Omega, \mathcal{F}, \mathbb{P}_x)$.

We define $\mathcal{L}^2(\Gamma) \equiv \mathcal{L}^2(\mathbb{R}^d, \Gamma)$ to be the Hilbert space of (equivalence classes of) Borel-measurable square-integrable functions $f, g : \mathbb{R}^d \to \mathbb{R}$ with respect to the inner product

$$\langle f, g \rangle_{\mathcal{L}^2(\Gamma)} = \int_{\mathbb{R}^d} f(y) g(y) \, \mathrm{d}\Gamma(y)$$

where the measure Γ is given by $\mathrm{d}\Gamma(y) = \gamma(y) \, \mathrm{d}y$, with $\mathrm{d}y$ denoting Lebesgue measure on $\mathbb{R}^d$. We denote the corresponding norm by $\| \cdot \|_{\mathcal{L}^2(\Gamma)}$.

Define $\varphi : \mathbb{R}^d \to \mathbb{R}$ by

$$\varphi := \frac{\pi}{\exp(2A)},$$

which is smooth and positive. By construction we have that φ is integrable with respect to Γ:

$$\int_{\mathbb{R}^d} \varphi(y) \, \mathrm{d}\Gamma(y) = \int_{\mathbb{R}^d} \pi(y) \, \mathrm{d}y < \infty.$$

We will generally be working in the function space $\mathcal{L}^2(\Gamma)$, as this is the space on which the generator of the killed diffusion can be realised as a self-adjoint operator, which we will do explicitly in Section 3.3.3. As such, we will want consider densities with respect to Γ – rather than Lebesgue measure – and hence we will work with φ, rather than directly with π. Of course in the case of killed Brownian motion, $A \equiv 0$, π and φ coincide.

Following this line of thought, Assumption 3 states that indeed $\varphi \in \mathcal{L}^2(\Gamma)$:

$$\|\varphi\|^2_{\mathcal{L}^2(\Gamma)} = \int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))}\,\mathrm{d}y < \infty.$$

Without loss of generality we can rescale π so that this quantity is 1.

3.3.2 The killed Markov semigroup

Our results depend on the spectral theory of self-adjoint linear operators on Hilbert spaces. The proof of Theorem 3.1.1 avoids the heavy machinery of this theory by drawing on R. Tweedie's *R-theory*, which provides some of the results of operator theory most relevant to asymptotics of stochastic processes in a somewhat probabilistic package. We review the essential theory in Section 3.3.3, but it will be required only for the proof of Theorem 3.1.2.

The diffusion X killed at rate κ has a formal infinitesimal generator $-\tilde{L}^\kappa$ described by

$$\tilde{L}^\kappa = -\frac{1}{2\exp(2A)}\nabla \cdot \exp(2A)\nabla + \kappa = -\frac{1}{2}\Delta - \nabla A \cdot \nabla + \kappa. \tag{3.16}$$

Under Assumption 4, this formal differential operator can be realized as a positive self-adjoint operator L^κ on an $\mathcal{L}^2$ Hilbert space. It is this theory that we defer to Section 3.3.3.

Straightforward calculation shows that

$$\tilde{L}^\kappa \varphi = K\varphi. \tag{3.17}$$

So φ is an eigenfunction of the formal differential operator $\tilde{L}^\kappa$ with eigenvalue K.

Since we have assumed that φ is in $\mathcal{L}^2(\Gamma)$, we have that since κ is continuous and

not identically zero,

$$\begin{aligned} K\int \varphi(y)^2\,\mathrm{d}\Gamma(y) &= \int \varphi(y)\tilde{L}^{\kappa}\varphi(y)\,\mathrm{d}\Gamma(y) \\ &= -\frac{1}{2}\int |\nabla\varphi(x)|^2\,\mathrm{d}\Gamma(y) + \int \kappa(y)\varphi(y)^2\,\mathrm{d}\Gamma(y) \\ &\geq \int \kappa(y)\varphi(y)^2\,\mathrm{d}\Gamma(y) \\ &> 0. \end{aligned}$$

Thus we conclude that $K > 0$. These manipulations will be rigorously justified in subsequent sections when we formally construct the generator on the appropriate domain and establish that $\tilde{L}^{\kappa}$ acting on smooth compactly supported functions is essentially self-adjoint.

Recall from Tuominen and Tweedie [1979] that a finite non-negative measurable function f with $\int f(x)\,\mathrm{d}x > 0$ is said to be λ-*invariant* for a continuous-time semigroup $(P_t)_{t\geq 0}$ if for all $t > 0$,

$$f(x) = e^{\lambda t}P_t f(x) \quad \text{for almost every } x$$

and a σ-finite non-trivial measure ν is λ-*invariant* for continuous-time $(P_t)_{t\geq 0}$ if for all $t > 0$,

$$\nu(A) = e^{\lambda t}\nu P_t(A) \text{ for every measurable } A.$$

Analogous notions of R-invariance of functions and measures are similarly defined for discrete-time processes as well; the requirement $t > 0$ is replaced with $t \in \mathbb{N}$, and $e^{\lambda t}$ is replaced is replaced by R^t.

All we need for present purposes is the following lemma:

Lemma 3.3.1. *The sub-Markovian semigroup $\{P_t^{\kappa} : t \geq 0\}$ of the killed process X has a unique self-adjoint generator that is an extension of $-\tilde{L}^{\kappa}$ on smooth compactly supported functions. π is a λ-invariant measure for this semigroup, and φ a λ-invariant function, for $\lambda = K$.*

Except for some technical complications, which we will describe in the context of presenting the operator-theory framework in Section 3.3.3, this should be reasonably intuitive. We have already pointed out in (3.17) that φ is an eigenfunction of the generator with eigenvalue $-K$. Direct calculation shows that $\tilde{L}^{\kappa}$ is symmetric with respect to the measure Γ; that is, for $f, g \in \mathcal{L}^2(\Gamma)$ in the domain of $\tilde{L}^{\kappa}$ we have that

$$\langle \tilde{L}^{\kappa} f, g\rangle_{\mathcal{L}^2(\Gamma)} = \langle f, \tilde{L}^{\kappa} g\rangle_{\mathcal{L}^2(\Gamma)}.$$

Heuristically, since our assumptions ensure that the generator of the killed diffusion L^κ is symmetric, using (3.17) we obtain the following manipulations, for any non-negative test function $f \in \mathcal{L}^2(\Gamma)$,

$$\begin{aligned}\mathbb{E}_\pi[L^\kappa f(Y)] &= \int \pi(y) L^\kappa f(y)\,\mathrm{d}y = \int \varphi(y) L^\kappa f(y)\,\mathrm{d}\Gamma(y) \\ &= \int L^\kappa \varphi(y) f(y)\,\mathrm{d}\Gamma(y) = K \int \varphi(y) f(y)\,\mathrm{d}\Gamma(y) \\ &= K\mathbb{E}_\pi[f(Y)].\end{aligned}$$

Bearing in mind that L^κ is *minus* the generator of the killed diffusion, this shows that started in π the process will remain in π, except with a *mass loss at rate* K. That is to say, π is quasi-stationary. For an unkilled diffusion, if π were stationary, we would expect a similar expression to hold for any appropriate f, except with the right-hand side being exactly zero, reflecting the fact that the mass is preserved.

If we think of the adjoint operator — acting on measures — as acting on densities with respect to Γ, we have $(P_t^\kappa)^* g = (P_t^\kappa) g$. On the other hand, if g is a density with respect to Lebesgue measure the action is

$$P_t^\kappa g = \gamma P_t^\kappa (g/\gamma). \tag{3.18}$$

3.3.3 Operator theory

This section gives the mathematical background necessary for the proof of Theorem 3.1.2 in Section 3.5. Readers interested in the proof of Theorem 3.1.1 can move straight to Section 3.4.

Our operator $\tilde{L}^\kappa$ on $C_c^\infty(\mathbb{R}^d)$, smooth compactly supported functions, is a symmetric semibounded operator, and therefore has a self-adjoint extension, for instance the Friedrichs extension, see [Davies, 1995, Section 4.4]. As a matter of fact our operator is essentially self-adjoint – proven in Section 3.5.1 – and thus has a unique self-adjoint extension L^κ, so its completions are self-adjoint.

Recall that (3.16) describes the formal infinitesimal generator of our killed process. This gives rise to a closable densely-defined positive quadratic form $\tilde{q}^\kappa$ on $\mathcal{L}^2(\Gamma)$ given by

$$\tilde{q}^\kappa(f) = \frac{1}{2}\int_{\mathbb{R}^d} \nabla f \cdot \nabla f(y)\gamma(y)\,\mathrm{d}y + \int_{\mathbb{R}^d} \kappa(y)|f(y)|^2\gamma(y)\,\mathrm{d}y$$

for $f \in \mathcal{D}_\kappa$, where

$$\mathcal{D}_\kappa := \{f \in \mathcal{L}^2(\Gamma) : f \text{ continuously differentiable, } \tilde{q}^\kappa(f) < \infty\}.$$

We note that Assumption 4 is essential here. From a probabilistic point of view, we need $\tilde{\kappa}$ to be bounded below since a sensible killing rate must be non-negative (which amounts to putting a bound on the Radon–Nikodým derivative; see [Pollock et al., 2016, Appendix B]). From a functional-analytic point of view, we also need $\tilde{\kappa}$ to be bounded below since we require $\tilde{q}^\kappa$ to be closable. The semiboundedness assumption on $\tilde{\kappa}$ implies that for all compactly supported, twice differentiable $f \in C_c^2(\mathbb{R}^d)$, $\tilde{q}^\kappa(f)$ is a non-negative quadratic form associated to the symmetric operator $\tilde{L}^\kappa$. By Lemma 1.29, Assertion 2 of Ouhabaz [2005] we therefore conclude that the quadratic form $\tilde{q}^\kappa$ is closable.

Now let us denote the closure of $\tilde{q}^\kappa$ by q^κ. To this quadratic form there is associated a unique positive self-adjoint operator L^κ, with dense domain $\mathcal{D}(L^\kappa) \subset \mathcal{L}^2(\Gamma)$; see [Ouhabaz, 2005, Section 1.2.3]. For smooth compactly supported functions the action of L^κ is identical to that of $\tilde{L}^\kappa$.

Let $\Sigma(L^\kappa)$ denote the $\mathcal{L}^2(\Gamma)$-spectrum of L^κ. Since L^κ is self-adjoint and positive, we have that $\Sigma(L^\kappa) \subset [0, \infty)$. We have seen in (3.17) that $K \in \Sigma(L^\kappa)$; in particular $\Sigma(L^\kappa)$ is non-empty, so let us write λ_0^κ for the bottom of the spectrum. In fact, we have that $K = \lambda_0^\kappa$. This can be seen, for instance, using the arguments of Section 3.5.1, which are purely abstract and independent of this section, which shows that $\Sigma(L^\kappa)$ is identical to the spectrum of a positive self-adjoint operator translated by the constant K. Hence K must be the bottom of the spectrum.

It then follows from Theorem XIII.44 of [Reed and Simon, 1978, Section XIII.12], that λ_0^κ is a simple eigenvalue, with φ being its unique eigenfunction up to constant multiples. This is akin to a kind of Perron–Frobenius theorem on Hilbert spaces, *c.f.* Theorem 2.3.10.

We now make use of the spectral calculus for self-adjoint operators using projection-valued measures, as discussed in [Davies, 1995, Section 2.5]. This gives us the existence of a family of spectral projections $(E_\lambda^\kappa)_{\lambda \in [\lambda_0^\kappa, \infty)}$ and allows us to define $\phi(L^\kappa)$ for Borel-measurable $\phi : \mathbb{R} \to \mathbb{R}$, via

$$\begin{aligned}
\phi(L^\kappa) f &= \int_{\Sigma(L^\kappa)} \phi(\lambda) \, \mathrm{d}E_\lambda^\kappa f, \\
\mathcal{D}(\phi(L^\kappa)) &= \left\{ f \in \mathcal{L}^2(\Gamma) : \int_{\Sigma(L^\kappa)} |\phi(\lambda)|^2 \, \mathrm{d}\langle E_\lambda^\kappa f, f \rangle_{\mathcal{L}^2(\Gamma)} < \infty \right\}, \\
\|\phi(L^\kappa) f\|_{\mathcal{L}^2(\Gamma)}^2 &= \int_{\Sigma(L^\kappa)} |\phi(\lambda)|^2 \, \mathrm{d}\langle E_\lambda^\kappa f, f \rangle_{\mathcal{L}^2(\Gamma)} \, .
\end{aligned}$$

Now the Feynman–Kac representation states that for each $t > 0$

$$(e^{-tL^\kappa} f)(x) = \mathbb{E}_x[f(X_t)1_{\{\tau_\partial > t\}}]$$

for $f \in \mathcal{L}^2(\Gamma)$. Furthermore, for each $t > 0$ the operator e^{-tL^κ} is a contraction on $\mathcal{L}^2(\Gamma)$ (*cf.* the derivation in Demuth and van Casteren [2000]).

The spectral theorem allows us to write the diffusion semigroup as

$$P_t^\kappa f(x) = \mathbb{E}_x[f(X_t)1_{\{\tau_\partial > t\}}] = e^{-tL^\kappa} f(x) = \int_{\Sigma(L^\kappa)} e^{-t\lambda} \, \mathrm{d}E_\lambda^\kappa f(x)$$

for $f \in \mathcal{L}^2(\Gamma)$. The $(E_\lambda^\kappa)_{\lambda \in [\lambda_0^\kappa, \infty)}$ are orthogonal projections; in particular, $E_{\lambda_0^\kappa}^\kappa$ projects onto the span of φ. We can write

$$e^{-tL^\kappa} f = e^{-t\lambda_0^\kappa} \varphi \langle f, \varphi \rangle_{\mathcal{L}^2(\Gamma)} + \int_{\Sigma(L^\kappa)\setminus\{\lambda_0^\kappa\}} e^{-t\lambda} \, \mathrm{d}E_\lambda^\kappa f \, .$$

Thus

$$e^{t\lambda_0^\kappa} e^{-tL^\kappa} f = \varphi \langle f, \varphi \rangle_{\mathcal{L}^2(\Gamma)} + \int_{\Sigma(L^\kappa)\setminus\{\lambda_0^\kappa\}} e^{-t(\lambda - \lambda_0^\kappa)} \, \mathrm{d}E_\lambda^\kappa f \, . \tag{3.19}$$

For a given $f \in \mathcal{L}^2$, we are interested in the convergence to 0 of the integral term in (3.19). We note here that the convergence in this discussion is convergence in $\mathcal{L}^2(\Gamma)$. Ultimately we will be interested in convergence in $\mathcal{L}^1(\Gamma)$; we will return to this issue later.

We also want to rigorously check that $\varphi \in \mathcal{D}(L^\kappa)$. We know that $\varphi \in \mathcal{L}^2(\Gamma)$ and that φ is a formal eigenfunction of the formal differential operator $\tilde{L}^\kappa$ with eigenvalue K. Let us write $L_{\min}$ for the formal differential operator $\tilde{L}^\kappa$ acting on $\mathcal{D}(L_{\min}) = C_c^\infty(\mathbb{R}^d)$. We write $L_{\max}$ for the operator with domain

$$\mathcal{D}(L_{\max}) = \{f \in \mathcal{L}^2(\Gamma) : \tilde{L}^\kappa f \in \mathcal{L}^2(\Gamma)\}$$

where the derivatives are understood in the weak sense. Then for $f \in \mathcal{D}(L_{\max})$, $L_{\max} f = \tilde{L}^\kappa f$.

By definition of the adjoint $L_{\min}^*$,

$$\mathcal{D}(L_{\min}^*) = \{f \in \mathcal{L}^2(\Gamma) : \langle L_{\min} u, f \rangle = \langle u, g \rangle, \forall u \in C_c^\infty(\mathbb{R}^d), \text{ for some } g \in \mathcal{L}^2(\Gamma)\}.$$

Then by definition of the weak derivative, we see that $L_{\max} = L_{\min}^*$. (Note that here we are crucially using the fact that A is smooth – Assumption 2 – which ensures that for $u \in C_c^\infty(\mathbb{R}^d)$, $\gamma u \in C_c^\infty(\mathbb{R}^d)$ also.)

It follows from our arguments in Section 3.5 that $L_{\min}$ is essentially self-adjoint, hence it has a unique self-adjoint extension L^κ, which is equal to its closure $\bar{L}_{\min}$. This is identical to the operator arising from the quadratic form q^κ above.

We now argue that for any $f \in \mathcal{D}(L_{\max})$ — for instance our φ — we have that $f \in \mathcal{D}(\bar{L}_{\min}) = D(L^\kappa)$. This is because

$$L^*_{\min} = \bar{L}^*_{\min} = \bar{L}_{\min}.$$

The first equality follows from basic facts of closable operators and the second from essential self-adjointness. But we have seen that $L^*_{\min} = L_{\max}$ above. And so we conclude that $\bar{L}_{\min} = L_{\max}$.

This argument rigorously justifies our manipulations in Section 3.3.2 when showing that $K > 0$, since

$$\langle \varphi, L_{\max}\varphi \rangle = \langle \varphi, L^\kappa \varphi \rangle = q^\kappa(\varphi).$$

3.4 Proof of Theorem 3.1.1

We wish to apply the results of Tuominen and Tweedie [1979]. In order to do this we first need to check that $(P^\kappa_t)_{t\geq 0}$ is "simultaneously ϕ-irreducible" — that is, the resolvent kernel is strictly positive for discrete versions of the process discretised with respect to arbitrary time-steps. Ordinary ϕ-irreducibility holds for diffusions with smooth drift and locally bounded volatility by the Stroock–Varadhan Support Theorem, [Pinsky, 1995, Section 2.6]. Simultaneous ϕ-irreducibility follows then immediately from Theorem 1 of Tuominen and Tweedie [1979] since our process has a jointly continuous transition density with respect to the reversing measure; see Remark 1 after this proof.

We now show that $(P^\kappa_t)_{t\geq 0}$ is λ-positive, with $\lambda = K$, and that the K-invariant measure is precisely the target density π. This will then imply convergence to quasi-stationarity by an application of Theorem 7 of Tuominen and Tweedie [1979], which states that λ-positive processes, when $\lambda > 0$, exhibit quasi-limiting convergence as in (3.3), where the quasi-limiting distribution is the (unique) λ-invariant measure.

By Theorem 4(ii) of Tuominen and Tweedie [1979], showing $(P^\kappa_t)_{t\geq 0}$ is λ-positive is equivalent to showing that each (discrete-time) skeleton chain generated by P^κ_h, for any $h > 0$, is $e^{\lambda h}$-positive in the discrete-time sense, as defined in Tweedie [1974]. This involves showing that each skeleton chain is R-recurrent with $R = e^{\lambda h}$ and that the corresponding integral of the $e^{\lambda h}$-invariant function against the $e^{\lambda h}$-invariant measure is finite. So let us fix $h > 0$.

It follows from (3.17) and the Kolmogorov equations that

$$e^{hK} P^\kappa_h \varphi = \varphi.$$

This is exactly the definition of φ being e^{hK}-invariant for the discrete-time semigroup generated by P_h^κ. By (3.18) the measure π with Lebesgue density $\gamma\varphi$ is similarly e^{hK}-invariant for the discrete-time chain. (Definitions of λ-invariance are included in Section 3.3.2 for convenience.)

By Assumption 3,

$$\int_{\mathbb{R}^d} \varphi(y)\pi(\mathrm{d}y) = \int_{\mathbb{R}^d} \frac{\pi^2(y)}{\gamma(y)}\,\mathrm{d}y < \infty.$$

Thus by Proposition 3.1 and Proposition 4.3 of Tweedie [1974] the skeleton chain defined by operator P_h^κ, $(X_{nh})_{n=1}^\infty$, is R-recurrent, with $R = e^{hK}$. Theorem 7 of Tweedie [1974] then tells us that this skeleton chain is e^{hK}-positive. Since $h > 0$ was arbitrary, we obtain that $(P_t^\kappa)_{t\geq 0}$ is λ-positive, with $\lambda = K$. Theorem 4(iii) of Tuominen and Tweedie [1979] also tells us that φ and π are the unique K-invariant function and measure for (P_t^κ) respectively.

We are now in a position to utilise Theorem 7 of Tuominen and Tweedie [1979]. Since $K > 0$, killing happens almost surely, hence the key assumption (B) of Theorem 7 of Tuominen and Tweedie [1979] requires simply that $\int \pi(y)\,\mathrm{d}y < \infty$, which is certainly true. The conclusion of the theorem implies convergence to quasi-stationarity (3.3); that is, for any measurable $E \subset \mathbb{R}^d$ there is a set of starting points x of full Lebesgue measure such that

$$\lim_{t\to\infty} \mathbb{P}_x(X_t \in E \,|\tau_\partial > t) = \frac{\int_E \pi(y)\,\mathrm{d}y}{\int_{\mathbb{R}^d} \pi(y)\,\mathrm{d}y}.$$

In fact, this convergence holds for every starting point x. Since we have a continuous transition density $p^\kappa(t,x,y)$ (see Remark 1 after this proof), we have for any measurable set $E \subset \mathbb{R}^d$

$$\mathbb{P}_x(X_{t+1} \in E) = \int_{\mathbb{R}^d} p^\kappa(1,x,y)\mathbb{P}_y(X_t \in E)\,\mathrm{d}\Gamma(y).$$

Since we have convergence for y in some set of full measure, we obtain convergence for all $x \in \mathbb{R}^d$, which completes the desired result. □

Remarks

1. Assumption 3 can be interpreted in terms of spectral theory. It tells us that $\varphi \in \mathcal{L}^2(\mathbb{R}^d, \Gamma)$, so φ is also an eigenfunction of L^κ in the sense of $\mathcal{L}^2$ spectral theory. It is then possible to prove Theorem 3.1.1 analogously to Lemma 4.4 of Kolb and Steinsaltz [2012]. Following the derivation of Demuth and van

Casteren [2000], it follows that we have a continuous integral kernel $p^\kappa(t,x,y)$ with $\mathbb{E}_x[f(X_t)1_{\{\tau_\partial > t\}}] = \int p^\kappa(t,x,y)f(y)\,\mathrm{d}\Gamma(y)$. We can then apply Simon [1993] to see that $e^{t\lambda_0^\kappa}p^\kappa(t,x,y) \to c\varphi(x)\varphi(y)$ as $t \to \infty$, where $c = \|\varphi\|^{-2}_{\mathcal{L}^2(\Gamma)}$ and the proof of Theorem 3.1.1 can proceed analogously.

2. Our argument here relies fundamentally on self-adjointness of the operators and subsequent properties such as (3.17) , so there is no way we can circumvent the assumption of a gradient-form drift in (3.1). In one dimension this always holds, since we can simply take the integral of the drift function.

3.5 Rates of Convergence

Practitioners hoping to implement quasi-stationary Monte Carlo methods, such as the ScaLE Algorithm of Pollock et al. [2016], having been reassured that the procedure indeed converges to the correct distribution, will naturally inquire about the rate of convergence. Our result in this section draws heavily on the spectral theory for self-adjoint (unbounded) operators that we have outlined in Section 3.3.3.

When there is a *spectral gap*, that is, when $\lambda_1^\kappa > \lambda_0^\kappa$, the integral term will vanish at an exponential rate. Thus, it suffices to understand the spectrum $\Sigma(L^\kappa)$. To do this we will adapt an idea of Pinsky [2009], to transform our operator into one whose spectrum is already understood. Here it will be the infinitesimal generator of a certain Langevin diffusion.

3.5.1 Proof of Theorem 3.1.2

Consider the formal differential operator

$$\tilde{L}^{\tilde{\kappa}} = -\frac{1}{2\gamma}\nabla\cdot\gamma\nabla + \tilde{\kappa}$$

where $\tilde{\kappa}$ is defined in (3.5), acting on $C_c^\infty(\mathbb{R}^d)$, the set of smooth compactly-supported functions. This is very similar to the formal differential operator we began with in (3.16), differing only by an additive constant K, which will have the effect of merely translating the spectrum accordingly. $\tilde{L}^{\tilde{\kappa}}$ can be realised as a non-negative, self-adjoint operator $L^{\tilde{\kappa}}$ on $\mathcal{L}^2(\Gamma)$, by taking the Friedrichs extension of the appropriate quadratic form as before.

Now let $\mathcal{L}^2(\pi^2/\gamma) \equiv \mathcal{L}^2(\mathbb{R}^d, \pi^2/\gamma)$ denote the Hilbert space of (equivalence classes of) measurable functions $u, v : \mathbb{R}^d \to \mathbb{R}$ which are square-integrable with respect to

the inner product

$$\langle u, v \rangle_{\mathcal{L}^2(\pi^2/\gamma)} = \int_{\mathbb{R}^d} u(y)v(y)\frac{\pi^2(y)}{\gamma(y)}\,\mathrm{d}y.$$

The multiplication operator

$$Uf = \frac{\gamma}{\pi}f$$

is a bounded unitary transformation $U : \mathcal{L}^2(\Gamma) \to \mathcal{L}^2(\pi^2/\gamma)$, with inverse given by $U^{-1}u = \frac{\pi}{\gamma}u$.

We now define a second formal differential operator

$$\tilde{L}^Z = -\frac{1}{2}\Delta - \frac{1}{2}\nabla \log\left(\frac{\pi^2}{\gamma}\right) \cdot \nabla\ ,$$

which is minus the generator of the Langevin diffusion given in (3.8), targeting the density π^2/γ. $\tilde{L}^Z$ can similarly be realized as a positive, self-adjoint operator L^Z on $\mathcal{L}^2(\pi^2/\gamma)$. Our two formal operators are related through

$$\tilde{L}^{\tilde{\kappa}} = U^{-1}\tilde{L}^Z U.$$

We can also conjugate L^Z to obtain $U^{-1}L^Z U$, a self-adjoint operator on $\mathcal{L}^2(\Gamma)$.

Theorem 3.1.2 will be an immediate consequence if we show that in fact $U^{-1}L^Z U = L^{\tilde{\kappa}}$. This is the same as showing that the following diagram commutes:

$$\begin{array}{ccc} \tilde{L}^Z & \xrightarrow{\text{Friedrichs ext}} & L^Z \\ \Big\downarrow{\scriptstyle \text{Conjugate with } U} & & \Big\downarrow{\scriptstyle \text{Conjugate with } U} \\ \tilde{L}^{\tilde{\kappa}} & \xrightarrow{\text{Friedrichs ext}} & L^{\tilde{\kappa}} \end{array}$$

An operator is said to be *essentially self-adjoint* if it has a unique self-adjoint extension, which is given by the closure. From the background in Section 3.3.3 we see that the diagram commutes, and so Theorem 3.1.2 will follow, if we can show that $\tilde{L}^{\tilde{\kappa}}$ acting on $C_c^\infty(\mathbb{R}^d)$ is essentially self-adjoint. After all, the conjugate $U^{-1}L^Z U$ is a self-adjoint extension of $\tilde{L}^{\tilde{\kappa}}$; if the extension is unique it must be the same as $L^{\tilde{\kappa}}$.

We apply Theorem 2.13 of Braverman et al. [2002]. The smooth boundaryless manifold we are working in is simply $\mathbb{R}^d$, with smooth positive measure Γ. In their notation we take D to be $\frac{1}{\sqrt{2}}\nabla$, which is elliptic. The formal adjoint D^* is given by $-\frac{1}{\sqrt{2}}(\nabla \cdot + 2\nabla A\cdot)$. We set $V = \tilde{\kappa}$, and the resulting operator H_V is precisely $\tilde{L}^{\tilde{\kappa}}$.

The result follows immediately if V satisfies their Assumptions A and B, which ask for a decomposition of V into well-behaved non-negative parts and a mild technical

condition. Assumption A is immediate by writing

$$V = \underbrace{\tilde{\kappa} + K}_{V_+} + \underbrace{(-K)}_{V_-}$$

where clearly $V_+ \geq 0$ and $V_- \leq 0$. V_- trivially satisfies (ii) of Assumption A since it is constant.

Assumption B follows from their Theorem 2.3(ii), since our operator acts on scalar functions. The final condition of Theorem 2.13 is completeness of the metric g^{TM}, which is satisfied since it is equivalent to geodesic completeness of the manifold, which is true for $\mathbb{R}^d$.

Since unitary transformations leave spectra invariant it follows that the $\mathcal{L}^2(\Gamma)$ spectrum of $L^{\tilde{\kappa}}$ coincides with the $\mathcal{L}^2(\pi^2/\gamma)$ spectrum of L^Z, and hence the $\mathcal{L}^2$ spectra of $L^{\tilde{\kappa}}$ and L^Z coincide after translation by K.

We now would like to extend our proof of $\mathcal{L}^2$ convergence to $\mathcal{L}^1$ convergence in the case when there is a spectral gap. Let $\psi \in \mathcal{L}^1(\Gamma) \cap \mathcal{L}^2(\Gamma)$ be any initial density (with respect to the measure Γ). For the rest of this section, all norms and inner products will be with respect to $\mathcal{L}^2(\Gamma)$. Writing $\lambda_1^\kappa := \inf\left\{\Sigma(L^\kappa) \setminus \{\lambda_0^\kappa\}\right\}$, from our earlier results we have that

$$\begin{aligned}
\|e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi - \langle\psi,\varphi\rangle\varphi\|^2 &= \left\| \int_{\lambda_1^\kappa}^\infty e^{-t(\lambda-\lambda_0^\kappa)}\, \mathrm{d}E_\lambda^\kappa \psi \right\|^2 \\
&= \int_{\lambda_1^\kappa}^\infty e^{-t\cdot 2(\lambda-\lambda_0^\kappa)}\, \mathrm{d}\langle E_\lambda^\kappa\psi, \psi\rangle \\
&\leq \|\psi\|^2 \cdot e^{-t\cdot 2(\lambda_1^\kappa - \lambda_0^\kappa)}. \qquad (3.20)
\end{aligned}$$

We now link this to $\mathcal{L}^1$ convergence. Let $H \subset \mathbb{R}^d$ be a compact set. From the Cauchy–Schwarz inequality we know that

$$\begin{aligned}
\int_H |e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi(y) - \langle\psi,\varphi\rangle\varphi(y)|\, \mathrm{d}\Gamma(y) & \\
&\leq \|e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi - \langle\psi,\varphi\rangle\varphi\| \cdot \Gamma(H)^{1/2} \\
&\leq \|\psi\| \cdot \Gamma(H)^{1/2} \cdot e^{-t(\lambda_1^\kappa - \lambda_0^\kappa)}.
\end{aligned}$$

So we have the appropriate convergence in $\mathcal{L}^1(\Gamma)$ on compact sets. We could similarly obtain convergence for test functions $f \in \mathcal{L}^2(\Gamma)$, that is

$$\left| \left\langle e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi, f\right\rangle - \langle\psi,\varphi\rangle\langle\varphi, f\rangle \right| \leq \|\psi\| \cdot \|f\| \cdot e^{-t(\lambda_1^\kappa - \lambda_0^\kappa)}. \qquad (3.21)$$

We see that when Γ is a finite measure, we will obtain $\mathcal{L}^1$ convergence at this rate on all measurable sets, not just compact ones. This is the case when the (unkilled) diffusion has a strong inward drift.

Now assume that $\Gamma(\mathbb{R}^d) < \infty$ and fix some $E \subset \mathbb{R}^d$. Writing $\mathbb{P}_\psi$ for the law of the killed process starting from ψ, we have (recalling that $\int_{\mathbb{R}^d} \pi(x)\,\mathrm{d}x = \langle \varphi, 1\rangle$),

$$\begin{aligned}\left|\mathbb{P}_\psi(X \in E \,|\, \tau_\partial > t) - \pi(E)\right| &= \left| \int_E \left(\frac{e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi(y)}{e^{t\lambda_0^\kappa}\mathbb{P}_\psi(\tau_\partial > t)} - \frac{\varphi(y)}{\int_{\mathbb{R}^d}\pi(x)\,\mathrm{d}x}\right) \mathrm{d}\Gamma(y)\right| \\ &= \frac{1}{\langle\psi,\varphi\rangle \int \pi(x)dx}\left| \int_E \left(e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi(y) - \langle\psi,\varphi\rangle\varphi(y)\right) \mathrm{d}\Gamma(y)\right. \\ &\quad \left. - \left(\int_E e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi(y)\,\mathrm{d}\Gamma(y)\right)\left(\frac{e^{t\lambda_0^\kappa}\mathbb{P}_\psi(\tau_\partial > t) - \langle\psi,\varphi\rangle\int\pi(x)dx}{e^{t\lambda_0^\kappa}\mathbb{P}_\psi(\tau_\partial > t)}\right)\right|.\end{aligned}$$

Note that

$$\frac{\int_E e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi(y)\,\mathrm{d}\Gamma(y)}{e^{t\lambda_0^\kappa}\mathbb{P}_\psi(\tau_\partial > t)} = \frac{\langle e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi, \mathbf{1}_E\rangle}{\langle e^{t\lambda_0^\kappa} e^{-tL^\kappa}\psi, \mathbf{1}\rangle} \le 1,$$

so

$$\left|\mathbb{P}_\psi(X \in E \,|\, \tau_\partial > t) - \pi(E)\right| \le \frac{2\|\psi\|\Gamma\left(\mathbb{R}^d\right)^{1/2}}{\langle\psi,\varphi\rangle\int\pi(x)dx} e^{-t(\lambda_1^\kappa - \lambda_0^\kappa)}. \tag{3.22}$$

It remains to derive the rate of pointwise convergence for

$$e^{t\lambda_0^\kappa} p^\kappa(t,x,y) \to \varphi(x)\varphi(y) \quad \text{as } t \to \infty.$$

This argument does not require us to assume $\Gamma(\mathbb{R}^d) < \infty$. Following the approach of Simon [1993], for $x, y \in \mathbb{R}^d$ let us write $g_x(y) := e^{\lambda_0^\kappa} p^\kappa(1,x,y)$. First note that $g_x \in \mathcal{L}^2(\Gamma)$:

$$\begin{aligned}\|g_x\|^2_{\mathcal{L}^2(\Gamma)} &= e^{2\lambda_0^\kappa} \int p^\kappa(1,x,y) p^\kappa(1,x,y)\ \mathrm{d}\Gamma(y) \\ &= e^{2\lambda_0^\kappa} \int p^\kappa(1,x,y) p^\kappa(1,y,x)\ \mathrm{d}\Gamma(y) \\ &= e^{2\lambda_0^\kappa} p^\kappa(2,x,x) \\ &< \infty,\end{aligned}$$

using symmetry and the semigroup property. By the invariance of φ,

$$\langle g_x, \varphi\rangle = e^{\lambda_0^\kappa} \int p^\kappa(1,x,z)\varphi(z)\,\mathrm{d}\Gamma(z) = \varphi(x).$$

Now for $t > 2, x, y \in \mathbb{R}^d$

$$\begin{aligned}e^{t\lambda_0^\kappa} p^\kappa(t,x,y) &= \int\int g_x(z) e^{\lambda_0^\kappa(t-2)} p^\kappa(t-2,z,w) g_y(w)\ \mathrm{d}\Gamma(z)\,\mathrm{d}\Gamma(w) \\ &= \left\langle e^{-(t-2)L^\kappa} e^{(t-2)\lambda_0^\kappa} g_x\ ,\ g_y\right\rangle.\end{aligned}$$

By (3.21) this converges to

$$\langle g_x, \varphi\rangle\langle g_y, \varphi\rangle = \varphi(x)\varphi(y),$$

with rate given by

$$\left|e^{t\lambda_0^\kappa}p^\kappa(t,x,y)-\varphi(x)\varphi(y)\right| \leq e^{2\lambda_0^\kappa}\left(p^\kappa(2,x,x)p^\kappa(2,y,y)\right)^{1/2}e^{-t(\lambda_1^\kappa-\lambda_0^\kappa)}. \tag{3.23}$$

If the drift is bounded then the transition density is bounded as well, so this is bounded by $Ce^{-t(\lambda_1^\kappa-\lambda_0^\kappa)}$ for a universal constant C. □

3.6 Discussion

In this chapter we have proven natural sufficient conditions for the quasi-limiting distribution of a diffusion of the form (3.1) killed at an appropriate state-dependent rate to coincide with a target density π. We have also quantified the rate of convergence to quasi-stationarity by relating the rate of this convergence to the rate of convergence to stationarity of a related unkilled process.

As mentioned in the introduction, this framework is foundational for the recently-developed class of quasi-stationary Monte Carlo algorithms to sample from Bayesian posterior distributions, introduced in Pollock et al. [2016]. This framework promises improvement over more traditional MCMC approaches particularly for Bayesian inference on large datasets, since the killed diffusion framework enables the use of *subsampling* techniques. As detailed in [Pollock et al., 2016, Section 4], these allow the construction of estimators which scale exceptionally well as the size of the underlying dataset grows.

Quasi-stationary Monte Carlo methods are likely to be particularly effective compared to established Monte Carlo methods for Bayesian inference for *tall data*; that is, where parameter spaces have moderate dimension (allowing diffusion simulation to be feasible) but where data sizes are high. This includes the 'Big-Data' context where data size is so large it cannot even be stored locally on computers implementing the algorithm. This is because subsampling can take place 'offline' with only the subsets being stored locally. This adds significantly to the potential applicability of quasi-stationary Monte Carlo methods.

Our approach in this present work is also slightly more general than that of Pollock et al. [2016] in that we allow for a non-zero drift term in our diffusion (3.1). This raises the question of how to select among several possible drift functions the one that results in the most practical computational outcomes? While a detailed answer to this question is beyond the scope of this present work, we suggest the following guidelines. There is a critical trade-off between overall killing and the essential rate of convergence described in (3.3). As mentioned in the example of Section 3.2, higher

rates of killing will tend to *increase* the essential rate of convergence, while increasing the computational burden imposed by simulating killing events. Depending on the implementation details, this trade-off could go either way in terms of optimality. When scalable estimators for the killing events are available, such as in Pollock et al. [2016], it would be sensible to choose a drift that makes the killing rate high, for instance choosing a Brownian motion, so $A \equiv 0$. Of course, any Gaussian process allows straightforward simulation of the unkilled dynamics, and the choice of Brownian motion also simplifies Assumptions 3 and 4. Formally answering this question of the choice of drift would be an interesting avenue for future exploration.

We comment briefly now on some of our assumptions. Assumption 3 is generally straightforward to verify, especially in light of the stronger 'rejection sampling' formulation in (3.4). For instance, if A is uniformly bounded below then Assumption 3 holds if π is a bounded density function.

Assumption 4 is generally the most challenging. When $A \equiv 0$ this is mostly straightforward to verify, especially since densities on $\mathbb{R}^d$ are often convex in the tails. Verifying Assumption 4 in general can be done using the equivalent expression for $\tilde{\kappa}$ in (3.7), by comparing the decay of derivatives $2A$ with those of $\log \pi$. Indeed, ensuring a practically useful form of $\tilde{\kappa}$ – so that verification of Assumption 4 is straightforward – could influence the choice of A in the first place.

In practice, Assumption 4 also involves computing a lower bound for $\tilde{\kappa}$. It is actually not necessary to compute the precise value of $\inf_{y \in \mathbb{R}^d} \tilde{\kappa}(y)$; our results still hold if K in (3.6) is replaced by any constant such that the resulting κ is non-negative everywhere. Intuitively, taking a larger constant K amounts to merely adding additional killing events according to a homogeneous, independent Poisson process.

Depending on the choices of π and A, $\tilde{\kappa}$ can be a convex function in the tails, even in cases of non-zero A, as in our example of Section 3.2. A precise recipe for computing K in general is currently unavailable; readers interested in these more implementational details are encouraged to look at Pollock et al. [2016].

We conclude this discussion by indicating some potential future directions. As mentioned above, there are important questions of how to choose the underlying diffusion to optimize the computation for a given target density. One could also consider extensions of this work to entirely different underlying processes, such as jump diffusions or Lévy processes. Finally, another potential question is the exploration of alternative approaches to that described in Pollock et al. [2016] for the simulation of quasi-stationary distributions, such as the stochastic approximation approaches as

discussed in Blanchet et al. [2016] and Benaïm et al. [2018]. This is the topic of Chapter 4.

Chapter 4

An approximation scheme for quasi-stationary distributions of killed diffusions

4.1 Introduction

In this chapter we consider a stochastic approximation algorithm to approximate the quasi-stationary distribution of a killed diffusion on a compact manifold. This is motivated by one particular QSMC method: Regenerating ScaLE (ReScaLE), whose computational properties were explored in the work of Kumar [2019].

The work of this chapter has recently been accepted for publication in *Stochastic Processes and their Applications* as Wang et al. [2019b]. This work was done in collaboration with my supervisors Prof. Gareth Roberts and Prof. David Steinsaltz.

4.2 Introduction

In this work we assume that we are working on a compact, boundaryless, connected, d-dimensional smooth Riemannian manifold M, as in Benaïm et al. [2002]. We have a particle $X = (X_t)_{t\geq 0}$ evolving on M in continuous time, according to the solution to the stochastic differential equation (SDE)

$$\mathrm{d}Y_t = \nabla A(Y_t)\,\mathrm{d}t + \mathrm{d}W_t \tag{4.1}$$

between regeneration events, where $A : M \to \mathbb{R}$ is a smooth function and W is a standard Brownian motion on M. Regeneration events occur at a state-dependent rate $\kappa(X_{t-})$, where $\kappa : M \to [0, \infty)$ is a given smooth, positive function, which we will refer to as the *killing rate*. At a regeneration event, say at time T, the particle is

instantaneously 'killed' and 'reborn': its location is drawn (independently) from its normalized weighted empirical occupation measure μ_T, where for all $t \geq 0$, μ_t is given by

$$\mu_t = \frac{r\mu_0}{r + \int_0^t \eta_s \, \mathrm{d}s} + \frac{\int_0^t \eta_s \delta_{X_{s-}} \, \mathrm{d}s}{r + \int_0^t \eta_s \, \mathrm{d}s}$$

where $r > 0$, $\eta_\cdot : \mathbb{R}_+ \to \mathbb{R}_+$ and a probability measure μ_0 on M are fixed. The resulting process X is clearly non-Markovian. The addition of the μ_0 term has the benefit of regularising the μ_t around $t = 0$, as well as providing practical flexibility for the resulting Monte Carlo algorithm. The weight function $\eta_\cdot$ similarly provides additional practical flexibility; a straightforward choice would be constant $\eta_s \equiv 1$, but see Remark 4.3.2 for a nonconstant alternative.

4.2.1 Main results

The goal of this chapter is to characterize the asymptotic behavior of the measure-valued process $(\mu_t)_{t\geq 0}$, and to show that it converges to the quasi-stationary distribution for the original killed process. We proceed in two steps, following the "ODE method" (*cf.* Benaïm [1999], Kushner and Yin [2003]), which has been used to prove convergence of similar reinforced processes, for instance in Benaïm and Cloez [2015], Benaïm et al. [2018, 2002], Kurtzmann [2010]. The ODE method proceeds by two key steps. First, showing that a certain deterministic semiflow Φ converges to the appropriate limit, in our case the quasi-stationary distribution π. Second, showing that, following a suitable deterministic time change $\zeta_t := \mu_{h(t)}$, the stochastic evolution of the measures $(\zeta_t)_{t\geq 0}$ shadows Φ in an appropriate sense, to be defined below. From these two properties almost-sure convergence of μ_t to the quasi-stationary distribution can be deduced.

The present chapter extends previous related work in considering a continuous-time diffusive process on a compact manifold (rather than, say, a discrete-time Markov chain), which experiences 'soft killing' according to a smooth state-dependent killing rate (rather than instantaneous 'hard killing' at a boundary). In this setting of soft killing, a generic killing time τ_∂ has the form

$$\tau_\partial = \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s) \, \mathrm{d}s \geq \xi \right\}, \tag{4.2}$$

where $\xi \sim \mathrm{Exp}(1)$ is independent of the process X, as seen previously, rather than the first hitting time of some forbidden set of states as in the hard killing case.

Theorem 4.2.1. *Under Assumptions 5, 6, 7 and 8 detailed in Section 4.3.1, we have with probability 1, that for each $T > 0$,*

$$\lim_{t\to\infty} \sup_{s\in[0,T]} d_w\big(\zeta_{t+s}, \Phi_s(\zeta_t)\big) = 0 \tag{4.3}$$

where d_w is a metric that metrises weak- convergence of probability measures given in* (4.4) *and Φ is the semiflow defined in Section 4.4.*

Remark 4.2.2. In Benaïm [1999], a map $t \mapsto \zeta_t$ that satisfies (4.3) for each $T > 0$ is called an *asymptotic pseudo-trajectory* for the measure-valued semiflow Φ.

In Section 4.4 we define the semiflow Φ, and proceed to show that it has a global attractor π, which is the unique quasi-stationary distribution of the diffusion (4.1) killed at rate κ.

In particular, Theorem 4.2.1 leads to the following corollary:

Corollary 4.2.3. *Under the conditions of Theorem 4.2.1, we have that almost surely, $\lim_{t\to\infty} \mu_t = \pi$ in the sense of weak-* convergence.*

That is, we have $\lim_{t\to\infty} \mu_t(f) = \pi(f)$ for any continuous $f : M \to \mathbb{R}$.

While our results above hold for any appropriate given killing rate which satisfies our assumptions, in the ReScaLE algorithm the killing rate κ is chosen so that the quasi-stationary distribution π *equals* the Bayesian posterior distribution of interest; see the expression (3.6), *cf.* the work of Chapter 3. Corollary 4.2.3 tells us that in this setting we can draw approximate samples from π by running the regenerating process X and outputting μ_t for a large t as a proxy for π.

Figure 4.1 shows the output of two independent simulations on the unit circle $M = \mathbb{R}/(2\pi\mathbb{Z})$ parameterised by $\theta \in [0, 2\pi)$, which is amenable to straightforward simulation and visualisation. The underlying diffusion is a Brownian motion ($A \equiv 0$ in (4.1)), and the quasi-stationary distribution is trimodal with density function $\pi(\theta) = (0.3+\sin^2(1.5\theta))/(1.6\pi)$ with respect to the Riemannian measure. The corresponding killing rate, calculated using (3.6) is $\kappa(\theta) = 2.25(2\cos^2(1.5\theta) - 1)/(0.3 + \sin^2(1.5\theta)) + K$, with $K = 1.75$.

The simulations were done using a simple Euler discretisation scheme where time was discretised into intervals of length 0.05. The plots are oriented so that $\theta = 0$ is due east, with $\theta = \pi/2$ being due north, and so on. We have chosen $r = 1000$, μ_0 to be the uniform distribution over M and $\eta_s \equiv 1$. The plots shows $\int_0^t \delta_{X_s-} \,\mathrm{d}s/t$ for $t = 25, 100, 1000, 10^6$, split into 50 evenly-spaced bins. The quasi-stationary distribution π is the dashed line.

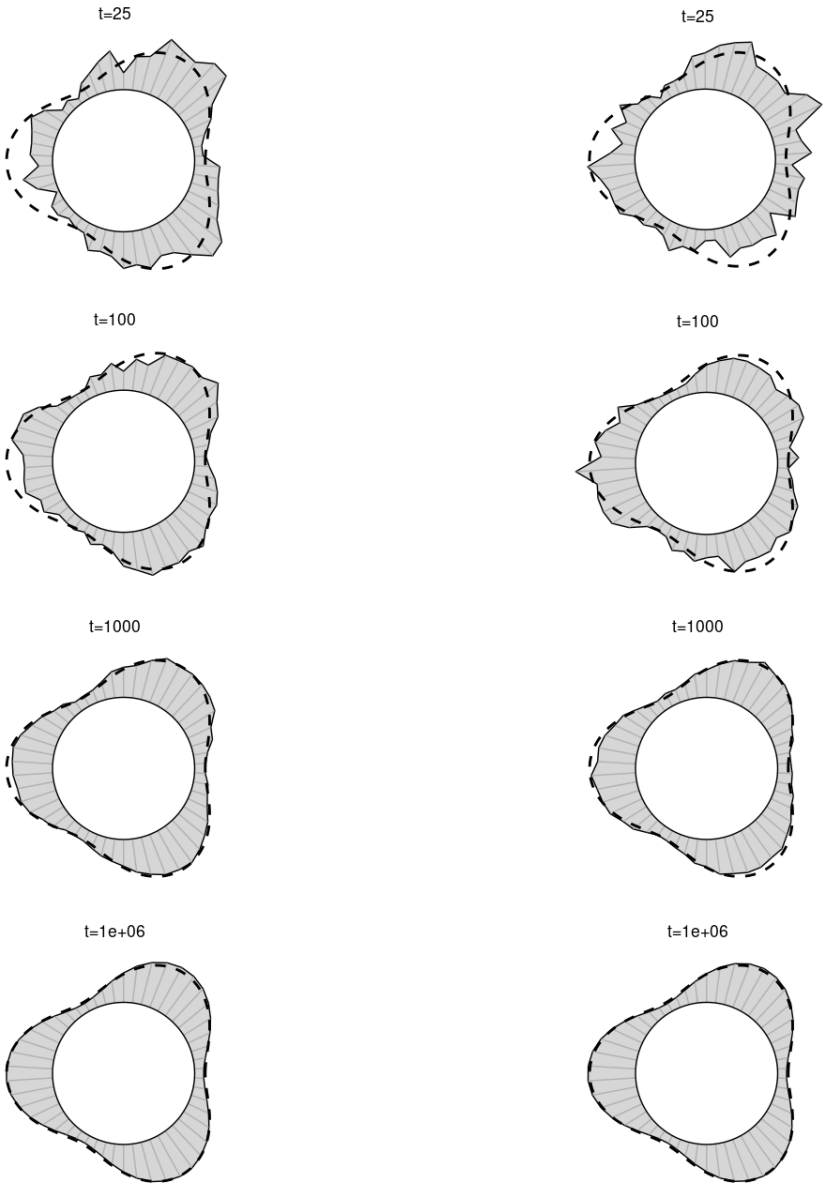


Figure 4.1: Example on the circle. The circle is parameterised by $\theta \in [0, 2\pi)$, with due east corresponding to $\theta = 0$. The underlying diffusion is a Brownian motion, and we have chosen $r = 1000$, μ_0 to be the uniform distribution on M and $\eta_s \equiv 1$. The quasi-stationary distribution $\pi(\theta) = (0.3 + \sin^2(1.5\theta))/(1.6\pi)$ is the dashed line. We have plotted $\int_0^t \delta_{X_s-} \, \mathrm{d}s/t$ discretised into 50 evenly-spaced bins, for $t = 25, 100, 1000, 10^6$. The two columns are two independent runs.

We see that there is a significant amount of variability for the small time values, as the initial rebirths are drawn mostly from the uniform distribution μ_0. These discrepancies are largely smoothed out by $t = 1000$ and certainly by $t = 10^6$, $\int_0^t \delta_{X_s-} \, \mathrm{d}s/t$ closely resembles the quasi-stationary distribution π.

This present work closely follows the approaches of Benaïm et al. [2002] and the recent Benaïm et al. [2018] in order to prove our limiting result Theorem 4.2.1. Benaïm et al. [2002] shows an analogous characterization for the normalized occupation measures of a self-interacting diffusion on a compact space, where the empirical occupation measure influences the present behavior of the diffusion through its drift term. In our work, the influence of the occupation measures is felt through jumps, which occur at a state-dependent rate. In Benaïm et al. [2018], the authors prove a discrete-time analogue of our above result; their underlying Markov process is a discrete-time Markov chain evolving on a compact space, rather than a diffusion. In [Benaïm et al., 2018, Section 8.3], the authors do suggest a continuous-time extension of their work: a diffusive process that is killed instantaneously when hitting the boundary of an open set. This is not the same as the present work; we are assuming a boundaryless manifold, and instead of hard killing at a boundary, killing occurs at a smooth state-dependent rate κ as in (4.2). The key difference in the proposed setting of [Benaïm et al., 2018, Section 8.3] is that in their case, the state space is no longer compact, and hence additional arguments ensuring almost sure tightness of the empirical occupation measures are needed.

We follow generally the path mapped out by Benaïm et al. [2002, 2018], Kurtzmann [2010], often referred to as the "ODE method", *cf.* Benaïm [1999], Kushner and Yin [2003]. To understand the particular continuous-time dynamics we employ techniques similar to those of Wang et al. [2019a], Kolb and Steinsaltz [2012]. For example, to handle the killing mechanism we make use of the transition subdensity of the killed diffusion and the corresponding resolvent operator; see Lemmas 4.3.4, 4.3.5 and 4.3.8. The continuous-time setting also provides a natural interpretation of the weights η_s in terms of the distribution of the rebirth 'times'; see Remark 4.3.2. Another contribution of this work is the analysis of the deterministic semiflow in Section 4.4, where we transform to an auxiliary Markov process and apply an appropriate drift condition.

The earliest version of a similar analysis that we are aware of is the convergence proof of Aldous et al. [1988] for the finite-state-space discrete-time setting. Benaïm and Cloez [2015] and Blanchet et al. [2016] expanded this work to derive rates of convergence and central limit theorems. These analyses rely upon special techniques,

such as those in Kushner and Yin [2003], applicable only to finite-dimensional probability distributions.

The structure of the chapter is as follows. In Section 4.3 we lay out the mathematical setting and describe the assumptions that we are making. We also define the fixed rebirth process and prove some of its key properties which are crucial for later defining the semiflow. In Section 4.4 we define and analyse the deterministic semiflow Φ and prove that π is a global attractor. In Section 4.5 we prove that our normalized weighted occupation measures almost surely comprise an asymptotic pseudo-trajectory for Φ, concluding the convergence proof.

4.3 Preliminaries

4.3.1 General background

We first describe our assumptions and some notation. These assumptions are *assumed to hold throughout the chapter.*

Assumption 5. *M is a d-dimensional boundaryless C^∞ compact connected Riemannian manifold.*

We will denote the corresponding Riemannian measure by $\mathrm{d}x$ or $\mathrm{d}y$, and the Riemannian inner product of points x and y by $x \cdot y$. Let $C(M)$ denote the Banach space of (bounded) real-valued continuous functions on M equipped with the sup norm, $\|\cdot\|_\infty$. Let $\mathcal{P}(M)$ denote the space of Borel probability measures on M, equipped with the topology of weak-* convergence, namely convergence along all bounded continuous test functions. (This is conventionally called simply weak convergence in probability theory, but we follow the terminology of Benaïm et al. [2018] to avoid confusion when working with the Banach space $C(M)$ and its dual.) Thus $\mathcal{P}(M)$ is a compact metrisable space, by the Prokhorov theorem, *cf.* [Dudley, 2002, Theorem 11.5.4].

As M is compact, $C(M)$ is separable, so we may choose a sequence of smooth functions $f_1, f_2, \ldots$ which are dense in $\{f \in C(M) : \|f\|_\infty \leq 1\}$. We define the metric $d_w : \mathcal{P}(M) \times \mathcal{P}(M) \to \mathbb{R}_+$ on $\mathcal{P}(M)$ by

$$d_w(\nu_1, \nu_2) = \sum_{i=1}^{\infty} \frac{1}{2^i} |\nu_1(f_i) - \nu_2(f_i)|. \tag{4.4}$$

We also define the total-variation norm for a signed measure μ on M by

$$\|\mu\|_1 = \sup\{|\mu(f)| : f \in C(M), \|f\|_\infty \leq 1\}.$$

Let Ω be the Skorokhod space of càdlàg paths $\omega : \mathbb{R}_+ \to M$, and let $\mathcal{F}$ be the cylinder σ-algebra. Let $X = (X_t)_{t\geq 0}$ be the coordinate process, $X_t(\omega) = \omega(t)$, and let $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration of X.

Assumption 6. *The function $A : M \to \mathbb{R}$ is C^∞.*

In particular, this implies that A is bounded (since M is compact), and that the SDE (4.1) on M has a unique strong solution for any initial position.

Given a measurable killing *rate* $\kappa : M \to [0, \infty)$, the corresponding killing time τ_∂ is defined as

$$\tau_\partial := \inf\left\{t \geq 0 : \int_0^t \kappa(Y_s)\,\mathrm{d}s \geq \xi\right\},$$

where ξ is an independent exponential random variable with rate 1.

Assumption 7. *The killing rate $\kappa : M \to [0, \infty)$ is C^∞ and is uniformly bounded away from zero: there exists some constant $\underline{\kappa} > 0$ such that*

$$0 < \underline{\kappa} \leq \kappa(x) \quad \forall x \in M. \tag{4.5}$$

As a continuous function on a compact space κ is necessarily bounded above, say

$$\kappa(x) \leq \bar{\kappa} < \infty \quad \forall x \in M.$$

Given a killing rate κ which is not strictly bounded away from 0, we can always add a positive constant everywhere; this will not affect the quasi-stationary behavior of the process and will ensure that (4.5) holds. The upper bound on the killing rate will certainly guarantee that the resulting process will be almost surely nonexplosive.

We will later see that given κ, the diffusion (4.1) killed at rate κ has a unique quasi-stationary distribution (Proposition 4.3.12).

The question of existence and uniqueness of quasi-stationary distributions of killed diffusions has been studied in depth, for instance in Kolb and Steinsaltz [2012], [Collet et al., 2013, Chapter 6]. This context of quasi-stationary Monte Carlo methods – where we are starting from a density π and wish to construct a killed diffusion whose quasi-stationary distribution coincides with π – is the topic of Chapter 3.

From (3.6) of Chapter 3, if we were to start with a smooth positive density π, the appropriate choice of κ is given for each $y \in M$ by

$$\kappa(y) := \frac{1}{2}\left(\frac{\Delta\pi}{\pi} - \frac{2\nabla A \cdot \nabla\pi}{\pi} - 2\Delta A\right)(y) + K.$$

Here ∇ and Δ are the gradient and Laplacian operators on M, and K is a constant chosen so that κ satisfies (4.5). In this work we will not necessarily assume we are starting from π, and take κ to be a general killing rate satisfying Assumption 7.

Fix $\mu_0 \in \mathcal{P}(M)$ and $r > 0$. We fix a weight function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following assumptions: Define functions $g : \mathbb{R}_+ \to \mathbb{R}_+$ and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$g(t) := \int_0^t \eta_s \, \mathrm{d}s, \quad \alpha_t := \frac{\eta_t}{r + g(t)}, \quad t \geq 0.$$

Assumption 8. *η is continuously differentiable with $\eta_t > 0$ for all $t > 0$, and $g(t) \to \infty$ as $t \to \infty$. α is differentiable and satisfies $\alpha_t \to 0$ as $t \to \infty$, $\int_0^\infty \alpha_s^2 \, \mathrm{d}s < \infty$ and*

$$\sum_{n=1}^{\infty} \int_{h(n)}^{\infty} \alpha_s^2 \, \mathrm{d}s < \infty, \tag{4.6}$$

where h is defined below.

Since g is strictly increasing (as $\eta_t > 0$), continuously differentiable and increases to ∞, it is a diffeomorphism of $\mathbb{R}_+ \to \mathbb{R}_+$. Thus it has a well-defined continuously differentiable inverse g^{-1}. The function $h : \mathbb{R}_+ \to \mathbb{R}_+$ is then given by

$$h(t) := g^{-1}(r\mathrm{e}^t - r), \quad t \geq 0.$$

This function h will be the time change that we shall employ in Section 4.5.

Remark 4.3.1. It follows from Assumption 8 that $\int_0^t \alpha_s \, \mathrm{d}s = \log(1 + g(t)/r) \to \infty$ as $t \to \infty$. Thus these conditions on α are analogous to the typical discrete-time assumptions on the step sizes in traditional stochastic approximation; *cf.* [Kushner and Yin, 2003, Chapter 5]. Since $t \mapsto \int_{h(t)}^\infty \alpha_s^2 \, \mathrm{d}s$ is monotone decreasing, a sufficient condition for (4.6) to hold is that

$$\int_0^\infty \mathrm{d}t \int_{h(t)}^\infty \mathrm{d}s \, \alpha_s^2 = \int_0^\infty \alpha_s^2 \log(1 + g(s)/r) \, \mathrm{d}s < \infty.$$

Define the normalized empirical occupation measures $(\mu_t(\omega))_{t \geq 0}$ by

$$\mu_t(\omega) = \frac{r\mu_0 + \int_0^t \eta_s \delta_{\omega(s-)} \, \mathrm{d}s}{r + \int_0^t \eta_s \, \mathrm{d}s}$$

where $\int_0^t \eta_s \delta_{\omega(s-)} \, \mathrm{d}s(A) = \int_0^t \eta_s 1_A(\omega(s-)) \, \mathrm{d}s$ for each measurable $A \subset M$. In general we will omit the dependence on ω.

Remark 4.3.2. For simplicity one may take $\eta_t \equiv 1$, then $g(t) = t$ and $\alpha_t = 1/(r+t)$ for all $t \geq 0$. Sampling from $Z \sim \int_0^t \eta_s \delta_{X_{s-}} \, \mathrm{d}s / \int_0^t \eta_s \, \mathrm{d}s$ is then equivalent to simulating $V \sim \mathrm{Unif}([0,1])$ and setting $Z = X_{(Vt)-}$. More generally, for $k \geq 0$ one can take $\eta_t = t^k$ for each $t \geq 0$. It is not difficult to check that for this choice of η Assumption 8 is satisfied, and simulating $Z \sim \int_0^t \eta_s \delta_{X_{s-}} \, \mathrm{d}s / \int_0^t \eta_s \, \mathrm{d}s$ is equivalent to the case of constant η except with $V \sim \mathrm{Beta}(k+1, 1)$. Heuristically, choosing a larger value of k prioritizes the more recent times. The choice of the parameter k to accelerate convergence is itself an interesting question, which will be explored in future work. Preliminary simulations involving Brownian motions and unimodal targets seem to suggest $k = 10$ might be a reasonable choice.

For any $\mu \in \mathcal{P}(M)$, define the operator $-L_\mu$ on twice-differentiable functions by

$$-L_\mu f(x) = \frac{1}{2}\Delta f(x) + \nabla A \cdot \nabla f(x) + \kappa(x) \int_M \big(f(y) - f(x)\big) \, \mu(\mathrm{d}y). \tag{4.7}$$

Here we choose to use the negative operator in order to comport with the convention adopted in Kolb and Steinsaltz [2012] and Chapter 3, where it was chosen to make the corresponding self-adjoint operators positive.

We can define probability measures $(\mathbb{P}_x : x \in M)$ with the following properties:

- $\mathbb{P}_x(X_0 = x) = 1$.
- For all smooth $f \in C(M)$

  $$N_t^f := f(X_t) - f(x) - \int_0^t (-L_{\mu_s}) f(X_{s-}) \, \mathrm{d}s$$

  is a $\mathbb{P}_x$-martingale with respect to $(\mathcal{F}_t)$.

The existence of such probability measures can be done analogously to Proposition 2.5 of Benaïm et al. [2002]. In our case, it is in fact simpler and we can construct $\mathbb{P}_x$ by explicitly defining the killing events, as follows. Let $\xi_1, \xi_2, \ldots$ be an i.i.d. sequence of Exp(1) random variables. Set $T_0 = 0$, $X_0 = x$ and given T_n and X_{T_n} inductively define

$$\tau_{n+1} := \inf\left\{ t \geq 0 : \int_0^t \kappa\big(Y_s^{(n+1)}\big) \, \mathrm{d}s \geq \xi_{n+1} \right\} \tag{4.8}$$

where for each $n = 0, 1, 2, \ldots$, $(Y_s^{(n+1)})_{s \geq 0}$ is an independent realisation of the solution to the SDE (4.1) started from $Y_0^{(n+1)} = X_{T_n}$. For a careful treatment of defining diffusions on manifolds, the reader is referred to Stroock [2000].

Then update the path $X_{T_n+s} = Y_s^{(n+1)}$ for $s \in [0, \tau_{n+1})$. Set $T_{n+1} = T_n + \tau_{n+1}$ and independently draw $X_{T_{n+1}} \sim \mu_{T_{n+1}}$, which only depends on the path before time T_{n+1}.

Remark 4.3.3. The practical simulation of this process, namely of the SDE dynamics (4.1) and the killing times (4.8), may seem difficult at first glance. For the killing times, in the present compact setting simulation is actually straightforward, since one can directly employ Poisson thinning (see [Devroye, 1986, Chater VI.2.4]) as κ is bounded. Even in noncompact spaces simulation of such times may be performed without error through *layered processes*; this is the case for the ScaLE algorithm Pollock et al. [2016]. To simulate the SDE, it turns out that through the techniques of *exact simulation* Beskos et al. [2006], in many settings the SDE (4.1) can be simulated on fixed time horizons without error and without resorting to discretization. A thorough computational analysis of the resulting ReScaLE algorithm, including applications to large data sets and comparisons to existing methods is the subject of Kumar [2019].

4.3.2 Fixed Rebirth processes

We now define the fixed rebirth processes and derive some useful properties. These will be crucial later for defining the deterministic semiflow. It will be convenient to work on M with the measure

$$\Gamma(\mathrm{d}y) = \gamma(y)\,\mathrm{d}y,$$

where

$$\gamma(y) = \exp(2A(y)).$$

Let us write L^κ for minus the generator of the diffusion Y from (4.1) killed at rate κ and let $\mathcal{D}(L^\kappa) \subset C(M)$ be its domain. Such f are twice continuously differentiable, and we have

$$-L^\kappa f = \frac{1}{2}\Delta f + \nabla A \cdot \nabla f - \kappa f.$$

Then we have the identity

$$\mathrm{e}^{-tL^\kappa} f(x) = \mathbb{E}_x[f(Y_t)1_{\{\tau_\partial > t\}}] = \mathbb{E}_x\left[f(Y_t)\mathrm{e}^{-\int_0^t \kappa(Y_s)\,\mathrm{d}s}\right]. \tag{4.9}$$

Here (and throughout) τ_∂ denotes a general killing time, defined analogously to (4.8) and Y evolves according to our SDE (4.1) without any killing. The exponentiation of the operator $-L^\kappa$ is rigorously justified through the spectral theorem for self-adjoint operators, and the resulting (sub-Markovian) semigroup $(\mathrm{e}^{-tL^\kappa})_{t\geq 0}$ is strongly continuous on $C(M)$. Details may be found in Demuth and van Casteren [2000].

Similarly to Wang et al. [2019a], Kolb and Steinsaltz [2012], we show the existence of a continuous, positive transition subdensity for the SDE (4.1) killed at rate κ.

Lemma 4.3.4. *The SDE* (4.1) *killed at rate* κ *has a* C^∞ *positive transition subdensity* $p^\kappa(t,x,y)$ *with respect to* Γ*, that is,*

$$\mathrm{e}^{-tL^\kappa} f(x) = \int_M f(y)\, p^\kappa(t,x,y)\Gamma(\mathrm{d}y).$$

Proof. First, we note that the *unkilled* diffusion has a smooth positive transition density $p^0(t,x,y)$ with respect to Γ. The existence of this density is described briefly in Example 9 of [Demuth and van Casteren, 2000, Chapter 1.C], and the details can be found in [Bismut, 1984, Chapter II]. In particular, the assumptions of Bismut [1984] are that the manifold M is a C^∞ compact connected finite-dimensional Riemannian manifold (as in our Assumption 5) and that the drift – our ∇A – is a C^∞ vector field, which we are assuming (Assumption 6).

In order to obtain the transition subdensity for the killed diffusion, we make use of (4.9). By conditioning on the end point we can write

$$\mathrm{e}^{-tL^\kappa} f(x) = \mathbb{E}_x\left[f(Y_t)g(t,x,Y_t)\right] = \int_M f(y)g(t,x,y)p^0(t,x,y)\Gamma(\mathrm{d}y),$$

where for each $t>0$, $x,y\in M$,

$$g(t,x,y) := \mathbb{E}_x\left[\mathrm{e}^{-\int_0^t \kappa(Y_s)\,\mathrm{d}s}\middle| Y_t = y\right].$$

By [Bismut, 1984, Chapter II.d], since κ is bounded C^∞, $g(t,x,y)$ is continuous in x,y and in fact since κ is smooth and nonnegative, $g(t,x,y)$ is jointly continuous over t,x,y and smooth as a function of x or y. Since κ is nonnegative and bounded above, we have the bounds $\mathrm{e}^{-t\bar{\kappa}} \le g(t,x,y) \le 1$. In particular $g(t,x,y)$ is positive, and hence setting for each $t>0$, $x,y\in M$,

$$p^\kappa(t,x,y) = p^0(t,x,y)g(t,x,y), \tag{4.10}$$

we obtain the positive C^∞ transition subdensity $p^\kappa(t,x,y)$ with respect to Γ of diffusion (4.1) killed at rate κ. □

Lemma 4.3.5. *The resolvent operator* $\mathcal{R} : C(M) \to C(M)$ *given by*

$$\mathcal{R}f(x) = \int_0^\infty \mathrm{d}t \int_M \Gamma(\mathrm{d}y)\, p^\kappa(t,x,y) f(y)$$

is a well-defined bounded, positive linear operator.

Remark 4.3.6. We have defined $\mathcal{R}$ as an operator on the Banach space $C(M)$. Its dual operator acts on the space of finite signed Borel measures on M (*cf.* [Dudley, 2002, Section 7.4]). Following the standard probabilistic notation we will denote its dual action on a measure μ by simply $\mu\mathcal{R}$. That is, $\mu\mathcal{R}$ is the measure defined by

$$\mu\mathcal{R}(f) = \int_M \mu(\mathrm{d}x)\mathcal{R}f(x).$$

Proof. $\mathcal{R}$ is clearly linear and maps nonnegative functions to nonnegative functions. It maps continuous functions to continuous functions since $p^\kappa(t,x,y)$ and γ are continuous. Thus $\mathcal{R}$ is a positive linear operator mapping $C(M) \to C(M)$. For the constant function $1 : x \mapsto 1$ we have

$$\mathcal{R}1(x) = \int_0^\infty \mathrm{d}t \int_M \Gamma(\mathrm{d}y)\, p^\kappa(t,x,y) = \int_0^\infty \mathrm{d}t\, \mathbb{P}_x(\tau_\partial > t) = \mathbb{E}_x[\tau_\partial]. \tag{4.11}$$

Since we are assuming that the killing rate is everywhere bounded below by $\underline{\kappa}$, it follows that we have the uniform bound over $x \in M$,

$$\mathbb{E}_x[\tau_\partial] \le 1/\underline{\kappa}.$$

Thus since $\mathcal{R}$ is positive, it follows that $\mathcal{R}$ is bounded. □

We note for future reference that

$$\frac{1}{\bar{\kappa}} \le \inf_{x\in M} \mathcal{R}1(x) \le \sup_{x\in M} \mathcal{R}1(x) \le \frac{1}{\underline{\kappa}}. \tag{4.12}$$

Remark 4.3.7. Heuristically, the resolvent describes the average cumulative occupation measure of the killed diffusion over a single lifetime.

Fix a probability measure $\mu \in \mathcal{P}(M)$. We now define the fixed rebirth process with rebirth distribution μ, abbreviated to FR(μ), to be a Markov process with càdlàg paths, evolving according to the SDE (4.1) between regeneration events, which occur at rate $\kappa(X_t^\mu)$. At such an event the location is drawn independently from distribution μ. It can be constructed explicitly as in the construction at the end of Section 4.3.1 in the simpler case when $X_{T_{n+1}} \sim \mu$ for each n. Let $(P_t^\mu)_{t\ge 0}$ denote the semigroup of this process.

Recall that Y denotes the unkilled process that evolves according to the SDE (4.1). Since the FR(μ) process exhibits a natural renewal behavior, by conditioning on the first arrival time τ_∂, we see that

$$P_t^\mu f(x) = \int_0^t \mathbb{E}_x\left[\kappa(Y_s)\mathrm{e}^{-\int_0^s \kappa(Y_u)\,\mathrm{d}u}\right] \mu P_{t-s}^\mu f \,\mathrm{d}s + \mathbb{E}_x\left[f(Y_t)1_{\{\tau_\partial > t\}}\right]. \tag{4.13}$$

The second term can be expressed equivalently as in (4.9).

Since κ is continuous and bounded, and the unkilled diffusion semigroup of Y is strongly continuous (since A is smooth), it follows directly from the relation (4.13) that $(P_t^\mu)_{t\geq 0}$ is also a strongly continuous semigroup on $C(M)$. This implies that FR(μ) is a Feller–Markov process. Thus by the Hille–Yosida theorem (Theorem 1.7, Guionnet and Zegarlinski [2002]), since $(P_t^\mu)_{t\geq 0}$ *is* a Feller–Markov semigroup, it has an infinitesimal generator $-L_\mu$ defined on a dense domain $\mathcal{D}(L_\mu) \subset C(M)$.

The action of the generator $-L_\mu$ on smooth functions agrees with (4.7). Since κ is bounded and continuous, $\mathcal{D}(L_\mu)$ will consist of twice continuously differentiable functions on M, and in fact $\mathcal{D}(L_\mu)$ is independent of μ.

Lemma 4.3.8. *Given $\mu \in \mathcal{P}(M)$, an invariant measure for the FR(μ) process is given by*

$$\Pi(\mu)(f) = \frac{\mu\mathcal{R}f}{\mu\mathcal{R}1}. \tag{4.14}$$

Proof. Let $f \in \mathcal{D}(L_\mu)$, then it follows that $f \in \mathcal{D}(L^\kappa)$. We wish to show that $\mu\mathcal{R}L_\mu f = 0$ (Proposition 9.2 of [Ethier and Kurtz, 1986, Chapter 9]). Note by (4.9) that we can write

$$\mathcal{R}f = \int_0^\infty e^{-tL^\kappa} f \,\mathrm{d}t.$$

Then

$$\begin{aligned} -\mu\mathcal{R}L_\mu f &= -\mu \int_0^\infty e^{-tL^\kappa} L_\mu f \,\mathrm{d}t \\ &= \mu \int_0^\infty e^{-tL^\kappa} \bigg(-L^\kappa f + \kappa \int f(y)\,\mu(\mathrm{d}y) \bigg) \\ &= \mu \bigg([e^{-tL^\kappa} f]_0^\infty + \mu(f) \int_0^\infty e^{-tL^\kappa} \kappa \,\mathrm{d}t \bigg) \end{aligned}$$

where we used the backward equation (see, for instance, [Ethier and Kurtz, 1986, Chapter 1], Proposition 1.5),

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{-tL^\kappa} f) = -e^{-tL^\kappa} L^\kappa f.$$

Note that by Tonelli's theorem we can exchange the order of integration to find

$$\int_0^\infty e^{-tL^\kappa} \kappa \,\mathrm{d}t = \mathbb{E}\bigg[\int_0^{\tau_\partial} \kappa(Y_s)\,\mathrm{d}s \bigg] = \mathbb{E}[\xi] = 1$$

where $\xi \sim \mathrm{Exp}(1)$ by the definition of our killing construction.

Thus putting the terms together we have that

$$-\mu\mathcal{R}L_\mu f = \mu(-f + \mu(f)1) = -\mu(f) + \mu(f) = 0.$$

Thus it follows that $\Pi(\mu)$, which is the normalized version of the measure $\mu\mathcal{R}$, is an invariant probability measure for the FR(μ) process. □

Proposition 4.3.9. *We have the bound*

$$\|\nu_1 P_t^\mu - \nu_2 P_t^\mu\|_1 \leq 2e^{-t\underline{\kappa}}$$

for any $\nu_1, \nu_2, \mu \in \mathcal{P}(M)$. In particular choosing $\nu_2 = \Pi(\mu)$ gives the bound

$$\|\nu P_t^\mu - \Pi(\mu)\|_1 \leq 2e^{-t\underline{\kappa}} \tag{4.15}$$

for any $\nu, \mu \in \mathcal{P}(M)$. It follows that $\Pi(\mu)$ is the unique invariant probability measure for the FR(μ) process.

Proof. This will follow straightforwardly from the coupling inequality, see, for instance, [Roberts and Rosenthal, 2004, Section 4.1]. This states that $\|\mathcal{L}(X) - \mathcal{L}(Y)\|_1 \leq 2\mathbb{P}(X \neq Y)$ for a coupling $\mathbb{P}$ of random variables X, Y with laws $\mathcal{L}(X), \mathcal{L}(Y)$ respectively.

Since we are assuming the killing rate κ is bounded below by $\underline{\kappa}$ and the rebirth distribution is fixed, we can couple two processes started from different initial distributions at the first arrival time of a homogeneous Poisson process with rate $\underline{\kappa}$. □

Lemma 4.3.10. *The map $\Pi : \mathcal{P}(M) \to \mathcal{P}(M)$ is continuous in the topology of weak-* convergence.*

Proof. We know that $\mathcal{R}$ is a bounded linear operator on $C(M)$ by Lemma 4.3.5. Thus as noted in Remark 4.3.6 it acts dually on the space of finite signed Borel measures, and is continuous on the dual space. So it must also be weak-* continuous. Continuity of $\mu \mapsto \Pi(\mu)$ follows. □

Remark 4.3.11. In fact, since $\mathcal{P}(M)$ is a compact metric space it follows that $\Pi : \mathcal{P}(M) \to \mathcal{P}(M)$ is uniformly continuous.

Proposition 4.3.12. *$\mu \in \mathcal{P}(M)$ satisfies the fixed point equation*

$$\mu = \Pi(\mu) \tag{4.16}$$

if and only if μ is quasi-stationary for the diffusion Y killed at rate κ.

There exists a unique quasi-stationary distribution π for the diffusion Y killed at rate κ. Furthermore, π has a strictly positive C^∞ density with respect to the Riemannian measure, which will also be denoted by π.

Proof. Suppose μ is invariant for L_μ. This means that for all smooth f,

$$\mu L_\mu f = 0;$$

that is, by (4.14)

$$\frac{1}{2}\mu(\Delta f) + \mu(\nabla A \cdot \nabla f) + \mu(\kappa)\mu(f) - \mu(\kappa f) = 0,$$

which is equivalent to

$$\mu L^\kappa f = \mu(\kappa)\mu(f).$$

Since $\mu(\kappa) > 0$ this tells us that μ is a quasi-stationary distribution for X.

Conversely, suppose μ is quasi-stationary for Y. Then

$$\mu L^\kappa f = \lambda_0 \mu(f)$$

for all smooth f and some $\lambda_0 > 0$. Then choosing $f \equiv 1$, we find that $\mu(\kappa) = \lambda_0$, from which it follows that $\mu L_\mu f = 0$. Hence μ is stationary for L_μ.

Existence and uniqueness of the quasi-stationary distribution π follows from Theorem 1.1 of Champagnat and Villemonais [2016]. For each $t > 0$, the transition subdensity $p^\kappa(t, x, y)$ is strictly bounded away from 0; we have that for fixed $t > 0$, we can find some $\epsilon > 0$ such that $p^\kappa(t, x, y) > \epsilon$, for all $x, y \in M$. Then the Assumption (A) of Theorem 1.1 of Champagnat and Villemonais [2016] is easily verified.

In addition, it follows from the existence of the transition subdensity $p^\kappa(t, x, y)$ that π must also be absolutely continuous with respect to the Riemannian measure: Since it is the quasi-stationary distribution of the diffusion Y, by basic properties of quasi-stationary distributions (e.g. Theorem 2.2 of Collet et al. [2013]) there exists some $\lambda_0^\kappa > 0$ such that for each measurable $A \subset M$ and $t > 0$,

$$\pi(A) = \mathbb{P}_\pi(Y_t \in A \,|\tau_\partial > t) = \int_M \pi(\mathrm{d}x) \int_A \mathrm{d}y\, \gamma(y) \mathrm{e}^{t\lambda_0^\kappa} p^\kappa(t, x, y).$$

In particular this implies that π is absolutely continuous with respect to the Riemannian measure; hence π has a density with respect to it, which we will also denote by π. Thus the density π satisfies for each $t > 0$,

$$\pi(y) = \int_M \pi(\mathrm{d}x)\, \mathrm{e}^{t\lambda_0^\kappa} p^\kappa(t, x, y)\gamma(y), \tag{4.17}$$

for almost every $y \in M$. But since $p^\kappa(t, x, y)$ is positive and smooth, it follows that the density π is continuous – so (4.17) holds for all $y \in M$ – and then the density π is smooth, and also positive everywhere. □

Remark 4.3.13. As noted in Section 4.3, in the context of Monte Carlo the quasi-stationary distribution π will be chosen to coincide with a distribution of interest by choosing κ according to (3.6).

4.4 Deterministic Flow

4.4.1 Basic properties

We are now in a position to define the deterministic measure-valued flow that will characterise the asymptotic behavior of the normalized occupation measures $(\mu_t)_{t\geq 0}$.

Recall, as in Benaïm [1999], that on a metric space E a *semiflow* Φ is a jointly continuous map

$$\Phi : \mathbb{R}_+ \times E \to E,$$
$$(t, x) \mapsto \Phi(t, x) = \Phi_t(x)$$

such that Φ_0 is the identity on E and $\Phi_{t+s} = \Phi_t \circ \Phi_s$ for all $s, t \in \mathbb{R}_+$.

We would like to define a semiflow Φ on the space $E = \mathcal{P}(M)$ of probability measures with the topology of weak-* convergence, which is a metric space. In particular we want $t \mapsto \Phi_t(\mu)$ to solve the measure-valued ordinary differential equation (ODE)

$$\dot{\nu}_t = -\nu_t + \Pi(\nu_t), \quad \nu_0 = \mu, \tag{4.18}$$

in the weak sense, meaning that for any test function $f \in C(M)$

$$\Phi_t(\mu)f = \mu f + \int_0^t \big(-\Phi_s(\mu)f + \Pi(\Phi_s(\mu))f \big)\, \mathrm{d}s.$$

We define such a semiflow by adapting the approach of [Benaïm et al., 2018, Section 5] to our present setting. As noted in Lemma 4.3.5, the operator $\mathcal{R}$ is a bounded linear operator mapping from the Banach space $C(M)$ to itself. This allows us to define, for any $t \geq 0$, the bounded linear operator $e^{t\mathcal{R}} : C(M) \to C(M)$, whose dual acts on the space of finite signed Borel measures, equipped with the total-variation norm.

This allows us to define, for each $t \geq 0$, the probability measures

$$\tilde{\Phi}_t(\mu) := \frac{\mu e^{t\mathcal{R}}}{\mu e^{t\mathcal{R}} 1}.$$

The map $t \mapsto \tilde{\Phi}_t(\mu)$ satisfies the weak measure-valued ODE

$$\dot{\nu}_t = \nu_t \mathcal{R} - (\nu \mathcal{R} 1)\nu_t = (\nu_t \mathcal{R} 1)(-\nu_t + \Pi(\nu_t)), \quad \nu_0 = \mu.$$

To get a solution to our actual ODE (4.18) we employ a suitable time change, imitating Blanchet et al. [2016] and Benaïm et al. [2018]. Similarly to Benaïm et al. [2018], for $t \geq 0$ set

$$s_\mu(t) := \int_0^t \tilde{\Phi}_s(\mu)\mathcal{R}1 \, \mathrm{d}s$$

so $\dot{s}_\mu(t) = \tilde{\Phi}_t(\mu)\mathcal{R}1 = \mathbb{E}_{\tilde{\Phi}_t(\mu)}[\tau_\partial] > 0$, so $t \mapsto s_\mu(t)$ is strictly increasing.

Lemma 4.4.1. *$\dot{s}_\mu(t) \geq 1/\bar{\kappa} > 0$ for all $\mu \in \mathcal{P}(M)$ and $t > 0$. Thus in particular $s_\mu(t) \to \infty$ as $t \to \infty$ for any $\mu \in \mathcal{P}(M)$.*

Proof. Immediate since the function $x \mapsto \mathbb{E}_x[\tau_\partial]$ is uniformly bounded below by $1/\bar{\kappa}$ since the killing rate is bounded above by $\bar{\kappa}$. □

For a fixed $\mu \in \mathcal{P}(M)$, since s_μ is a strictly increasing, differentiable map $\mathbb{R}_+ \to \mathbb{R}_+$ we can define an inverse mapping τ_μ, and compose

$$\Phi_t(\mu) = \tilde{\Phi}_{\tau_\mu(t)}(\mu).$$

Recall that we have equipped the space $\mathcal{P}(M)$ with the weak-* topology, which is a compact metric space.

Proposition 4.4.2. *Φ is an injective semiflow on $\mathcal{P}(M)$, and for each $\mu \in \mathcal{P}(M)$ $t \mapsto \Phi_t(\mu)$ is the unique weak solution to* (4.18).

Proof. This is identical to the proof of [Benaïm et al., 2018, Proposition 5.1]. □

4.4.2 Stability of π

Recall that we are interested in working with respect to the measure $\Gamma(\mathrm{d}y) = \gamma(y)\,\mathrm{d}y$ where $\gamma = \exp(2A)$. From Proposition 4.3.12, we see that there is a (unique) quasi-stationary distribution π which is a fixed point of Π, which also has a density with respect to the Riemannian measure, which we also denote by π. Set

$$\varphi := \pi/\gamma$$

which is a smooth function.

We now list some basic facts about π and φ which can all be verified through routine manipulations. There exists $\lambda_0^\kappa > 0$, which describes the asymptotic rate of killing, such that

$$L^\kappa \varphi = \lambda_0^\kappa \varphi.$$

φ is bounded above and bounded away from 0 since by Proposition 4.3.12, π is both bounded above and bounded away from zero.

Writing

$$\beta := \frac{1}{\lambda_0^\kappa}$$

it follows that

$$\begin{aligned} \mathcal{R}\varphi &= \beta\varphi, \\ \pi\mathcal{R} &= \beta\pi \end{aligned} \tag{4.19}$$

where this final identity holds both in terms of π as a measure, and pointwise as density functions.

We wish now to analyse the asymptotic behavior of the semiflow Φ defined in Section 4.4. To do this we will derive a drift condition for a reweighted version of the Markov process.

Since φ is bounded above and away from zero, consider the bounded linear operator $L : C(M) \to C(M)$ given by

$$L = \frac{1}{\varphi}(\mathcal{R} - \beta)\varphi.$$

Thus by exponentiation L generates a Markov semigroup $(\mathcal{K}^t)_{t\geq 0}$. For each $t \geq 0$,

$$\mathcal{K}^t f = \frac{1}{\varphi} \exp(t(\mathcal{R} - \beta))(\varphi f).$$

We can also define the kernels $(\tilde{\mathcal{K}}^t)_{t\geq 0}$,

$$\tilde{\mathcal{K}}^t := \varphi \mathcal{K}^t (f/\varphi) = \exp(t(\mathcal{R} - \beta))f.$$

For the process defined by $(\mathcal{K}^t)_{t\geq 0}$ it can easily be seen from (4.19) that the measure $\pi\varphi$ given by $(\pi\varphi)(f) = \int f(x)\varphi(x)\pi(x)\,\mathrm{d}x$ is an invariant measure. Since φ is bounded, we see that $\pi\varphi$ is in fact a finite measure. In what proceeds, without loss of generality we rescale φ so that $\pi\varphi(1) = 1$.

We would like to show that this process is V-uniformly ergodic, with $V = 1/\varphi$. We do this using a drift condition from Theorem 5.2 of Down et al. [1995]: for a continuous-time irreducible aperiodic Markov process with extended generator L, assume it satisfies for constants $b, c > 0$ and petite set C

$$LV \leq -cV + b1_C, \tag{4.20}$$

then the process is V-uniformly ergodic. Heuristically, a petite set is a set from which the Markov process leaves with a common minorizing measure. The precise definition is carefully presented in [Down et al., 1995, Section 3]. For our present purposes, since we have a positive continuous transition subdensity $p^\kappa(t, x, y)$, it is enough to note that compact sets (and hence the entire space) are petite.

Proposition 4.4.3. *The drift condition* (4.20) *holds.*

Proof. Recall we have set $V = 1/\varphi$. Then for $\beta = 1/\lambda_0^\kappa$,

$$LV = \frac{1}{\varphi}(\mathcal{R} - \beta)1 = \frac{1}{\varphi}(\mathbb{E}_\cdot[\tau_\partial] - \beta) = -\beta V + V\mathbb{E}_\cdot[\tau_\partial] \leq -\beta V + b$$

where b is an upper bound for $V(x)\mathbb{E}_x[\tau_\partial]$ for all x, which exists since V is bounded above and κ is bounded away from 0. Note our entire space is compact, hence petite. Thus the drift condition holds. $\square$

Remark 4.4.4. We note that since the entire space M is compact hence petite, the drift condition (4.20) can be trivially satisfied by choosing $V \equiv 1$, the constants $b = c = 1$ and $C = M$. However we have kept the choice $V = \varphi$ in the proof since this choice is suggestive of how one might generalise this to noncompact spaces.

By [Down et al., 1995, Theorem 5.2] this implies V-uniform ergodicity: There exist constants D and $0 \leq \rho < 1$ such that for all $x \in M$:

$$\sup_{|g|\leq V} |\mathcal{K}^t(x,g) - \pi\varphi(g)| \leq V(x) D\rho^t.$$

Multiplying through by $\varphi(x)$ and relabeling φg as f, we see that the condition $|g| \leq V = 1/\varphi$ is equivalent to $|f| \leq 1$, hence proving uniform ergodicity: For any $x \in M$

$$\sup_{|f|\leq 1} |\tilde{\mathcal{K}}^t(x,f) - \varphi(x)\pi(f)| \leq D\rho^t.$$

Thus we will have, for any initial distribution μ,

$$\sup_{|f|\leq 1} |\mu\tilde{\mathcal{K}}^t f - \mu(\varphi)\pi(f)| \leq D\rho^t. \tag{4.21}$$

Proposition 4.4.5. *We have convergence $\tilde{\Phi}_t(\mu) \to \pi$ as $t \to \infty$ in total variation distance, uniformly in μ .*

Proof. Let $\varphi_* := \min_{x\in M} \varphi(x)$, which is positive, since φ is a continuous positive function on a compact set. We find that for any $t \geq (\log 2D - \log \varphi_*)/\log \rho^{-1}$ we have by (4.21)

$$\mu\tilde{\mathcal{K}}^t 1 \geq \mu(\varphi) - D\rho^t \geq \frac{\varphi_*}{2}.$$

Since

$$\tilde{\Phi}_t(\mu) = \frac{\mu\tilde{\mathcal{K}}^t}{\mu\tilde{\mathcal{K}}^t 1},$$

we have then for any probability measure μ and continuous f with $|f| \leq 1$

$$\begin{aligned}
|\tilde{\Phi}_t(\mu)f - \pi(f)| &\leq \left(\mu\tilde{\mathcal{K}}^t 1\right)^{-1} \left[\left|\mu\tilde{\mathcal{K}}^t f - \mu(\varphi)\pi(f)\right| + |\pi(f)| \cdot \left|\mu(\varphi) - \mu\tilde{\mathcal{K}}^t 1\right|\right] \\
&\leq \frac{2}{\varphi_*} \cdot \left[1 + |\pi(f)|\right] D\rho^t \\
&\leq \frac{4D}{\varphi_*}\rho^t.
\end{aligned}$$

□

Finally, we see that this convergence carries over to the semiflow Φ.

Theorem 4.4.6. *We have convergence $\Phi_t(\mu) \to \pi$ as $t \to \infty$ uniformly in μ in total variation distance.*

Proof. This follows from Proposition 4.4.5 and the fact that $\dot{s}_\mu(t)$ is bounded above by $1/\underline{\kappa}$ uniformly in μ and t, as in Lemma 4.3.5. The boundedness of this derivative ensures that its inverse τ_μ satisfies for all $\mu \in \mathcal{P}(M)$ and $t \geq 0$

$$\tau_\mu(t) \geq \underline{\kappa} t.$$

□

Remark 4.4.7. In the language of Benaïm [1999], this shows that π is a *global attractor* of the semiflow Φ.

4.5 Asymptotic Pseudo-Trajectories

4.5.1 Basic properties and definitions

We now wish to relate the stochastic behavior of the normalised weighted empirical measures

$$\mu_t = \frac{r\mu_0 + \int_0^t \eta_s \delta_{X_{s-}} \, \mathrm{d}s}{r + \int_0^t \eta_s \, \mathrm{d}s}$$

to the deterministic behavior of the flow defined in Section 4.4.

Definition. ([Benaïm, 1999, Section 3]) For a metric space (E, d), given a semiflow Φ on E, a continuous function $w : [0, \infty) \to E$ is an *asymptotic pseudo-trajectory of* Φ if for all $T > 0$,

$$\lim_{t\to\infty} \sup_{s\in[0,T]} d\big(w(t+s), \Phi_s(w(t))\big) = 0.$$

Recall that by Assumption 8, since g is continuously differentiable, strictly increasing and $g(t) \to \infty$ as $t \to \infty$, it is a diffeomorphism of $\mathbb{R}_+ \to \mathbb{R}_+$. Thus we can define its inverse function $g^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$, which is also continuously differentiable and satisfies $g^{-1}(t) \to \infty$ as $t \to \infty$. Let

$$\zeta_t := \mu_{h(t)}$$

for all $t \geq 0$, where we defined $h(t) := g^{-1}(re^t - r)$. We will show that almost surely $t \mapsto \zeta_t$ is an asymptotic pseudo-trajectory of the semiflow Φ defined in Section 4.4.

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t = \alpha_t\big(-\mu_t + \delta_{X_{t-}}\big), \tag{4.22}$$

applying the chain rule and product rule for derivatives yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta_t = \big(-\zeta_t + \Pi(\zeta_t)\big) + \big(\delta_{X_{h(t-)}} - \Pi(\zeta_t)\big). \tag{4.23}$$

Looking at the first bracket, we recognize the flow from Section 4.4. Thus to formally show that ζ_t approximates the flow, we need to control the second bracket. We also note here for future reference that

$$\begin{aligned}\frac{\partial}{\partial t}\Pi(\mu_t)f &= \alpha_t\bigg(-\frac{\mu_t\mathcal{R}f}{\mu_t\mathcal{R}1} + \frac{\mathcal{R}f(X_{t-})}{\mu_t\mathcal{R}1} + \frac{\mu_t\mathcal{R}f}{\mu_t\mathcal{R}1} - \frac{\mu_t\mathcal{R}f\cdot\mathcal{R}1(X_{t-})}{(\mu_t\mathcal{R}1)^2}\bigg)\\ &= \frac{\alpha_t}{\mu_t\mathcal{R}1}\bigg(\mathcal{R}f(X_{t-}) - \mathcal{R}1(X_{t-})\Pi(\mu_t)f\bigg).\end{aligned} \tag{4.24}$$

We formalize this intuition by the approach of [Benaïm et al., 2002, Proposition 3.5] and [Kurtzmann, 2010, Lemma 5.4]. It is proven in Theorem 3.2 of Benaïm [1999] that asymptotic pseudo-trajectories must be uniformly continuous. Conversely, Theorem 3.2 of Benaïm [1999] also tells us that a uniformly continuous path ζ is an asymptotic pseudo-trajectory if and only if every limit point of the time shifts $\Theta^t\zeta$ in the topology of uniform convergence on compact sets is itself a trajectory of the flow. We define $\mathcal{M}(M)$ to be the space of Borel signed measures on M, equipped with the weak-* topology, which can be metrized analogously to (4.4). Let $C(\mathbb{R}_+, \mathcal{P}(M))$ and $C(\mathbb{R}_+, \mathcal{M}(M))$ be the spaces of continuous paths mapping $\mathbb{R}_+$ into $\mathcal{P}(M)$ and $\mathcal{M}(M)$ respectively, each equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$. As usual, for each $t \geq 0$ we define $\Theta^t : C(\mathbb{R}_+, \mathcal{P}(M)) \to C(\mathbb{R}_+, \mathcal{P}(M))$ to be the shift map given by

$$\big[\Theta^t\zeta\big]_s = \zeta_{t+s}, \quad s \geq 0.$$

Defining the *retraction* $\hat{\Phi} : C(\mathbb{R}_+, \mathcal{P}(M)) \to C(\mathbb{R}_+, \mathcal{P}(M))$ as in Benaïm [1999] and [Benaïm et al., 2002, Proposition 3.5] by

$$\hat{\Phi}(\zeta)(s) = \Phi_s(\zeta_0), \quad s \geq 0,$$

showing that ζ is an asymptotic pseudo-trajectory of Φ is then equivalent to showing that the limit points of $\{\Theta^t\zeta\}_{t\geq 0}$ are fixed points of the retraction $\hat{\Phi}$.

We also define in analogue to Benaïm et al. [2002] the operator $L_F : C(\mathbb{R}_+, \mathcal{P}(M)) \to C(\mathbb{R}_+, \mathcal{M}(M))$ by

$$L_F(\nu)(s) = \nu(0) + \int_0^s F(\nu(u))\,\mathrm{d}u, \quad s \geq 0,$$

with $F(\nu) := -\nu + \Pi(\nu)$, which will be used in the subsequent proof.

Define the collection of signed measures

$$\epsilon_t(s) := \int_t^{t+s} \left(\delta_{X_{h(u-)}} - \Pi(\zeta_u)\right) \mathrm{d}u$$

for all $t, s \geq 0$. We note that for each $t \geq 0$, $\epsilon_t(\cdot) \in C(\mathbb{R}_+, \mathcal{M}(M))$.

Theorem 4.5.1. *Suppose $\zeta : \mathbb{R}_+ \to \mathcal{P}(M)$ is a continuous path as described above. Then ζ is an asymptotic pseudo-trajectory for Φ if and only if we have the following condition:*

For any $T > 0$ and smooth $f \in C(M)$

$$\lim_{t\to\infty} \sup_{s\in[0,T]} |\epsilon_t(s)f| = 0. \tag{4.25}$$

Remark 4.5.2. Theorem 3.2 of Benaïm [1999] assumes relative compactness of the image of the path ζ. By Prokhorov's theorem (see [Dudley, 2002, Theorem 11.5.4]), relative compactness in $\mathcal{P}(M)$ is equivalent to tightness, which trivially holds in our present compact setting.

Proof. Given continuous f with $\|f\|_\infty \leq 1$, we have

$$|\zeta_{t+s}f - \zeta_t f| \leq 2|s|$$

since

$$\frac{\mathrm{d}(\zeta_t f)}{\mathrm{d}t} = -\zeta_t f + f(X_{h(t-)}).$$

Hence we have uniform continuity.

Suppose the condition (4.25) holds for any $T > 0$ and smooth $f \in C(M)$. This says that $\epsilon_t(\cdot)$ converges to 0 in $C(\mathbb{R}_+, \mathcal{M}(M))$. Consider the family $\{\Theta^t\zeta\}_{t\geq 0}$. Suppose $\tilde{\zeta}$ is a limit point of this family in $C(\mathbb{R}_+, \mathcal{P}(M))$. Analogously to Benaïm et al. [2002], we can use (4.23) and the definition of ϵ_t to write

$$\Theta^t\zeta = L_F(\Theta^t\zeta) + \epsilon_t(\cdot). \tag{4.26}$$

Since we are assuming precisely (4.25), and since L_F is continuous, taking $t \to \infty$ in (4.26) along the appropriate subsequence we obtain $\tilde{\zeta} = L_F(\tilde{\zeta})$. This shows that the limit path $\tilde{\zeta}$ is a fixed point of L_F, that is,

$$\tilde{\zeta}_s = \tilde{\zeta}_0 + \int_0^s F(\tilde{\zeta}_u)\,\mathrm{d}u, \quad s \geq 0.$$

By uniqueness of the flow this implies that $\tilde{\zeta}_s = \Phi_s(\tilde{\zeta}_0)$ for all $s \geq 0$, that is, $\tilde{\zeta}$ is a fixed point of the retraction $\hat{\Phi}$. This concludes the proof of the sufficiency of condition (4.25).

Conversely, suppose ζ is an asymptotic pseudo-trajectory for Φ. By definition, this means that for each $T > 0$, as $t \to \infty$,

$$\sup_{s\in[0,T]} d_w\left(\zeta_{t+s}, \Phi_s(\zeta_t)\right) \to 0. \tag{4.27}$$

By the representation (4.26) we would like to show that for each $T > 0$,

$$\sup_{s\in[0,T]} d_w\left(\Theta^t\zeta(s), L_F(\Theta^t\zeta)(s)\right) = \sup_{s\in[0,T]} d_w\left(\zeta_{t+s},\, \zeta_t + \int_0^s F(\zeta_{t+u})\,\mathrm{d}u\right) \to 0$$

By (4.27), it is sufficient to control $d_w(L_F(\Theta^t\zeta_s), \Phi_s(\zeta_t))$ uniformly over $s \in [0, T]$.

By the definition of the flow Φ, this will follow if we can control

$$\sup_{s\in[0,T]} d_w\left(\int_0^s F(\zeta_{t+u})\,\mathrm{d}u, \int_0^s F(\Phi_u(\zeta_t))\,\mathrm{d}u\right). \tag{4.28}$$

Since F is uniformly continuous (since $\Pi : \mathcal{P}(M) \to \mathcal{P}(M)$ is uniformly continuous; Remark 4.3.11), (4.28) can be made arbitrarily small by choosing t sufficiently large because of (4.27).

Thus by (4.26) we conclude that $\epsilon_t \to 0$ in $C(\mathbb{R}_+, \mathcal{M}(M))$, that is, (4.25) must hold for any $T > 0$ and smooth $f \in C(M)$ as required. □

The goal now is to establish the requirements of Theorem 4.5.1 almost surely. So we need to establish (4.25).

4.5.2 Poisson Equation

To show that (4.25) holds, inspired by [Benaïm et al., 2002, Section 5.2] and [Benaïm et al., 2018, Section 6.2], we will measure the discrepancy between the cumulative occupancy and the quasi-stationary distribution by a solution to the Poisson equation

$$-f + \Pi(\mu)f = -L_\mu g. \tag{4.29}$$

Fix any $\mu \in \mathcal{P}(M)$. Define for $f \in C(M)$,

$$Q_\mu f := \int_0^\infty \left(P_t^\mu f - \Pi(\mu) f\right) \mathrm{d}t.$$

Lemma 4.5.3. *For any $\mu \in \mathcal{P}(M)$, Q_μ is a bounded linear operator mapping $C(M)$ to itself, and*

$$\|Q_\mu f\|_\infty \leq \frac{2\|f\|_\infty}{\underline{\kappa}}$$

for any $\mu \in \mathcal{P}(M)$ and $f \in C(M)$.

Proof. $Q_\mu f$ is continuous since f is continuous and (P_t^μ) is Feller. From (4.15), we have that

$$|Q_\mu f(x)| \leq \int_0^\infty \|P_t^\mu(x) - \Pi(\mu)\|_1 \|f\|_\infty \,\mathrm{d}t \leq 2\|f\|_\infty \int_0^\infty e^{-t\underline{\kappa}} \,\mathrm{d}t = \frac{2\|f\|_\infty}{\underline{\kappa}}$$

uniformly over $x \in M$ and $\mu \in \mathcal{P}(M)$. □

Recall that $-L_\mu$ is the generator of the FR(μ) process. By the basic properties of infinitesimal generators (see, for instance, Guionnet and Zegarlinski [2002]), we have that for any $t > 0$, $f \in C(M)$, that $\int_0^t P_s^\mu f \,\mathrm{d}s \in \mathcal{D}(L_\mu)$, and $(-L_\mu)\int_0^t P_s^\mu f \,\mathrm{d}s = P_t^\mu f - f$. Then

$$P_t^\mu f(x) = f(x) + (-L_\mu) \int_0^t P_s^\mu f(x) \,\mathrm{d}s. \tag{4.30}$$

Proposition 4.5.4. *Given $f \in C(M)$, $Q_\mu f \in \mathcal{D}(L_\mu)$, and $g = Q_\mu f$ solves the Poisson equation* (4.29).

Proof. Without loss of generality, assume $\Pi(\mu) f = 0$; just replace f with $f - \Pi(\mu) f$. Then taking $t \to \infty$ in (4.30) and using Proposition 4.3.9 we see by closure of $-L_\mu$ that $Q_\mu \in \mathcal{D}(L_\mu)$, and that $\Pi(\mu) f = f - L_\mu Q_\mu f$. Thus (4.29) is satisfied by $g = Q_\mu f$. □

Remark 4.5.5. By basic properties of generators of Feller processes (*cf.* [Revuz and Yor, 1991, Section VII.1]), we similarly have that for $f \in \mathcal{D}(L_\mu)$,

$$-f + \Pi(\mu) f = -Q_\mu L_\mu f. \tag{4.31}$$

Remark 4.5.6. Note that we can think of the operator

$$K_\mu : f \mapsto f - \Pi(\mu) f$$

as a projection operator since it is linear and idempotent ($K_\mu^2 = K_\mu$).

4.5.3 Bounding the discrepancy

Using our solution to the Poisson equation, we now rewrite the discrepancy term, and decompose it using Itô's formula. Making the change of variables $u \leftarrow h(u)$ we write

$$\begin{aligned}\epsilon_t(s)f &= \int_t^{t+s} (f(X_{h(u-)}) - \Pi(\zeta_u)f)\,\mathrm{d}u \\ &= \int_{h(t)}^{h(t+s)} (-L_{\mu_u})Q_{\mu_u}f(X_{u-})\,\alpha_u \mathrm{d}u.\end{aligned}$$

Consider now a time-dependent function $f : \mathbb{R} \times M \to \mathbb{R}$, written as $f_s(x)$. For simplicity we will take the notation ∇ and ∇^2 to refer to the gradient with respect to the coordinates of M, and f'_s to be the time derivative $\partial f_t/\partial t \,\big|_{t=s}$. We now apply Itô's formula for general semimartingales (as formulated, for instance, as Theorem 14.2.4 of Cohen and Elliott [2015]), taking advantage of the fact that all quadratic covariation terms are 0 for the Brownian motion that is driving our process X_t:

$$\begin{aligned}f_t(X_t) - f_0(X_0) = \int_0^t \nabla f_u(X_{u-}) \cdot \mathrm{d}X_u + \frac{1}{2}\int_0^t \nabla^2 f_u(X_{u-})\,\mathrm{d}u + \int_0^t f'_u(X_{u-})\,\mathrm{d}u \\ + \sum_{0<u\leq t} \big(f_u(X_u) - f_u(X_{u-}) - \nabla f_u(X_{u-}) \cdot \Delta X_u\big).\end{aligned}$$

Using the formula (4.7) for L_μ we find

$$\begin{aligned}f_t(X_t) - f_0(X_0) = \int_0^t (-L_{\mu_u})f_u(X_{u-})\,\mathrm{d}u + \int_0^t \nabla f_u(X_{u-}) \cdot \mathrm{d}W_u + \int_0^t f'_u(X_{u-})\,\mathrm{d}u \\ + \sum_{0<u\leq t} \big(f_u(X_u) - f_u(X_{u-})\big) - \int_0^t \kappa(X_{u-}) \int \big(f_u(y) - f_u(X_{u-})\big)\mu_u(\mathrm{d}y)\,\mathrm{d}u,\end{aligned}$$

where W_u is a Wiener process on M.

We apply this formula now to the function $f_s(x) = Q_{\mu_s}f(x)\alpha_s$. By Proposition 4.5.4, $Q_{\mu_s}f \in \mathcal{D}(L_{\mu_s})$, and so in particular $Q_{\mu_s}f$ is twice continuously differentiable. Thus $f_s(x)$ is indeed twice continuously differentiable with respect to x. We rearrange the terms to obtain

$$\begin{aligned}\epsilon_s(t)f &= \int_{h(t)}^{h(t+s)} (-L_{\mu_u})Q_{\mu_u}f(X_{u-})\,\alpha_u \mathrm{d}u \\ &= \epsilon_s^{(1)}(t)f + \epsilon_s^{(2)}(t)f + \epsilon_s^{(3)}(t)f + \epsilon_s^{(4)}(t)f + \epsilon_s^{(5)}(t)f,\end{aligned} \tag{4.32}$$

where

$$\begin{aligned}
\epsilon_t^{(1)}(s)f &= Q_{\mu_{h(t+s)}}f(X_{h((t+s)-)})\,\alpha_{h(t+s)} - Q_{\mu_{h(t)}}f(X_{h(t-)})\,\alpha_{h(t)},\\
\epsilon_t^{(2)}(s)f &= \int_{h(t)}^{h(t+s)} Q_{\mu_u}f(X_{u-})\frac{\mathrm{d}\alpha_u}{\mathrm{d}u}\,\mathrm{d}u,\\
\epsilon_t^{(3)}(s)f &= -\int_{h(t)}^{h(t+s)} \frac{\partial}{\partial u}\Big(Q_{\mu_u}f\Big)(X_{u-})\,\alpha_u\,\mathrm{d}u,\\
\epsilon_t^{(4)}(s)f &= -N^f_{h(t+s)} + N^f_{h(t)},\\
\epsilon_t^{(5)}(s)f &= -J^f_{h(t+s)} + J^f_{h(t)},
\end{aligned}$$

and N and J are the local martingales

$$\begin{aligned}
N_t^f &:= \int_0^t \nabla Q_{\mu_u}f(X_{u-})\,\alpha_u\cdot\mathrm{d}W_u,\\
J_t^f &:= \sum_{0<u\le t}\Big(Q_{\mu_u}f(X_u)\alpha_u - Q_{\mu_u}f(X_{u-})\alpha_u\Big)\\
&\quad - \int_0^t \mathrm{d}u\,\kappa(X_{u-})\int\mu_u(\mathrm{d}y)\Big(Q_{\mu_u}f(y)\alpha_u - Q_{\mu_u}f(X_{u-})\alpha_u\Big).
\end{aligned}$$

Theorem 4.5.7. *The conditions of Theorem 4.5.1 hold almost surely. This implies that $t\mapsto\zeta_t$ is almost surely an asymptotic pseudo-trajectory for Φ.*

Proof. To establish (4.25) we will use the decomposition (4.32) and consider the five error terms individually.

4.5.3.1 $\epsilon_t^{(1)}(s)f$ and $\epsilon_t^{(2)}(s)f$

Using the bound from Lemma 4.5.3, we see that

$$\begin{aligned}
|\epsilon_t^{(1)}(s)f| &\le \|Q_{\mu_{h(t+s)}}f\|_\infty\alpha_{h(t+s)} + \|Q_{\mu_{h(t)}}f\|_\infty\alpha_{h(t)}\\
&\le 4\underline{\kappa}^{-1}\|f\|_\infty(\alpha_{h(t+s)} - \alpha_{h(t)}),\\
|\epsilon_t^{(2)}(s)f| &\le \int_{h(t)}^{h(t+s)}\|Q_{\mu_u}f\|_\infty\frac{\mathrm{d}\alpha_u}{\mathrm{d}u}\,\mathrm{d}u\\
&= 2\underline{\kappa}^{-1}\|f\|_\infty(\alpha_{h(t+s)} - \alpha_{h(t)}).
\end{aligned}$$

Since $h(t)\to\infty$ as $t\to\infty$ and $\alpha_t\to 0$ as $t\to\infty$ by Assumption 8, we see that both terms decay to 0 as $t\to\infty$.

4.5.3.2 $\epsilon_t^{(3)}(s)f$

By the same argument as in the proof of Lemma 5.5 of Benaïm et al. [2002] we have

$$\frac{\partial}{\partial t}Q_{\mu_t} = -\left(\frac{\partial}{\partial t}K_{\mu_t} + Q_{\mu_t}\frac{\partial}{\partial t}(-L_{\mu_t})\right)Q_{\mu_t}, \tag{4.33}$$

where $K_\mu f = f - \Pi(\mu)f$. While our $-L_\mu$ is not the same as their operator A_μ, the same proof holds, since it relies only upon the Poisson equation (4.29) and (4.31).

Applying (4.22) to the definition (4.7) of L_{μ_t} we obtain

$$-\frac{\partial}{\partial t}L_{\mu_t}f(x) = \kappa(x)\alpha_t\Big(f(X_{t-}) - \mu_t(f)\Big).$$

From (4.24) we obtain an upper bound

$$\left\|\alpha_t^{-1}\frac{\partial K_{\mu_t}}{\partial t}Q_{\mu_t}f\right\|_\infty \le \left|\frac{1}{\mu_t\mathcal{R}1}\right|\left(\left\|\mathcal{R}Q_{\mu_t}f(X_{t-})\right\|_\infty + \left|\Pi(\mu_t)\mathcal{R}1(X_{t-})\right|\right),$$

so by Lemma 4.5.3 and the bounds (4.12) we see that there is a constant C (depending on the upper and lower bounds on κ) such that

$$\sup_{x\in M}\left|\frac{\partial Q_{\mu_t}}{\partial t}f(x)\right| \le C\|f\|_\infty\alpha_t.$$

This implies that

$$|\epsilon_t^{(3)}(s)f| \le C\|f\|_\infty\int_{h(t)}^\infty \alpha_u^2\,\mathrm{d}u.$$

Since the definite integral from 0 to ∞ is finite by Assumption 8, it follows immediately that $\epsilon_t^{(3)}(s)f$ decays to 0 as $t\to\infty$.

4.5.3.3 $\epsilon_t^{(4)}(s)f$

We now turn to the first of the two martingale terms. Similarly to Proposition 5.3 of Kurtzmann [2010], our goal is to control the quadratic variation and then apply the Burkholder–Davis–Gundy inequality.

The quadratic variation of the martingale N_t^f is bounded by

$$\int_0^t \|\nabla Q_{\mu_s}f(X_s)\|_\infty^2\,\alpha_u^2\,\mathrm{d}s\,.$$

We can bound this by means of the inequality

$$\|\nabla P_t^\mu f\|_\infty \le \frac{C_1\|f\|_\infty}{\sqrt{t}}, \quad \text{for } t \text{ sufficiently small,} \tag{4.34}$$

for some constant C_1. Such an inequality was shown to hold in the proof of Lemma 5.1 of Benaïm et al. [2002] for diffusive processes on compact manifolds without jumps. To see that this inequality also holds in the present setting, consider the expression for the semigroup (4.13):

$$P_t^\mu f(x) = \int_0^t \mathbb{E}_x \left[\kappa(Y_s) 1_{\{\tau_\partial > s\}}\right] \mu P_{t-s}^\mu f \, \mathrm{d}s + \mathbb{E}_x \left[f(Y_t) 1_{\{\tau_\partial > t\}}\right].$$

With this representation, it suffices to show that an inequality of the form (4.34) for the killed semigroup given in Lemma 4.3.4. This will ensure that our resulting bound (4.34) is uniform over μ.

By (4.10), since $|g(t,x,y)| \leq 1$ uniformly, and since we know that such an inequality holds for diffusive processes on compact manifolds without jumps, it suffices to show that

$$\|\nabla_x g(t,\cdot,y)\|_\infty \leq \frac{C}{\sqrt{t}} \tag{4.35}$$

for some constant C uniformly over y, for all t sufficiently small. By the Girsanov–Cameron–Martin formula Elworthy [1982] we can write

$$g(t,x,y) = \mathbb{Q}_t \left[\mathrm{e}^{-\int_0^t \kappa(Y_s)\,\mathrm{d}s} G(Y)\right]$$

where for $s \in [0,t]$, $Y_s = (1-s/t)x + (s/t)y + \hat{W}_s$, where under $\mathbb{Q}_t$, $\hat{W}$ is a standard Brownian bridge with $\hat{W}_0 = \hat{W}_t = 0$. $G(Y)$ is given by

$$G(Y) = \exp\left(A(y) - A(x) - \frac{1}{2}\int_0^t \Delta A(Y_s)\,\mathrm{d}s - \frac{1}{2}\int_0^t \|\nabla A(Y_s)\|^2\,\mathrm{d}s\right),$$

since our drift is assumed to be of a gradient form ∇A. Having now expressed Y with explicit dependence upon the initial position x, we can calculate $\nabla_x g(t,\cdot,y)$. Since the functions κ and A are assumed smooth (hence they and their derivatives are bounded uniformly), we can conclude that a bound of the form (4.35) holds. In fact, the bound we obtain is actually uniform over small t, and does not blow up as $t \to 0$.

Armed with (4.34), we then have

$$\int_0^\infty \|\nabla P_t^\mu f\|_\infty \,\mathrm{d}t \leq C_2 \|f\|_\infty, \tag{4.36}$$

for some constant C_2, obtained by considering separately the integrals over $(0,t_0)$ and (t_0,∞). Applying (4.34) bounds the former, while the semigroup property allows the

latter piece to be bounded by

$$\int_{t_0}^{\infty} \|\nabla P_t^\mu f\|_\infty \,\mathrm{d}t \le \int_0^{\infty} \|\nabla P_{t_0}^\mu (P_t^\mu (K_\mu f))\|_\infty \,\mathrm{d}t \le \frac{C_1}{\sqrt{t_0}} \int_0^{\infty} \|P_t^\mu (K_\mu f)\|_\infty \,\mathrm{d}t.$$

Here we replaced f with $K_\mu f = f - \Pi(\mu) f$ since $\Pi(\mu) f$ is a constant so $\nabla \Pi(\mu) f = 0$. This final term can be bounded just as in Lemma 4.5.3.

Since $\nabla Q_\mu f = \int_0^\infty \nabla P_t^\mu \,\mathrm{d}t$, (4.36) immediately implies a universal bound

$$\|\nabla Q_{\mu_u} f\|_\infty \le C_3 \|f\|_\infty.$$

for smooth f. This gives us a bound for the quadratic variation of

$$C_3^2 \|f\|_\infty^2 \int_0^t \alpha_s^2 \,\mathrm{d}s.$$

The Burkholder–Davis–Gundy inequality then implies the existence of a constant C_4 such that for any $\delta > 0$

$$\mathbb{P}_x \left(\sup_{s \in [0,T]} |\epsilon_t^{(4)}(s) f| \ge \delta \right) \le \frac{C_4 \|f\|_\infty^2}{\delta^2} \int_{h(t)}^{\infty} \alpha_s^2 \,\mathrm{d}s.$$

Using (4.6) from Assumption 8, it follows from the Borel–Cantelli lemma that almost surely, for any $\delta > 0$ and $T > 0$,

$$\limsup_{n \to \infty} \sup_{s \in [0,T]} |\epsilon_n^{(4)}(s) f| \le \delta. \tag{4.37}$$

Thus for any $\delta > 0$ we may find N sufficiently large so that for any $t \ge N - T$,

$$\sup_{s \in [0,T]} \left|\epsilon_t^{(4)}(s) f\right| \le \sup_{s \in [0,1]} \left|\epsilon_{\lfloor t \rfloor}^{(4)}(s) f\right| + \sup_{s \in [0,T+1]} \left|\epsilon_{\lfloor t \rfloor}^{(4)}(s) f\right| + \sup_{s \in [0,1]} \left|\epsilon_{\lfloor t+T \rfloor}^{(4)}(s) f\right| \le \delta.$$

Since $\delta > 0$ is arbitrary, it follows that almost surely for every T

$$\lim_{t \to \infty} \sup_{s \in [0,T]} |\epsilon_t^{(4)}(s) f| = 0.$$

4.5.3.4 $\epsilon_t^{(5)}(s) f$

The final error term is a jump martingale term. We use the same approach as above for $\epsilon_t^{(4)}(s) f$. For this jump martingale the quadratic variation is

$$\sum_{0 < u \le t} \left(Q_{\mu_u} f(X_u) \alpha_u - Q_{\mu_u} f(X_{u-}) \alpha_u \right)^2,$$

where the sum is over jump-points $u \in (0, t]$. The squared jump at time u is bounded by

$$16\|f\|_\infty^2 \underline{\kappa}^{-2} \alpha_u^2,$$

making use of Lemma 4.5.3. The expected quadratic variation is then the expectation of the predictable variation, which is bounded by

$$\mathbb{E}\left[\int_0^t \kappa(X_u) \alpha_u^2 \, \mathrm{d}u \right] \leq \bar{\kappa} \int_0^t \alpha_u^2 \, \mathrm{d}u. \tag{4.38}$$

Thus the total quadratic variation is bounded, and as before we conclude that almost surely

$$\lim_{t \to \infty} \sup_{s \in [0,T]} |\epsilon_t^{(5)}(s) f| = 0.$$

4.5.3.5 Concluding the proof

We have shown that the five discrepancy terms all converge to 0 uniformly on compact sets as $t \to \infty$ almost surely. Thus condition (4.25) holds almost surely and the proof is complete. □

4.5.4 Proof of Corollary 4.2.3

This follows from Theorem 4.2.1 and Remark 4.4.7, since limit sets of asymptotic pseudo-trajectories are *attractor free sets*, and so will be contained in the global attractor of Φ, which is $\{\pi\}$. These relationships are spelled out in [Benaïm, 1999, Section 5]. □

4.6 Subsequent work

After the work described in this chapter was first presented online in August 2018, several related pieces of work have since been done. In Mailler and Villemonais [2018], the authors consider *measure-valued Pólya processes* (MVPP). These processes are measure-valued urns, whereby at each discrete time step a point is drawn from the urn, then depending on the point drawn a new (measure-valued) entry is added to the urn, which itself could be random.

In the case when the weights $\eta_t \equiv 1$ are constant, the ReScaLE algorithm as studied in this chapter is precisely such an MVPP. Mailler and Villemonais [2018] prove a general convergence result for the almost sure convergence of MVPPs, using the techniques of stochastic approximation.

In the context of stochastic approximation of quasi-stationary distributions as in this chapter, they prove convergence of the algorithm (when the weights $\eta_t \equiv 1$ are constant) on possibly *noncompact state spaces*, provided the killing κ is uniformly bounded, and there is at least linear drift back to the center of the space. Namely, the diffusion is the solution of the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \mathrm{d}B_t,$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ satisfies

$$\limsup_{\|x\|\to\infty} \frac{\langle b(x), x\rangle}{\|x\|} < -\frac{3}{2}\|\kappa\|_\infty^{1/2}.$$

This work demonstrated – for the first time – that indeed extensions to noncompact spaces are possible for such algorithms. However, several desirable situations are still excluded. For example, the case of Brownian motion on $\mathbb{R}^d$ with isotropic Gaussian quasi-stationary distribution falls out of this framework, since, firstly, it has no drift, and secondly, its killing rate is quadratic (hence unbounded). This general case of potentially unbounded killing with or without drift conditions remains an open question.

In very recent work, Benaïm et al. [2019], the authors consider a similar algorithm to the one studied in this chapter, except there is no soft killing, but hard killing at the boundary of an open bounded domain $D \subset \mathbb{R}^d$. They similarly prove almost sure weak-* convergence of the normalised empirical occupation measures to the quasi-stationary distribution.

The key innovation in this work is to handle the fact that D, since open, is not compact, and so the sequence of empirical occupation measures is not automatically tight. This is done by considering the 'distance from the boundary' as a function on $[0, \infty)$ and making use of an elegant coupling.

Chapter 5

The Restore sampler

In the previous two chapters we have focused on QSMC methods, and in particular in Chapter 4 on the method known as ReScaLE. QSMC methods have shown promise for Bayesian inference in tall data settings, however several issues remain.

For ReScaLE, one prominent issue is the necessity of storing the trajectory of the process, in order to draw samples from the empirical occupation measure. This is implemented typically using the theory of diffusion bridges; see Kumar [2019] for the details. While the resulting algorithm is entirely valid, this is the major bottleneck for current implementations of ReScaLE. This motivates the following: instead of using the empirical occupation measure, we 'freeze' the rebirth distribution, and use what we called the FR(μ) process in Chapter 4. Of course, we will need to check that we can still derive a valid sampler in the sense of π-stationarity.

The topic is this chapter is to describe when such a process can be used as a continuous-time Monte Carlo sampler and to provide some supporting theory and simple examples.

The work of this chapter was done in collaboration with my supervisors Prof. Gareth Roberts and Prof. David Steinsaltz, and in addition with Dr. Murray Pollock (University of Warwick).

5.1 Introduction

In this chapter we introduce the Restore sampler (Randomly Exploring STOchastically REnewing). The Restore sampler has three ingredients:

- (the law of) a continuous-time Markov process Y on a state space E, referred to as the *interarrival process*;
- a locally bounded nonnegative function $\kappa : E \to \mathbb{R}^+ = [0, \infty)$, the *killing rate*;

- a probability measure μ on E; the *rebirth distribution*.

The Restore process X is a continuous-time Markov process heuristically defined as follows.

Run the continuous-time Markov process Y. In addition, at rate $t \mapsto \kappa(Y_t)$, the particle teleports: the particle is instantaneously killed and then reborn at a new location drawn (independently) from μ. Then continue to evolve the continuous-time Markov process Y from there.

The Restore process is a continuous-time Markov process which combines *local dynamics* given by the stochastic process Y and *global dynamics* given by the rebirth distribution μ. Given Y and μ, we will show that in certain situations it is possible to select the killing rate κ in such a way that the resulting Restore process X will possess a given target density π as its invariant distribution. Thus the Restore process can be used as the basis for a new Monte Carlo algorithm for sampling from a given density function π.

An example of how one might simulate the Restore process via Poisson thinning, [Devroye, 1986, Chapter 6.2], in the case when the killing rate κ is uniformly bounded above by M, is given in Algorithm 2. $\mathcal{L}(Y_t|x)$ denotes the law of the process Y_t, given $Y_0 = x$. The output is a (random) collection of times, along with the value of the Restore process at those times. Values of the process at intermediate times – if desired – can typically be obtained, depending on the process Y.

Algorithm 2 Bounded Restore: $\kappa \leq M$.

initialise: $X_0 = x_0, t_0 = 0, i = 0$
while $t_i < T$ **do**
    $i \leftarrow i + 1$
    $t_i \leftarrow t_{i-1} + \tau_{i-1}$, where $\tau_{i-1} \sim \text{Exp}(M)$
    simulate $Z_i \sim \mathcal{L}(Y_{\tau_{i-1}}|Y_0 = X_{i-1})$
    with probability $1 - \kappa(Z_i)/M$
        $X_i \leftarrow Z_i$
    else
        $X_i \sim \mu$
end while
return pairs (t_i, X_i)

We see the novelty and potential of the Restore sampler lies in the following main points:

- The Restore sampler is a general recipe to combine the local dynamics given by the interarrival process Y and the global dynamics given by the rebirth

distribution μ in such a way that the resulting process has a specified invariant distribution. It is flexible, suitable for both discrete and continuous state spaces. In particular, it is able to do this in settings where *neither the local dynamics nor the global dynamics are themselves π-invariant*. For example, we will see in Section 5.6 an example where the process is a highly unstable diffusion process.

- After the work has been done in setting up the process – notably ensuring that the killing rate is nonnegative – the resulting algorithm is simple to implement. See, for instance, Algorithm 2.

- The Restore process naturally exhibits regenerations at each rebirth event, that is, times at which the process 'starts afresh'. This means that the resulting process is easy to analyse mathematically. For instance, under some additional assumptions a central limit theorem and coupling from the past implementation to obtain exact draws from π are easily derived. Practically speaking, the regenerative structure means that the process does not suffer from the issue of *burn-in* and can be straightforwardly implemented in parallel.

- The Restore sampler also provides a simple recipe for introducing rejection-free moves to existing samplers. This can be done in cases where standard MCMC algorithms may exhibit poor mixing. This is discussed in Section 5.3.3.

5.1.1 Previous work

We briefly mention some connections with existing methods; see also Sections 2.2.2, 2.2.3, 2.5.

Metropolis–Hastings algorithms, as discussed in Section 2.2.1, can typically be classified as being either *local* or *global*. Local versions are algorithms where the proposal kernel $q(x, \cdot)$ is in some sense localised around x, so the process tends to make small moves. Examples include random walk Metropolis (RWM), where $q(x, \cdot)$ is a Gaussian centered around x, or the MALA algorithm, Roberts and Rosenthal [1998]. Global algorithms, by contrast, are ones where the proposals $q(x, \cdot)$ have little dependence on x, such as the independence sampler, where the proposal kernel $q(x, \cdot) \equiv q(\cdot)$ is entirely independent of x.

Local algorithms are typically effective at exploring individual modes of the target π, but often struggle to move between modes. On the other hand, global algorithms may be able to move around the state space at large, but struggle with exploration of individual modes and may exhibit very high rejection rates. As such, local and

global algorithms are often combined to create hybrid chains, see for instance Tierney [1996], by combining several π-invariant MCMC algorithms.

By contrast, the Restore sampler gives a natural way to combine local dynamics – the Markov process Y – and global dynamics – jumps distributed according to μ – in such a way that the resulting process is π-invariant. Crucially, for Restore neither the local nor global dynamics need themselves be π-invariant, increasing the flexibility substantially.

The Restore sampler also has direct connections with many previous samplers. In the most basic case, when the interrarival process Y is constant, the Restore sampler reduces to the classical rejection sampler or importance sampler, see Section 5.4.1.

As a nonreversible continuous-time processes with a state-dependent Poisson clock, the Restore sampler has a natural affinity with the class of PDMP methods, as decribed in Section 2.2.3 and with QSMC methods, Section 2.4 and Chapter 3. In each case a state-dependent Poisson clock is used to drive the underlying Markov process towards equilibium. For PDMP methods, when the clock rings the particle's velocity jumps, resulting in a 'bounce' for the process. For QSMC methods, the particle is killed, and in the case of ReScaLE (Chapter 4) reborn according to its empirical occupation measure. For Restore, when the clock rings the particle is also killed and reborn, but according to a fixed measure μ.

Thus since the rebirth measure is fixed these times are *regeneration times*, as discussed in Section 2.2.2. Thus the Restore sampler gains the benefits which accrue from possessing such times, for instance lack of burn-in issues, parallelisability, and so on. However, if the process is constructed poorly, it may exhibit a rapid succession of short lifetimes which can pose a computational burden; its effect on the variance is studied in Section 5.4.2.

The Restore sampler can also be seen as a continuous-time version of the hybrid sampler in Murdoch [2000], where an independence sampler is combined with a random walk Metropolis sampler, which can be seen as 'global' and 'local' moves respectively. However, the crucial difference with Restore is that in our case, neither the global nor local dynamics need to be themselves π-invariant.

As an instance of a regenerative process (see Section 2.5), the Restore sampler also continues this line of inquiry into such processes. To the best of our knowledge, this is the first time a continuous-time regenerative process has been used directly in the context of Monte Carlo. As mentioned in Section 2.5, Darroch and Seneta [1965] noted that for (discrete-time, finite state space) regenerative processes, the invariant

distribution could be 'made into almost any distribution'. The work of this chapter demonstrates that for continuous time and general state spaces, this is also the case.

Finally, we reiterate the connection between the Restore sampler and the PageRank algorithm, Page et al. [1999]. Recall that for the discrete-time random PageRank surf on a finite graph G, at each step with a fixed probability $d \in [0,1]$ the surfer moves to a random uniform neighbouring vertex, otherwise teleports to a new location drawn from a fixed probability measure μ. The Restore sampler generalises this by moving into general state spaces in continuous time, and by allowing the 'teleport' rate to vary with spatial location.

We turn now to formally define the Restore process and state the main results of the chapter.

5.2 The Restore process

First we formalise the heuristic definition of the Restore process given in the introduction. We define the process first in a general, abstract framework.

Let (E, m) be a measure space, where the state space E is a Radon topological space with its Borel σ-algebra $\mathcal{E}$ and m is a Radon measure on $\mathcal{E}$. Examples of such spaces are second countable, locally compact Hausdorff spaces such as Euclidean space $\mathbb{R}^d$ equipped with Lebesgue measure. We assume that we are given a *right process* $Y = (\Omega', \mathcal{F}', \mathcal{F}'_t, Y_t, \mathbb{P}^0_x)$ evolving on E. The precise definition of a right process can be found in [Sharpe, 1988, Chapter 20]. Intuitively, right processes are an abstract class of continuous-time, right-continuous strong Markov processes, which includes most practically useful processes, such as diffusions, Feller processes, right-continuous ODE flows and jump processes.

As usual, for a general initial distribution μ, we write $\mathbb{P}^0_\mu = \int_E \mu(\mathrm{d}x)\, \mathbb{P}^0_x$.

Let $\kappa : E \to \mathbb{R}^+ = [0, \infty)$ be a locally bounded measurable function, the *killing rate*. Define the *killing time* τ_∂ as follows:

$$\tau_\partial := \inf \left\{ t \geq 0 : \int_0^t \kappa(Y_s)\, \mathrm{d}s \geq \xi \right\}, \tag{5.1}$$

where ξ is an exponential random variable with parameter 1 independent of Y. Set $\inf \varnothing = \infty$.

Fix a probability measure μ on (E, m), which will be the *rebirth distribution*. Fix an initial point $x_0 \in E$. Let $(Y^{(0)}, \tau^{(0)})$ be a realisation of (Y, τ_∂) under $\mathbb{P}^0_{x_0}$, that is, with $Y_0 = x_0$. Let $(Y^{(i)}, \tau^{(i)})$ for $i = 1, 2, \ldots$ be i.i.d. realisations of (Y, τ_∂) under $\mathbb{P}^0_\mu$, namely with $Y_0 \sim \mu$.

Set $T_0 = 0$, and for each $n = 1, 2, \ldots$, set

$$T_n = \sum_{i=0}^{n-1} \tau^{(i)}.$$

Then we define the Restore process $X = (X_t)_{t \geq 0}$ to be the process given by

$$X_t = \sum_{i=0}^{\infty} 1_{[T_i, T_{i+1})}(t)\, Y^{(i)}_{t-T_i}.$$

This defines a process $X = (\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \mathbb{P}_x)$ with state space (E, m). For its semigroup we will write $\{P^\mu_t : t \geq 0\}$. For an arbitrary initial distribution ν, as usual we set $\mathbb{P}_\nu = \int \mathrm{d}\nu(x)\mathbb{P}_x$.

We will refer to this process as the *Restore process with interarrival dynamics Y, killing rate κ and rebirth distribution μ.*

Proposition 5.2.1. *Assume that Y is a right process on the Radon space (E, m) with Radon measure m, $\kappa : E \to \mathbb{R}^+$ is a locally bounded measurable function and μ is a probability measure on E. Then the resulting Restore process $X = (\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \mathbb{P}_x)$ with interarrival dynamics Y, locally bounded nonnegative killing rate κ and rebirth distribution μ defines a right process with state space (E, m). In particular, X is right-continuous and strong Markov. Moreover, $T_n \to \infty$ almost surely.*

Proof. The killing time can be realised using the approach of [Sharpe, 1988, Chapter 61], as a subprocess generated by the decreasing right multiplicative functional m (see [Sharpe, 1988, Chapter 57]), defined by

$$m_t := \exp\left(-\int_0^t \kappa(Y_s)\,\mathrm{d}s\right), \quad t \geq 0.$$

Since Y is right-continuous and κ is locally bounded, m is indeed a right multiplicative functional. Then by Theorem 61.5 of Sharpe [1988], the killed process is another right process.

Then the above construction of X by 'gluing together' copies of the killed process can be rephrased using the technique of *concatenation of processes* in [Sharpe, 1988, Chapter 14]. Since the killed process is a right process, Exercise 14.17 of Sharpe [1988] shows that the resulting Restore process X is also a right process.

The final statement follows from the fact that we are assuming κ is locally bounded, hence $\mathbb{E}_\mu[\tau_\partial] > 0$ and it follows that $T_n \to \infty$ almost surely, since the lifetimes are independent and identically distributed. □

In order for the Restore process to be useful in the context of Monte Carlo, we would like to choose the dynamics in such a way that the resulting process has a given density function π as its invariant distribution.

5.2.1 Chapter overview

We begin by stating a general version of our invariance result, which in Section 5.3 will be specialised and proven in three particular settings: countable state spaces, for symmetric diffusions and jump processes; see Theorems 5.3.2, 5.3.7, 5.3.11. In particular the generator Q and its adjoint Q^* will be given full definitions later.

Framework result. *Assume that we are given a positive target density π with respect to m on E, a rebirth density μ with respect to m on E, and an interarrival right process Y with infinitesimal generator Q with adjoint Q^*. We assume that there is a constant $C > 0$ such that the killing rate $\kappa : E \to \mathbb{R}^+$ given by*

$$\kappa(x) := \frac{Q^*\pi(x)}{\pi(x)} + C\frac{\mu(x)}{\pi(x)}, \quad x \in E, \tag{5.2}$$

is nonnegative for each $x \in E$.

Under a range of settings and assumptions, to be detailed in Section 5.3, the resulting Restore process with interarrival dynamics Y, killing rate κ and rebirth distribution μ has invariant density π.

Remark 5.2.2. Because of the presence of the constant C, in practice we do not require μ or π to be normalised in order to compute κ.

Remark 5.2.3. Heuristically, this result is true because the generator of the Restore process L^μ is like

$$L^\mu f(x) = Lf(x) + \kappa(x) \int \big(f(y) - f(x)\big)\, \mu(\mathrm{d}y),$$

and the choice of κ in (5.2) is one that makes $(L^\mu)^*\pi = 0$. However, we have not found a single formalisation that covers all the settings of interest: discrete state, continuous jump process, and diffusion. This is because the Restore process is not necessarily a Feller process; given a function f with compact support, the semigroup $P_t^\mu(f)$ at a fixed time does not necessarily decay to zero outside compact sets, as the process can jump very quickly back to the centre of space. Hence we address these settings individually in Section 5.3. The particular theorems are formulated and proved as Theorems 5.3.2, 5.3.7, and 5.3.11.

Recall that a nonnegative random variable is *non-lattice* if it is not concentrated on a set of the form $\{\delta, 2\delta, \dots\}$ for any $\delta > 0$.

Theorem 5.2.4. *Suppose that the Restore process X, as in the conclusion of Proposition 5.2.1, is defined on a metric space E, its semigroup P_t^μ maps continuous functions to continuous functions for each $t \geq 0$, has a unique stationary distribution π, that $\mathbb{E}_\mu[\tau_\partial] < \infty$, and that the lifetimes are non-lattice. Then for any bounded function $f : E \to \mathbb{R}$, we have almost sure convergence of the ergodic averages*

$$\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s \to \pi[f],$$

as $t \to \infty$.

Proof. By Theorem 1.2 of [Asmussen, 2003, Chapter 6], and uniqueness of the stationary distribution, it follows that

$$\pi[f] = \frac{\mathbb{E}_\mu[\int_0^{\tau^{(0)}} f(X_s)\,\mathrm{d}s]}{\mathbb{E}_\mu[\tau^{(0)}]}.$$

The result then follows from standard renewal arguments; splitting f into positive and negative parts, and utilising the renewal structure of the Restore process. □

This invariance allows us to use the Restore process as the basis of a new Monte Carlo algorithm for sampling from π. Under some stronger assumptions in Section 5.4, we will establish the following.

- A *central limit theorem* for the Restore process. This is straightforwardly derived, following Hobert et al. [2002], since the Restore process naturally exhibits regeneration times, at which the subsequent evolution is independent of the past and identically distributed. This enables us to obtain variance estimates.

- A *coupling from the past* implementation, as pioneered by Propp and Wilson [1996]. When applicable, this allows us to obtain exact draws from the target π. We also show that the classical rejection sampler is a special case of this coupling from the past implementation.

In Section 5.5, we discuss some practical issues related to the simulation of Restore.

- We discuss one possible choice of rebirth distribution μ which minimises the killing rate, and the tuning of the constant C in (5.2).

- When the killing rate κ is unbounded, simulation of the event times is in general difficult. We provide a result in the diffusion setting concerning the bias incurred when implementing Restore with a *truncated* version of the killing rate.

In Section 5.6 we give some simple examples of the Restore sampler to demonstrate its potential.

5.3 Invariance

Suppose we are given a probability measure π on E, the target density. We would like to construct a Restore process X whose invariant distribution coincides with π. In this section we give conditions in several settings under which this is possible.

We consider the following settings: when (E, m) is a countable space, when the interarrival process Y is a symmetric diffusion, and when the interarrival process is a jump process.

5.3.1 Countable state space setting

Suppose E is a countable set equipped with counting measure m. This is certainly a Radon space (with the discrete topology). We view probability measures on (E, m) as (possibly infinite) row vectors, which are Radon–Nikodym derivatives with respect to m.

We assume that the underlying process Y on E is nonexplosive and is defined by a conservative rate matrix $Q^0 = (q^0(i, j) : i, j \in E)$. Such processes are certainly right-processes. We do not need to assume reversibility of Y. Let μ be the rebirth distribution.

Given a constant $C > 0$, set

$$\kappa(i) := \frac{(\pi Q^0)(i)}{\pi(i)} + C\frac{\mu(i)}{\pi(i)}, \quad i \in E. \tag{5.3}$$

Assumption 9 (Assumptions for discrete spaces)**.** *(E, m) is a countable state space equipped with counting measure m. π is a strictly positive probability density on E. The underlying process Y is irreducible and nonexplosive and is defined by a conservative rate matrix Q^0. We assume that the constant $C > 0$ in (5.3) is chosen such that $\kappa(i) \geq 0$ for all $i \in E$.*

Let us write for $i \in E$, $\lambda(i) := -q(i, i)$ for the holding rates of the Q^0 chain. Note then that we can rewrite κ as

$$\kappa(i) = \frac{\sum_{j \neq i} \pi(j) q^0(j, i) - \lambda(i)\pi(i) + C\mu(i)}{\pi(i)}, \quad i \in E. \tag{5.4}$$

Remark 5.3.1. Given this alternative expression for κ, we see that a sufficient condition for $\kappa(i) \geq 0$ is that

$$C\mu(i) \geq \lambda(i)\pi(i), \quad i \in E.$$

Alternatively, if the underlying process is already π-invariant, so $\pi Q^0 \equiv 0$, then the first term in (5.3) is identically zero and any $C > 0$ and μ are valid.

Under Assumption 9, the resulting Restore process X with interarrival dynamics Y, killing rate κ and rebirth distribution μ is another nonexplosive continuous-time Markov chain on (E, m) and we can write down its rate matrix explicitly.

Define a new transition rate matrix as follows: $Q^\kappa = (q^\kappa(i,j) : i,j \in E)$ is a rate matrix corresponding to the rebirth events given by

$$\begin{aligned} q^\kappa(i,i) &= -\kappa(i)(1-\mu(i)), \quad i \in E, \\ q^\kappa(i,j) &= \kappa(i)\mu(j), \qquad\quad i,j \in E, i \neq j. \end{aligned}$$

Then the Restore process X with interarrival dynamics Y, killing rate κ and rebirth distribution μ has explicit rate matrix given by

$$Q^\mu = Q^0 + Q^\kappa.$$

Theorem 5.3.2 (Discrete space Restore)**.** *Suppose that Assumption 9 holds. Then the Restore process X with interarrival dynamics Y, killing rate κ and rebirth distribution μ has unique invariant distribution π.*

Proof. Since Y is irreducible, it follows that the Restore process X is irreducible. We can directly verify that

$$\pi Q^\mu = 0.$$

Since X is nonexplosive, by Proposition 5.2.1, it is fully determined by its rate matrix Q^μ. Since it is also irreducible, it has unique invariant distribution π. □

5.3.1.1 Example

Suppose we are given an irreducible discrete-time stochastic matrix $P^0 = (p^0(i,j) : i,j \in E)$, with $p^0(i,i) = 0$ for each $i \in E$, and fix a target density π. We give now a recipe to construct a Restore process with π as its invariant measure.

Fix a positive rebirth distribution μ. We will use P^0 to define the jump chain of the continuous-time Markov chain Y, so we need to define the holding rates $(\lambda(i), i \in E)$. We will make use of the expression (5.4), and without loss of generality we will take

$C = 1$ (since we could just multiply the holding rates by a fixed constant). Note that $q^0(i,j) = \lambda(i)p^0(i,j)$ for each $i, j \in E$ with $i \neq j$, and $q^0(i,i) = -\lambda(i)$ for $i \in E$.

To ensure nonnegativity of κ, we choose the holding rates

$$\lambda(i) := \frac{\mu(i)}{\pi(i)}, \quad i \in E.$$

This now fully specifies the rate matrix of the chain Y. We need to assume that the chain Y is nonexplosive; this holds if for instance, the $\lambda(i)$ are uniformly bounded above or P^0 defines a recurrent discrete-time Markov chain.

In this case the killing rate κ can be written as

$$\kappa(i) = \frac{\sum_{i \neq j} \pi(j)\lambda(j)p^0(j,i)}{\pi(i)} = \frac{\sum_{j \neq i} \mu(j)p^0(j,i)}{\pi(i)}.$$

As a concrete example, take (E, m) to be the integers $\mathbb{Z}$ with counting measure. Take the interarrival process Y to be the simple symmetric random walk on $\mathbb{Z}$ in continuous time, which transitions from state $i \in \mathbb{Z}$ to states $i-1$ and $i+1$ each with rate 1, and has no other transitions. Fix a target distribution π on $\mathbb{Z}$. In this setting,

$$\pi Q^0(i) = \pi(i-1) - 2\pi(i) + \pi(i+1), \quad i \in \mathbb{Z},$$

and so the first term in (5.3) will be negative if and only if

$$\pi(i) > \frac{\pi(i+1) + \pi(i-1)}{2},$$

which will typically hold only for a finite number of $i \in \mathbb{Z}$.

In this case a straightforward sufficient condition for κ to be nonnegative is to choose a μ and C such that

$$C\mu(i) \geq 2\pi(i), \quad i \in E.$$

5.3.2 Symmetric diffusions

We now consider Restore when the underlying process is a symmetric diffusion on $E = \mathbb{R}^d$. For a smooth C^∞ function $A : \mathbb{R}^d \to \mathbb{R}$ consider the stochastic differential equation (SDE)

$$\mathrm{d}Y_t = \nabla A(Y_t)\,\mathrm{d}t + \mathrm{d}B_t, \quad Y_0 = x, \tag{5.5}$$

on $\mathbb{R}^d$ where B is a standard Brownian motion on $\mathbb{R}^d$. Define the smooth function $\gamma : \mathbb{R}^d \to \mathbb{R}$ by

$$\gamma(y) = \exp(2A(y)), \quad y \in \mathbb{R}^d,$$

and define a measure Γ on $\mathbb{R}^d$ by

$$\mathrm{d}\Gamma(y) = \gamma(y)\,\mathrm{d}y,$$

where $\mathrm{d}y$ denotes Lebesgue measure on $\mathbb{R}^d$. We are thus working on $(E, m) = (\mathbb{R}^d, \Gamma)$. This is an example of a Radon space with a Radon measure.

Assumption 10 (Underlying process)**.** *$A : \mathbb{R}^d \to \mathbb{R}$ is a smooth C^∞ function, and the SDE* (4.1) *has a unique weak solution. The process Y has a continuous symmetric transition density $p^0(t, x, y)$ on $(0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d$ with respect to Γ, which satisfies the BASSA conditions of Demuth and van Casteren [2000], which are restated in Appendix A. In particular, the diffusion is Feller, hence a right process.*

The semigroup of the diffusion Y is given for each $t \geq 0$ by

$$\mathbb{E}^0_x[f(Y_t)] = \int p^0(t, x, y) f(y)\,\mathrm{d}\Gamma(y) \tag{5.6}$$

for functions f where this integral makes sense. Under Assumption 10, the semigroup (5.6) maps $C_0(\mathbb{R}^d)$ – continuous functions vanishing at ∞ – into $C_0(\mathbb{R}^d)$ and is strongly continuous on $C_0(\mathbb{R}^d)$ with generator $-L^0$. Hence we can also write the semigroup as

$$\mathbb{E}^0_x[f(Y_t)] = [\exp(-tL^0)f](x).$$

The action of the generator on smooth compactly supported f is given by

$$-L^0 f = \frac{1}{2}\Delta f + \nabla A \cdot \nabla f.$$

Note that we are writing L^0 for *minus* the generator, as is done in Demuth and van Casteren [2000].

Under Assumption 10, the semigroup is also strongly continuous on

$$\mathcal{L}^p(\Gamma) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ measurable}, \int_{\mathbb{R}^d} |f(x)|^p\,\mathrm{d}\Gamma(x) < \infty \right\},$$

for each $1 \leq p < \infty$; see Theorem 2.5 of Demuth and van Casteren [2000], restated in Appendix A. When we want to emphasise the underlying function space we may write $-L^0_p$ for the corresponding generators on $\mathcal{L}^p(\Gamma)$ and $\mathcal{D}(L^0_p) \subset \mathcal{L}^p(\Gamma)$ for their respective dense domains.

We now assume that the target distribution and rebirth distributions are defined by density functions with respect to Γ denoted $\varphi, \mu \in \mathcal{L}^1(\Gamma)$ respectively:

Assumption 11 (Densities)**.** *The target density $\varphi \in \mathcal{L}^1(\Gamma)$, is positive on $\mathbb{R}^d$ and is twice continuously differentiable with $\int \varphi \, \mathrm{d}\Gamma = 1$. The rebirth density $\mu \in \mathcal{L}^1(\Gamma)$ and is nonnegative with $\int \mu \, \mathrm{d}\Gamma = 1$. Furthermore, $\varphi, \mu \in \mathcal{L}^2(\Gamma)$ and we have that $\varphi \in \mathcal{D}(L_2^0)$.*

Remark 5.3.3. Let us emphasise that we are writing φ and μ for densities with respect to the measure Γ, which may not necessarily be Lebesgue measure. Later on we will write $\pi := \varphi\gamma$ for the density with respect to Lebesgue measure.

For our proofs we take φ, μ to be normalised, but as noted previously this knowledge is not necessary in practice because of the constant C which appears in the killing rate.

Because L^0 is a self-adjoint operator on $\mathcal{L}^2(\Gamma)$, a sufficient condition for $\varphi \in \mathcal{D}(L_2^0)$ is that $L^0\varphi \in \mathcal{L}^2(\Gamma)$. This follows from the arguments presented in Section 3.3.3.

We can now define the killing rate κ. First, define the *partial killing rate* $\tilde{\kappa}$, via

$$\tilde{\kappa}(x) := \frac{1}{\varphi(x)} \left(\frac{1}{2} \Delta\varphi(x) + \nabla A \cdot \nabla\varphi(x) \right), \quad x \in \mathbb{R}^d.$$

This is consistent with the expressions of Chapter 3.

We define the actual killing rate κ as follows. Set for a given constant $C > 0$,

$$\kappa(x) := \tilde{\kappa}(x) + C\frac{\mu(x)}{\varphi(x)}, \quad x \in \mathbb{R}^d. \tag{5.7}$$

Assumption 12 (Killing rate)**.** *The function κ is continuous, and C is chosen such that $\kappa \geq 0$.*

Remark 5.3.4. By Proposition 4.3.12, we know that in many settings φ is quasi-stationary if and only if it is the invariant distribution of the Restore process with rebirth distribution φ. Thus our previous Assumption 4 in Chapter 3 can be seen as a special case of Assumption 12 in the case $\varphi = \mu$.

Under Assumptions 10, 11, 12, the process Y killed at rate κ (without rebirths) can be defined and analysed using Theorem 2.5 of Demuth and van Casteren [2000] (see Appendix A). This gives us a sub-Markovian semigroup $\{\exp(-tL^\kappa) : t \geq 0\}$ with symmetric, continuous kernel $p^\kappa(t, x, y)$. $p^\kappa(t, x, y)$ is thus a *sub*-density. The corresponding semigroup is strongly continuous on $C_0(\mathbb{R}^d)$ and on $\mathcal{L}^p(\Gamma)$ for any $1 \leq p < \infty$. The generator $-L^\kappa = -L^0 \dot{-} \kappa$ extends $-L^0 - \kappa$. As before, when we want to make explicit which $\mathcal{L}^p(\Gamma)$ space we are using, for $1 \leq p < \infty$, we will write $-L_p^\kappa$ for the generator of the strongly continuous semigroup on $\mathcal{L}^p(\Gamma)$, with corresponding domain $\mathcal{D}(L_p^\kappa) \subset \mathcal{L}^p(\Gamma)$.

Remark 5.3.5. It follows from Assumption 11 that $\varphi \in \mathcal{D}(L_2^\kappa)$, since both $\varphi, \mu \in \mathcal{L}^2(\Gamma)$ and formally $L^\kappa \varphi = C\mu$.

We have one final technical assumption.

Assumption 13 (Technical conditions on φ, μ)**.** *We have that*

$$\varphi \in \mathcal{D}(L_1^\kappa), \quad L_1^\kappa \varphi = C\mu. \tag{5.8}$$

Furthermore, μ is such that

$$\int \mathrm{d}\Gamma(x)\mu(x)\,\mathbb{E}_x^0\left[\sup_{t\in[0,1]}\left|\kappa(Y_t)\mathrm{e}^{-\int_0^t \kappa(Y_s)\,\mathrm{d}s}\right|\right] < \infty. \tag{5.9}$$

The condition (5.8) is fairly abstract, and so might be difficult to verify in a particular case, or in a general class of processes that one may want to consider. For many purposes we may replace it with the following sufficient condition: Set

$$\pi := \varphi\gamma.$$

We write $W^{2,1}(\mathbb{R}^d)$ for the Sobolev space of measurable functions on $\mathbb{R}^d$ whose first and second derivatives are integrable with respect to Lebesgue measure on $\mathbb{R}^d$.

Lemma 5.3.6. *Assume that Assumptions 10, 11, 12 hold. Suppose that the drift is at most linear in the tails: we can bound $|\nabla A(x)| \le K|x|$, for x outside some compact set, some $K > 0$. Assume that Assumption 10 holds. Suppose φ is smooth, and that $\pi \in W^{2,1}(\mathbb{R}^d)$. In addition, we require that*

$$\int_{\mathbb{R}^d} |\nabla A(x)\cdot\nabla\pi(x)|\,\mathrm{d}x < \infty, \int_{\mathbb{R}^d} |\Delta A(x)\pi(x)|\,\mathrm{d}x < \infty.$$

Then (5.8) *holds.*

Proof. See Appendix A □

Alternatively, another sufficient condition for (5.8) that Γ is a finite measure. This is the case whenever the underlying diffusion Y is positive recurrent, say a stable Ornstein–Uhlenbeck process. Then under $\varphi \in \mathcal{L}^2(\Gamma)$ and $\mu \in \mathcal{L}^2(\Gamma)$, $\varphi \in \mathcal{D}(L_1^\kappa)$ with $L_1^\kappa\varphi = C\mu$ is automatic, since in that case $\mathcal{L}^2$ convergence implies $\mathcal{L}^1$ convergence.

The condition (5.9) is needed so that we can differentiate under the integral. A necessary condition for (5.9) to hold is that $\int \mathrm{d}\Gamma(x)\mu(x)\kappa(x) < \infty$, so in particular μ cannot have tails which are too heavy relative to φ. From a computational point

of view, this is reasonable since otherwise the rebirth mechanism would be highly inefficient; the Restore process would tend to be killed very rapidly. Of course, a sufficient condition for (5.9) is that

$$\int \mathrm{d}\Gamma(x)\mu(x)\,\mathbb{E}_x^0\left[\sup_{t\in[0,1]}\kappa(Y_t)\right] < \infty.$$

Theorem 5.3.7. *Under Assumptions 10, 11, 12, 13, the Restore process X with interarrival dynamics Y, killing rate κ and rebirth distribution μ has invariant distribution φ.*

Proof. See Appendix A. □

5.3.2.1 Examples

We now give some examples of diffusions which satisfy the assumptions of Theorem 5.3.7.

Sufficient conditions ensuring BASSA (Assumption 10, Appendix A) are given in Example 2 of [Demuth and van Casteren, 2000, Chapter 1.C]. In our present setting when we consider diffusions defined by (4.1), these conditions can be written as

$$\exp(A(x)) \geq c^{-1}\exp(-c|x|^2), \quad \forall x \in \mathbb{R}^d, \tag{5.10}$$

$$c^{-1} \leq \exp(A(x) - A(y)) \leq c, \quad \forall x, y \in \mathbb{R}^d : |x-y| \leq c^{-1}(1+|x|)^{-c}, \tag{5.11}$$

for some $c > 0$.

Let $|\cdot|$ denote the ℓ_2 norm on $\mathbb{R}^d$.

Proposition 5.3.8. *The SDE* (4.1) *with $A = \alpha|x|^2$ for any $\alpha \in \mathbb{R}$ satisfies BASSA.*

Remark 5.3.9. In this case $\nabla A(x) = 2\alpha x$ is linear. $\alpha < 0$ corresponds to a (stable) Ornstein–Uhlenbeck process, $\alpha = 0$ is a Brownian motion and $\alpha > 0$ is an *unstable* Ornstein–Uhlenbeck process which drifts into the tails.

Proof. (5.10) clearly holds in this setting. The second condition (5.11) can be seen from the reverse triangle inequality:

$$\begin{aligned}\big||x|^2 - |y|^2\big| &= (|x|+|y|)\,\big||x|-|y|\big| \leq (|x|+|y|)|x-y| \\ &\leq \frac{|x|+|y|}{c(1+|x|)^c} \leq \frac{2|x|}{c(1+|x|)^c} + \frac{1}{c^2(1+|x|)^{2c}}.\end{aligned}$$

This is uniformly bounded over $x \in \mathbb{R}^d$ for $c > 1$. □

5.3.3 Jump processes

The Restore process is inherently a continuous-time process, and so the underlying process Y must be a continuous-time object. Suppose, however, we are given a a discrete-time Markov transition kernel P on (E, m), with action on measurable functions $f : E \to \mathbb{R}$ and measures ν on E given by

$$\begin{aligned} Pf(x) &= \int f(y)p(x,y)\,\mathrm{d}m(y), \quad x \in E, \\ \nu P(\mathrm{d}y) &= \int \nu(\mathrm{d}x)\,p(x,y)\,\mathrm{d}m(y), \end{aligned} \tag{5.12}$$

for some integral kernel $p(x, y)$ on $E \times E$, whenever these integrals make sense. Since we have an integral kernel $p(x, y)$, we will also think of νP as a measurable function given by

$$\nu P(y) := \int \nu(\mathrm{d}x)\,p(x,y)$$

for a measure ν on E, provided this makes sense.

It is straightforward to embed P into continuous time, by specifying a measurable function $\lambda : E \to \mathbb{R}^+$, the *holding rates*, as was seen in Section 5.3.1. We take the jump chain to be defined by the discrete-time Markov kernel P, and just take the holding times to be independent $\mathrm{Exp}(\lambda(x))$ times, when currently at state x.

Such a process will be a *continuous-time jump process* on E, meaning it has right-continuous, piecewise-constant sample paths. Provided they are nonexplosive, such processes are determined by the transition kernel of the jump chain and the holding rates. See for instance, [Ethier and Kurtz, 1986, Chapter 4.2].

Suppose π, μ are two densities on E with respect to m, the target density and rebirth density respectively, where we assume π is positive. Suppose we are given a transition kernel P on E and holding rates $\lambda : E \to \mathbb{R}^+$. We now construct the Restore process. Given a constant C, define the killing rate κ to be

$$\kappa(x) := \frac{\int m(\mathrm{d}y)\pi(y)\lambda(y)p(y,x) - \lambda(x)\pi(x) + C\mu(x)}{\pi(x)}, \quad x \in E. \tag{5.13}$$

Assumption 14 (Jump process Restore)**.** *P is a transition kernel with a density as in (5.12), $\lambda : E \to \mathbb{R}^+$ is measurable, strictly positive. π is a positive probability density with respect to m, μ is a probability density with respect to m, and $\int \lambda(x)\pi(x)m(\mathrm{d}x) < \infty$. The constant C is such that $\kappa \geq 0$ on E. κ is locally bounded, and we have that $\int(\lambda(x) + \kappa(x))^2\pi(x)m(\mathrm{d}x) < \infty$.*

Note that such jump processes are right processes (Exercise 14.18 of Sharpe [1988]).

This construction can also be extended to kernels P which do not possess a density as in (5.12). For example, the classical Metropolis–Hastings kernel is of the form

$$P(x, \mathrm{d}y) = \alpha(x, y)q(x, y)\,\mathrm{d}m(y) + (1 - j(x))\delta_x(\mathrm{d}y),$$

where $0 \leq \alpha(x, y) \leq 1$ are the acceptance probabilities, $q(x, y)$ is a transition density (so $\int q(x, y)\,\mathrm{d}m(y) = 1$ for each $x \in E$), and

$$j(x) := \int \alpha(x, y)q(x, y)\,\mathrm{d}m(y)$$

are the jump probabilities. Because of the presence of the delta mass $\delta_x(\mathrm{d}y)$, such kernels cannot possess straightforward densities. However in continuous-time, these rejected moves associated with the delta mass are not visible, and so we can modify the killing rate as follows: we replace the term $(\pi\lambda)P(x)$ in (5.13) by

$$\int \mathrm{d}m(y)\,\pi(y)\lambda(y)\alpha(y, x)q(y, x) + \lambda(x)(1 - j(x))\pi(x).$$

Returning to the construction of the Restore process, we will take the interarrival dynamics to be given by the jump process defined by P and λ, the killing rate to be κ and the rebirth distribution μ. The resulting Restore process X is another continuous-time jump process, and so we describe its jump chain and holding rates. This will provide a method to simulate the process.

At $x \in E$, the transition kernel $P^\mu(x, \mathrm{d}y)$ of the jump chain is given by

$$P^\mu(x, \mathrm{d}y) = \frac{\lambda(x)}{\lambda(x) + \kappa(x)}P(x, \mathrm{d}y) + \frac{\kappa(x)}{\lambda(x) + \kappa(x)}\mu(y)\,\mathrm{d}m(y).$$

The overall holding rates in continuous time are given by

$$\bar{\lambda}(x) = \lambda(x) + \kappa(x), \quad x \in E,$$

that is, at $x \in E$, by the Markov property, the time until the next jump is an $\mathrm{Exp}(\bar{\lambda}(x))$ time.

Remark 5.3.10. Rather than working in terms of densities, and a killing function κ, a more general approach would be to make use of a killing measure. We do not consider this particular generalisation in this work.

Theorem 5.3.11. *Assume that Assumption 14 holds, and that the interarrival dynamics defined by P and λ are nonexplosive. Then the resulting Restore process X is a nonexplosive jump process with invariant distribution π.*

Proof. Nonexplosivity follows from Proposition 5.2.1, and the fact that the interarrival process is assumed nonexplosive.

Let us write $\{Q_t^\mu : t \geq 0\}$ for the continuous-time semigroup for the Restore process X. Our goal is to show that for any continuous bounded function $f : E \to \mathbb{R}$, $\pi Q_t^\mu f = \pi(f)$, for each $t \geq 0$. To do this we compute the time derivative of the mapping $t \mapsto \pi Q_t^\mu f$ and show that it is 0. By time-homogeneity and the semigroup property, it is sufficient to compute this derivative at $t = 0$. This was the approach similarly used to prove π-invariance of the Bouncy Particle Sampler in the supplementary material of Bouchard-Côté et al. [2018].

By conditioning on the first jump, we obtain the following representation (*cf.* equation (4.24) of Moyal [1957]),

$$\begin{aligned} Q_t f(x) &= \mathrm{e}^{-\bar{\lambda}(x)t} f(x) + \int_0^t \mathrm{d}s\, \bar{\lambda}(x) \mathrm{e}^{-\bar{\lambda}(x)s} P^\mu[Q_{t-s} f](x) \\ &= \mathrm{e}^{-\bar{\lambda}(x)t} f(x) + \int_0^t \mathrm{d}s\, \bar{\lambda}(x) \mathrm{e}^{-\bar{\lambda}(x)(t-s)} P^\mu[Q_s f](x). \end{aligned}$$

From this representation we can calculate the derivative,

$$\begin{aligned} \frac{\mathrm{d}Q_t f(x)}{\mathrm{d}t} &= -\bar{\lambda}(x) \mathrm{e}^{-\bar{\lambda}(x)t} f(x) + \bar{\lambda}(x) P^\mu[Q_t f](x) \\ &\quad - \int_0^t \mathrm{d}s\, \bar{\lambda}^2(x) \mathrm{e}^{-\bar{\lambda}(x)(t-s)} P^\mu[Q_s f](x). \end{aligned}$$

From this we can easily calculate that at $t = 0$, $\frac{\mathrm{d}}{\mathrm{d}t} \pi Q_t f = 0$, using the definitions of $\bar{\lambda}$ and κ. The exchange of integration and differentiation is justified since we are assuming that $\pi(\bar{\lambda}^2) < \infty$. □

For applications we may further want to know when our sample paths can be used to approximate the stationary distribution.

Theorem 5.3.12. *Assume that Assumption 14 holds. In addition, suppose that the kernel P is Harris recurrent, and that the sum $\lambda + \kappa$ is uniformly bounded away from zero. Then for bounded measurable f, the ergodic averages $t^{-1} \int_0^t f(X_s)\, \mathrm{d}s$ converge almost surely to $\pi(f)$.*

The definition of Harris recurrence is given in [Meyn and Tweedie, 1993, Chapter 9].

Proof. Consider the probability density function with respect to m given by

$$\xi(x) = Z^{-1} \pi(x)(\lambda(x) + \kappa(x)), \quad x \in E,$$

where the normalising constant $Z := \int \pi\lambda \, \mathrm{d}m + C$. It is easy to see that this indeed defines a probability density, since

$$\int \kappa(x)\pi(x) \, \mathrm{d}m(x) = C.$$

A straightforward calculation then shows that for the transition kernel of the jump chain P^μ,

$$\xi P^\mu = \xi.$$

That is, ξ is an invariant distribution for the jump chain. Since the kernel P is Harris recurrent, the jump chain of X, defined by the kernel $P^\mu(x, \mathrm{d}y)$, is also Harris recurrent. This is because writing τ_A for the first return time to A and τ_∂ for the first rebirth event for the P^μ jump chain, we have that

$$\mathbb{P}_\mu(\tau_A < \tau_\partial) > 0,$$

provided A has positive mass, since P is assumed Harris recurrent and the killing rate κ is locally bounded. This entails then that from any initial position, hitting the set A is certain, implying Harris recurrence (see Proposition 9.1.1 of Meyn and Tweedie [1993]).

Since ξ is invariant for the jump chain, which is Harris recurrent, ξ also defines the almost sure limit of the ergodic averages of the jump chain, see Theorem 17.1.7 of Meyn and Tweedie [1993].

Moving to continuous time, since the process is nonexplosive, it follows that $\pi \propto \xi/\bar{\lambda}$ must be the almost sure limit of the ergodic averages for the continuous-time process. One way to see this is to decomponse the ergodic averages into sums of the form

$$\frac{1}{T_n} \sum_{m=0}^{n} f(X_m)\tau_m = \frac{1}{T_n} \sum_{m=0}^{n} \frac{f(X_m)}{\bar{\lambda}(X_m)} E_m,$$

where (X_m) denotes the jump chain, T_n are the jump times, $\tau_m \sim \mathrm{Exp}(\bar{\lambda}(X_m))$ are the holding times and (E_m) are an i.i.d. sequence of Exp(1) random variables. Consider then the augmented discrete-time Markov chain $(X_n, E_n)_{n\geq 0}$ with state space $E \times \mathbb{R}^+$. This chain is positive Harris, with invariant distribution given by $\xi(x)m(\mathrm{d}x) \otimes \mathrm{e}^{-y} \, \mathrm{d}y$ on $E \times \mathbb{R}^+$, $\mathrm{d}y$ denoting Lebesgue measure on $\mathbb{R}^+$. We can then apply the ergodic theorem (Theorem 17.1.7, Meyn and Tweedie [1993]) with function $g : E \times \mathbb{R}^+ \to \mathbb{R}$ given by $g(x, y) = (f(x)/\bar{\lambda}(x))y$.

□

In this setting, practical simulation of the Restore process is straightforward, *even when the killing rate is unbounded*, since the interarrival process Y is piecewise-constant. See Algorithm 3 for one possible implementation.

Algorithm 3 Jump process Restore.

initialise: $X_0 = x_0, t_0 = 0, i = 0$
while $t_i < T$ **do**
    $i \leftarrow i + 1$
    simulate $\tau_{i-1}^{(1)} \sim \mathrm{Exp}(\lambda(X_{i-1}))$, $\tau_{i-1}^{(2)} \sim \mathrm{Exp}(\kappa(X_{i-1}))$
    $\tau_{i-1} \leftarrow \tau_{i-1}^{(1)} \wedge \tau_{i-1}^{(2)}$
    $t_i \leftarrow t_{i-1} + \tau_{i-1}$
    if $\tau_{i-1}^{(1)} < \tau_{i-1}^{(2)}$ **then**
        $X_i \sim P(X_{i-1}, \cdot)$
    else
        $X_i \sim \mu$
end while
return pairs (t_i, X_i)

Algorithm 3 can be seen as a continuous-time variant of standard Metropolis–Hastings; at each iteration we 'propose' a move according to $P(X_{i-1}, \cdot)$, which is either accepted or rejected, depending on two exponential clocks. Upon rejecting a move, rather than remaining at X_{i-1} instead we move to a new location drawn from μ.

Before turning to an example, we close this section with a discussion of a technique presented in Douc and Robert [2011], which bears similarities with our present method. The approach of Douc and Robert [2011] is to rewrite the ergodic average estimate of a test function h arising from an MH chain $(X_i)_{i=1}^N$ as

$$\frac{1}{N} \sum_{i=1}^{M} \mathfrak{n}_i h(\mathfrak{z}_i),$$

where $(\mathfrak{z}_i)_{i=1}^M$ corresponds to the 'jump chain' of accepted moves of the MH chain, and $(\mathfrak{n}_i)_{i=1}^M$ are (conditionally) geometric random variables, corresponding to the number of times a proposed move from each $\mathfrak{z}_i$ is rejected.

This is similar to our approach, in the sense that we begin with a discrete-time process, with kernel P, which may not be π-invariant (much like their $\mathfrak{z}_i$), and we make use of weights to correct this bias. Although we work in continuous time rather than discrete time, our weights are holding times, hence (conditionally) exponentially distributed, akin to the $(\mathfrak{n}_i)$ which are (conditionally) geometricially distributed.

However, a crucial difference is that our approach is more flexible, in that we do not assume that the jump chain P is necessarily arising as the jump chain of some given MH kernel. To compensate for this fact, we additionally introduce regenerations into the process.

With regards to the main contribution of Douc and Robert [2011] in terms of variance reduction, it is an interesting question for future research whether or not an analogous 'Rao–Blackwellization' technique could be applied to Restore and subsequently attain a reduced variance.

5.3.3.1 Example

An example where Assumption 14 is easily checked is when P corresponds to a Markov chain which is already π-invariant, for instance the kernel of an appropriate MCMC algorithm targeting π. In this case we can easily embed P into continuous time without changing the asymptotic dynamics; just take the holding rates $\lambda \equiv 1$.

In this case the killing rate reduces to

$$\kappa(x) = C\frac{\mu(x)}{\pi(x)}, \quad x \in E,$$

and we see that any choice of $C > 0$ will ensure nonnegativity of κ. This gives a recipe to introduce rejection-free moves to a discrete sampler in continuous time.

For example, suppose we are in a setting where the target π is multimodal, and we have a rough idea of where the modes are (e.g. through preliminary runs of some Monte Carlo method), but our sampler struggles to move between the modes and find the relative weights. Then we could take μ to be supported on the modes we have found, and the resulting Restore sampler would then jump between them. For an example of this, see Section 5.6.3.

5.4 Limiting properties

In this section we consider some limiting properties of the Restore process. We will not *a priori* assume that X has invariant distribution π, but will work in the abstract framework of Proposition 5.2.1.

5.4.1 Coupling from the past

Recall the general Restore setting of Proposition 5.2.1: the underlying process Y is a right process evolving on a Radon space (E, m), we have a locally bounded measurable

function $\kappa : E \to \mathbb{R}^+$ and we a probability measure μ on E. We consider the Restore process X with these dynamics. We will write $\{P_t^\mu : t \geq 0\}$ for its semigroup.

Under the following (fairly strong) assumption, we will have direct access to the stationary distribution of the Restore process.

Assumption 15 (Coupling from the past)**.** *There exists some $\underline{\kappa} > 0$ such that m-almost everywhere,*

$$\kappa \geq \underline{\kappa} > 0.$$

We write $\|\cdot\|_\infty$ for the sup norm of a bounded function and $\|\cdot\|_1$ for the total variation norm signed measures; given a signed measure ν,

$$\|\nu\|_1 = \sup\{|\nu(f)| : f \text{ bounded, measurable }, \|f\|_\infty \leq 1\}.$$

Proposition 5.4.1 (Uniform ergodicity)**.** *Assume the basic conditions of Proposition 5.2.1 hold, and that X is irreducible. Under Assumption 15, the Restore process X is uniformly ergodic in that there exists a unique invariant distribution π such that*

$$\|\nu P_t^\mu - \pi\|_1 \leq 2\mathrm{e}^{-t\underline{\kappa}},$$

for any initial distribution ν and $t \geq 0$.

Proof. Fix any two arbitrary initial distributions ν_1, ν_2 on E. Then we can easily couple two copies of the Restore process X with initial distributions ν_1 and ν_2 respectively which meet at the first arrival time of a homogeneous Poisson process of rate $\underline{\kappa}$ and evolve identically thereafter. Thus by the well-known coupling inequality,

$$\|\nu_1 P_t^\mu - \nu_2 P_t^\mu\|_1 \leq 2\mathrm{e}^{-t\underline{\kappa}}. \tag{5.14}$$

The Markov property (i.e. the semigroup property) then shows that for any initial distribution ν, $(\nu P_t^\mu)_{t\geq 0}$ forms a Cauchy sequence in the space of probability measures equipped with the total variation norm. By completeness, there exists a limiting probability distribution π, which must also be a stationary distribution, by the Markov property and the fact that P_t^μ is a contraction in $\|\cdot\|_1$. That is, we have that $\pi P_t^\mu = \pi$ for any $t \geq 0$. By irreducibility, this invariant distribution is unique. Thus taking $\nu_1 = \nu$ and $\nu_2 = \pi$ in (5.14) the Proposition is proven. □

In fact under Assumption 15 we can do even better than uniform ergodicity and employ Coupling From the Past (CFTP), a technique pioneered by Propp and Wilson [1996] to obtain exact draws from the stationary distribution π.

Theorem 5.4.2 (Coupling from the past)**.** *Under the conditions of Proposition 5.2.1 and Assumption 15, consider the Restore process X with interarrival dynamics Y, modified killing rate*

$$\kappa' := \kappa - \underline{\kappa} \geq 0,$$

and rebirth distribution μ. Suppose X is irreducible and has initial distribution

$$X_0 \sim \mu.$$

Let $T \sim Exp(\underline{\kappa})$ be independent of X. Then

$$X_T \sim \pi,$$

where π is the unique invariant distribution of the process.

Proof. This follows from the technique of coupling from the past Propp and Wilson [1996]. For the Restore process with killing rate κ, at the arrival times of a homogeneous Poisson process of rate $\underline{\kappa}$ we can couple all processes started from all initial positions. Since the time reverse of a homogeneous Poisson process is also a homogeneous Poisson process, we can instead imagine initialising $X_0 \sim \mu$ and evolving an exponential time T into the future with modified killing rate κ'. □

In the case when κ is bounded above, one implementation is given in Algorithm 4 below.

Algorithm 4 Bounded Restore: $\kappa \leq M$, with CFTP.

draw run time: $T \sim \text{Exp}(\underline{\kappa})$
initialise: $X_0 \sim \mu, t_0 = 0, i = 0$
$i \leftarrow i + 1$
$t_i \leftarrow t_{i-1} + \tau_{i-1}$, where $\tau_{i-1} \sim \text{Exp}(M - \underline{\kappa})$
while $t_i < T$ **do**
    simulate $Z_i \sim \mathcal{L}(Y_{\tau_{i-1}} | X_{i-1})$
    with probability $1 - (\kappa(X_i) - \underline{\kappa})/(M - \underline{\kappa})$
        $X_i \leftarrow Z_i$
    else
        $X_i \sim \mu$
    $i \leftarrow i + 1$
    $t_i \leftarrow t_{i-1} + \tau_{i-1}$, where $\tau_{i-1} \sim \text{Exp}(M - \underline{\kappa})$
end while
simulate $Z \sim \mathcal{L}(Y_{T - t_{i-1}} | X_{i-1})$
return Z, which is drawn exactly from π

This CFTP implementation can be seen as a continuous-time version of the multigamma coupler of Murdoch and Green [1998] or of the hybrid scheme of [Murdoch, 2000, Section 3]. The multigamma coupler is a discrete-time MCMC algorithm that can be run when there is a uniform minorising function r for the proposal density; $f(y|x) \geq r(y)$ for all x, y. Then at each iteration, with a fixed probability $\rho := \int r(y)\,\mathrm{d}y$, all points are coupled together and drawn from $r(\cdot)$.

However unlike the multigamma coupler, here we are not identifying (via a uniform minorisation condition) times at which an existing sampler couples all paths, but rather building a new process with this property. The novelty here lies in that we are theoretically able to *choose* the rebirth distribution μ, subject to the assumptions detailed above. By contrast, in for the multigamma coupler and related methods, the regenerative distribution (denoted r in Murdoch and Green [1998]) is typically unknown or difficult to sample from. The approach of Murdoch [2000] is essentially to alternate between two different MCMC dynamics, such as random walk Metropolis–Hastings and the Independence sampler, in order to induce uniform ergodicity and gain the benefits from both local and global moves. This is not dissimilar to Restore, however with Restore we do not require the two dynamics to be individually π-invariant.

5.4.1.1 Example: classical rejection sampler

We show that the classical rejection sampler can be seen as a special case of the CFTP implementation of the Restore process with trivial interarrival process. A similar result can be established for the independence sampler version of the Metropolis–Hastings algorithm, as is shown in Murdoch and Green [1998].

Let π, μ be density functions on E with respect to m. We take Y to be the trivial stochastic process on E which given its initial position Y_0, has constant sample paths: almost surely, $Y_t = Y_0$ for all $t \geq 0$. Define the killing rate

$$\kappa(x) = C\frac{\mu(x)}{\pi(x)}, \quad x \in E, \tag{5.15}$$

for any constant $C > 0$. If we were to implement the classical rejection sampler targeting π from μ we would require the following condition:

$$\pi(x) \leq M\mu(x), \quad x \in E, \tag{5.16}$$

for some (finite) constant M. The classical rejection sampler targeting π from μ repeatedly draws X_n independently from μ, and accepts it with probability

$$\pi(X_n)/(M\mu(X_n)),$$

otherwise rejects it and tries again with a new $X_{n+1} \sim \mu$. The final accepted value X_n is an exact draw from π.

Theorem 5.4.3. *Under* (5.16)*, the CFTP implementation of the Restore process (Theorem 5.4.2) with constant interarrival dynamics, killing rate* κ *as in* (5.15) *and rebirth distribution* μ *is identical to classical rejection sampling targeting* π *from* μ*.*

Proof. We see that (5.16) holds if and only if Assumption 15 holds with

$$\underline{\kappa} = C/M.$$

Under this condition in the CFTP implementation (Theorem 5.4.2) we run the Restore process with killing rate

$$\kappa' = \kappa - \underline{\kappa} = C\frac{\mu}{\pi} - \frac{C}{M}$$

for a time $T \sim \text{Exp}(C/M)$.

We can simulate this Restore process iteratively by drawing for each $n = 1, 2, \ldots,$ $X_n \sim \mu$. We have two competing independent exponential clocks, $T \sim \text{Exp}(C/M)$ and $T_n \sim \text{Exp}(\kappa'(X_n))$.

If $T < T_n$, all trajectories have coupled and so we terminate the algorithm and output X_n, which is an exact draw from π. By the theory of competing exponentials this occurs with probability

$$\frac{C/M}{C/M + C\left(\frac{\mu(X_n)}{\pi(X_n)} - \frac{1}{M}\right)} = \frac{\pi(X_n)}{M\mu(X_n)}.$$

This is exactly the probability of acceptance for the classic rejection sampler.

If $T \geq T_n$ then we iterate again and draw $X_{n+1} \sim \mu$. By the memoryless property of the exponential distribution we have again two independent exponential clocks as before. □

If (5.16) doesn't hold, provided there is a unique invariant distribution π we can still use ergodic averages to estimate $\pi(f)$ for any bounded f. Suppose we run the Restore process with constant interarrival dynamics, killing rate κ as in (5.15) and rebirth distribution μ for n complete lifetimes. The corresponding ergodic average is

$$\frac{1}{T_n}\sum_{i=1}^{n} f(X_i)\tau^{(i)},$$

where $X_i \sim \mu$ are i.i.d., conditional on X_i, $\tau^{(i)} \sim \text{Exp}(C\mu(X_i)/\pi(X_i))$ are independent and $T_n = \sum_{i=1}^{n} \tau^{(i)}$. Thus the estimator of $\pi(f)$ can be seen as an importance sampling–type estimator with randomised importance weights; note $C\mathbb{E}[\tau^{(i)}|X_i] = \pi(X_i)/\mu(X_i)$.

5.4.2 Central Limit Theorem

In this subsection we give a central limit theorem for the Restore process. Our approach here is inspired by Hobert et al. [2002], who considered regenerative methods for MCMC (in discrete time).

We fix a measurable function $f : E \to \mathbb{R}$.

Assumption 16 (Central limit theorem)**.** *We assume the basic conditions of Proposition 5.2.1. Furthermore we assume that X is irreducible,*

$$\mathbb{E}_\mu[\tau_\partial^2] < \infty, \tag{5.17}$$

and that our function $f : E \to \mathbb{R}$ satisfies

$$\mathbb{E}_\mu\left[\left(\int_0^{\tau_\partial} f(X_s)\,\mathrm{d}s\right)^2\right] < \infty.$$

A sufficient condition for Assumption 16 to hold is that f is a bounded function and we have simply the second moment condition (5.17). In turn, a sufficient condition for (5.17) is that Assumption 15 holds, since in that case τ_∂ can be stochastically dominated by an $\mathrm{Exp}(\underline{\kappa})$ random variable.

Under Assumption 16 we will see that a central limit theorem holds. This can be easily done since the lifetimes of the Restore process, by construction, are independent and identically distributed.

As in the construction of Restore in Section 5.2, set $T_0 = 0$, let (T_n) be the successive rebirth times and let $(\tau^{(i)})$ be the lifetimes. We take the initial distribution $X_0 \sim \mu$. Set for each $i = 0, 1, 2, \ldots,$

$$Z_i := \int_{T_i}^{T_{i+1}} f(X_s)\,\mathrm{d}s.$$

By construction the (Z_i) are independent and identically distributed, with finite first and second moments.

We can apply the strong law of large numbers to the following numerator and denominator:

$$\frac{\int_0^{T_n} f(X_s)\,\mathrm{d}s}{T_n} = \frac{\sum_{i=0}^{n-1} Z_i}{\sum_{i=0}^{n-1} \tau^{(i)}} \to \frac{\mathbb{E}_\mu\left[\int_0^{\tau^{(1)}} f(X_s)\,\mathrm{d}s\right]}{\mathbb{E}_\mu[\tau^{(1)}]}$$

almost surely as $n \to \infty$.

Let us write

$$\pi(f) := \frac{\mathbb{E}_\mu\left[\int_0^{\tau^{(1)}} f(X_s)\,\mathrm{d}s\right]}{\mathbb{E}_\mu[\tau^{(1)}]}.$$

When the process is ergodic, this corresponds to the invariant distribution of the Restore process. It follows immediately that the random variables

$$Z_i - \tau^{(i)}\pi(f), \quad i = 0, 1, 2, \ldots$$

are independent and identically distributed and have mean 0 under $\mathbb{E}_\mu$.

Now we set, in analogue with the expression given in Hobert et al. [2002],

$$\sigma_f^2 := \frac{\mathbb{E}_\mu\left[\left(Z_1 - \tau^{(1)}\pi(f)\right)^2\right]}{\left(\mathbb{E}_\mu[\tau^{(1)}]\right)^2}. \tag{5.18}$$

This numerator is finite by Assumption 16.

Theorem 5.4.4 (Central limit theorem)**.** *We have that*

$$\sqrt{n}\left(\frac{\int_0^{T_n} f(X_s)\,\mathrm{d}s}{T_n} - \pi(f)\right) \xrightarrow{d} N\left(0, \sigma_f^2\right). \tag{5.19}$$

Proof. The left-hand side of (5.19) can be written

$$\sqrt{n}\left(\frac{\sum_{i=0}^{n-1} Z_i}{\sum_{i=0}^{n-1}\tau^{(i)}} - \pi(f)\right) = \frac{1}{\sqrt{n}} \cdot \frac{n}{\sum_{i=0}^{n-1}\tau^{(i)}}\left(\sum_{i=0}^{n-1}\left(Z_i - \tau^{(i)}\pi(f)\right)\right).$$

By the strong law of large numbers and the continuous mapping theorem, $n/\sum_{i=0}^{n-1}\tau^{(i)}$ converges almost surely to $(\mathbb{E}_\mu[\tau^{(i)}])^{-1}$, and in distribution also.

Hence by applying Slutsky's lemma and the central limit theorem to the independent and identically distributed mean zero random variables $(Z_i - \tau^{(i)})$, we see that (5.19) holds. □

Let us write $\bar{\tau}_n := n^{-1}\sum_{i=0}^{n-1}\tau^{(i)}$ and $\bar{f}_n := \frac{\int_0^{T_n} f(X_s)\,\mathrm{d}s}{T_n}$. Similar to Hobert et al. [2002], our σ_f^2 can be consistently estimated by

$$\hat{\sigma}_f^2 := \frac{\sum_{i=0}^{n-1}\left(Z_i - \bar{f}_n\tau^{(i)}\right)^2}{n\bar{\tau}_n^2}.$$

This is because the difference between $\hat{\sigma}_f^2$ and

$$\frac{\sum_{i=0}^{n-1}\left(Z_i - \tau^{(i)}\pi(f)\right)^2}{n\bar{\tau}_n^2}$$

converges to zero almost surely as $n \to \infty$, and the latter is a consistent estimator for σ_f^2.

We can use this to get an estimate of the efficiency of Restore. If we let

$$v_\pi(f) := \int \big(f(x) - \pi(f)\big)^2 \, \mathrm{d}\pi(x),$$

then we can set the effective sample size n_{eff} to be

$$n_{\text{eff}} := \frac{v_\pi(f)}{\sigma_f^2},$$

which we may be able to estimate.

We see from (5.18), that the denominator $(\mathbb{E}_\mu[\tau^{(1)}])^2$ will have a significant influence on the overall variance. If $\mathbb{E}_\mu[\tau^{(1)}]$ is small, the resulting variances of individual lifetimes may be unacceptably large, and as such practically speaking it is important to choose the rebirth distribution in such a way that the lifetimes are (on average) not too short. In particular, this means aiming to avoid regions of particularly high killing.

5.5 Simulation

We consider now some practical questions related to the Restore process.

5.5.1 Minimal rebirth distribution

In this section we assume that we are given some fixed interarrival process, a positive target density π on E and a rebirth density μ on E, which are both normalised.

The most difficult aspect of implementing the Restore sampler is checking that the killing rate is nonnegative, as in Assumption 12; we need for some $\underline{\kappa} \geq 0$

$$\kappa(x) = \tilde{\kappa}(x) + C\frac{\mu(x)}{\pi(x)} \geq \underline{\kappa} \geq 0, \quad x \in E, \tag{5.20}$$

where $\tilde{\kappa}$ is a function mapping $E \to \mathbb{R}$ which is specified by the underlying dynamics of the interarrival process.

In order to satisfy (5.20), given the target π and $\tilde{\kappa}$ we must choose an appropriate rebirth distribution μ and corresponding constant C. Heuristically, $\tilde{\kappa}(x)$ corrects for the local trend in $\pi(Y_{t+\delta t})$ when $Y_t = x$. As such, $\tilde{\kappa}$ will be negative around the modes of π (this is essentially stating the maximum principle for the generator of Y). Thus the support of μ must include neighbourhoods of the mode(s) of π.

One natural choice of rebirth distribution is to minimize the number of rebirth events by choosing the *minimal* μ that makes (5.20) possible. Namely, we would like to choose some minimal rebirth distribution μ^* and corresponding constant C^* such that the killing rate is given by

$$\kappa^* := \tilde{\kappa} + C^* \frac{\mu^*}{\pi} = \tilde{\kappa} \vee \underline{\kappa}. \tag{5.21}$$

This is entirely analogous to the choice of bounce rate for the Bouncy Particle Sampler of Bouchard-Côté et al. [2018] and the canonical switching rate of the Zig-Zag in Bierkens et al. [2019].

In order to satisfy (5.21), the appropriate choice of density μ^* with respect to the measure m on E is

$$\mu^*(x) := (C^*)^{-1}[0 \vee (\underline{\kappa} - \tilde{\kappa}(x))]\pi(x), \tag{5.22}$$

where

$$C^* := \int_E [0 \vee (\underline{\kappa} - \tilde{\kappa}(x))]\pi(x)\,\mathrm{d}m(x),$$

assuming that this quantity is finite. While this normalising constant will generally be intractable, the algorithm does not require its precise numerical value: we will only need to evaluate the right-hand side of (5.21) point-wise, and be able to sample from μ^*.

Proposition 5.5.1 (Minimal rebirth distribution)**.** *Let μ^*, C^* be defined as above for some fixed $\underline{\kappa} \geq 0$, where we assume μ^* is integrable and normalised. Let μ, C be any (normalised) probability measure on E and positive constant respectively such that* (5.20) *holds. Then μ^* minorises μ, in the sense that there exists some $\epsilon > 0$ such that for all measurable $B \subset E$,*

$$\mu(B) \geq \epsilon \mu^*(B), \tag{5.23}$$

and we have that

$$C \geq C^*.$$

Proof. From the assumption that (5.20) holds, we must have that $\kappa \geq \kappa^*$ pointwise, from which it follows that for each $x \in E$,

$$C\mu(x) \geq C^*\mu^*(x),$$

which establishes (5.23), and by integrating both sides over E it follows that $C \geq C^*$. □

How to obtain samples from μ^* is in general not obvious, and is reminiscent of sampling from minorising measures as in Murdoch and Green [1998]. μ^* is generally compactly supported and supported around the modes of π; its support is contained within the set

$$\{x \in E : \tilde{\kappa}(x) < \underline{\kappa}\},$$

and so often simulation is possible through straightforward rejection sampling.

Examples of the minimal rebirth distribution will be given for two isotropic densities on $\mathbb{R}^d$ in Sections 5.6.1.1 and 5.6.2.1, along with their limit as the dimension d increases.

5.5.2 Tuning the parameter C

Another practical consideration is the tuning of the parameter C. In order for Restore to have the correct invariant distribution, the constant C in (5.2) needs to be chosen so that the killing rate κ is nonnegative everywhere. However beyond this restriction, any value of C is permissible.

The constant C also has a useful practical interpretation: as is shown in proof of Theorem 5.3.7, $C = \mathbb{E}_\mu[\tau_\partial]$, the average lifetime when started from μ.

How one should choose C in practice to optimise convergence is an interesting question. In terms of computational cost, one would like to minimise the killing rate, since simulating the event times is one of the main computational costs. In this sense, choosing C as small as possible – for instance the minimal C^* along with the minimal rebirth distribution μ^* of Section 5.5.1 – would be one choice, when sampling from μ^* is feasible.

On the other hand, if one is implementing the CFTP version of Restore from Section 5.4.1, it is the lower bound $\underline{\kappa}$ which dictates the rate at which trajectories couple: we would prefer $\underline{\kappa}$ to be as *large* as possible to accelerate the coupling of trajectories. Making $\underline{\kappa}$ larger would require choosing a larger value of C. The interaction between these two objectives – minimising C to reduce the killing rate, while increasing C to maximise the lower bound – is nontrivial.

When the interarrival process is already π-invariant, tuning C is even more open, since *any* nonnegative value of C can be used. In this case, since we want to make use of both the local and global dynamics, a sensible way to tune C would be to choose it such that the average rate of teleports matches the rate of mixing of the interarrival process. In particular, looking at the average rate of teleports (the number of teleports divided by the total stochastic process time) is practically feasible, since it does not

assume that μ or π in (5.2) are normalised, and is immediately available from pilot runs. This approach is inspired by the recent Caputo and Quattropani [2019], who prove for the PageRank surfer on random (finite) graphs, the resulting mixing time depends on the interplay between the rate of mixing of the underlying walk and the teleport probability.

The interesting problem of formulating an optimal choice of C remains an open question.

5.5.3 Truncated killing

Another practical issue for implementing Restore is the issue of simulating the lifetimes $\tau^{(i)}$ as in (5.1). In case of countable states spaces, Section 5.3.1, or jump processes, 5.3.3, this is straightforward, since $t \mapsto \kappa(X_t)$ is piecewise constant. When κ is a uniformly bounded function simulation is also straightforward, since we can make use of Poisson thinning ([Devroye, 1986, Chapter 6.2]); see Algorithm 2.

We consider now the diffusion case, as in Section 5.3.2. In this case κ is typically unbounded. In some cases using layered processes it is still possible to simulate τ_∂ exactly, as with the techniques of Pollock et al. [2016]. These are technically demanding, so in this section we consider the alternative of *truncating* the killing rate. Namely, we fix some upper bound M, and work with the truncated killing rate

$$\kappa_M := \kappa \wedge M.$$

This will introduce some bias, a discrepancy between the invariant distribution and π, but we will show how this bias may be explicitly quantified.

In order to prove our result we will need to assume the following.

We assume that the interarrival process Y is a diffusion on $\mathbb{R}^d$ satisfying BASSA, and that κ is continuous. We also assume that Assumption 15 holds, namely that we have a lower bound

$$\kappa \geq \underline{\kappa} > 0.$$

Recall that under Assumption 15, τ_∂ can be stochastically dominated by an exponential random variable with rate $\underline{\kappa}$, and hence all moments of τ_∂ are finite.

In order to avoid pathologies we assume that

$$M > \inf_{x \in E} \kappa(x). \tag{5.24}$$

We consider now the Restore process X with interarrival process Y, rebirth distribution μ and truncated killing rate κ_M, for some given truncation level M satisfying (5.24).

Throughout this section we will only be interested in the behaviour of the Restore process *before the first rebirth event*. As the rebirth distribution μ will not play a significant role we will consider the unkilled process Y, and explicitly augment it with a first rebirth time. We will simply write $\mathbb{E}_x$ for the law of the unkilled process Y started from x, and consider the first arrival time τ_∂ to be a random variable defined by (5.1).

Let us write κ_M^{e} for the excess killing over level M, that is,

$$\kappa_M^{\mathrm{e}} := \kappa - \kappa_M.$$

τ_∂ is the first rebirth event of the Restore process with killing rate κ, defined by (5.1). Then by Poisson superposition, we can write

$$\tau_\partial = \tau_M \wedge \tau_M^{\mathrm{e}}, \tag{5.25}$$

where $\tau, \tau_M, \tau_M^{\mathrm{e}}$ are the first arrival times of inhomogeneous Poisson process with rate functions $t \mapsto \kappa(Y_t)$, $t \mapsto \kappa_M(Y_t)$ and $t \mapsto \kappa_M^{\mathrm{e}}(Y_t)$ respectively, where these Poisson processes are independent conditional on the path $t \mapsto Y_t$.

In particular, τ_M and τ_M^{e} can be written as

$$\tau_M = \inf\left\{ t \geq 0 : \int_0^t \kappa_M(Y_s)\,\mathrm{d}s \geq \xi_1 \right\}, \tag{5.26}$$

$$\tau_M^{\mathrm{e}} = \inf\left\{ t \geq 0 : \int_0^t \kappa_M^{\mathrm{e}}(Y_s)\,\mathrm{d}s \geq \xi_2 \right\}, \tag{5.27}$$

where $\xi_1, \xi_2 \sim \mathrm{Exp}(1)$ are independent of each other and of the underlying process Y.

Since we are assuming Assumption 15 holds, by the arguments of Section 5.4.1 it follows that the Restore process with killing rate κ has a unique invariant distribution π, and from Section 5.4.2, π can be written as

$$\pi(f) = \frac{\mathbb{E}_\mu\left[\int_0^{\tau_\partial} f(Y_s)\,\mathrm{d}s\right]}{\mathbb{E}_\mu[\tau_\partial]},$$

where here Y is the unkilled process and τ_∂ is defined as in (5.1). We can rewrite this by exchanging the order of integration. Consider the resolvent operator, which maps

measurable functions to measurable functions,

$$\begin{aligned}\mathcal{R}f(x) &:= \int_0^\infty \mathrm{d}t\, \mathbb{E}_x[f(Y_t)1\{\tau_\partial > t\}] \\ &= \int_0^\infty \mathrm{d}t\, \mathbb{E}_x\left[f(Y_t)1\{\tau_M > t\}1\{\tau_M^{\mathrm{e}} > t\}\right],\end{aligned}$$

where the second equality holds by (5.25). Note that given a bounded measurable function f, $\mathcal{R}f$ is also a bounded measurable function, since we can bound

$$|\mathcal{R}f(x)| \leq \|f\|_\infty \mathbb{E}_x[\tau_\partial] \leq \|f\|_\infty \underline{\kappa}^{-1}.$$

Thus by Fubini's theorem, we can write

$$\pi = \frac{\mu\mathcal{R}}{\mu\mathcal{R}1},$$

in analogue with expressions given in Wang et al. [2019b] and Benaïm et al. [2018].

Similarly, the Restore process with truncated killing rate κ_M is still uniformly ergodic and possesses a unique invariant distribution π_M, which can be represented in a similar way. Write

$$\pi_M = \frac{\mu\mathcal{R}_M}{\mu\mathcal{R}_M 1},$$

where for bounded measurable f,

$$\mathcal{R}_M f(x) = \int_0^\infty \mathrm{d}t\, \mathbb{E}_x[f(Y_t)1\{\tau_M > t\}].$$

Our goal now is to bound the total variation distance

$$\|\pi_M - \pi\|_1,$$

as a function of M.

Theorem 5.5.2. *We have the following bound on the bias.*

$$\|\pi_M - \pi\|_1 \leq \frac{4\int_0^\infty \mathbb{P}_\mu(\tau_M^{\mathrm{e}} \leq t)\exp(-t\underline{\kappa})\,\mathrm{d}t}{\mathbb{E}_\mu[\tau_\partial]}.$$

Proof. We have that

$$\begin{aligned}\pi_M - \pi &= \frac{(\mu\mathcal{R}1)\mu\mathcal{R}_M - (\mu\mathcal{R}_M 1)\mu\mathcal{R}}{(\mu\mathcal{R}_M 1)(\mu\mathcal{R}1)} \\ &= \frac{\mu\mathcal{R}1(\mu\mathcal{R}_M - \mu\mathcal{R}) + (\mu\mathcal{R}1 - \mu\mathcal{R}_M 1)\mu\mathcal{R}}{(\mu\mathcal{R}_M 1)(\mu\mathcal{R}1)}.\end{aligned}$$

So we would like to bound

$$|\mu\mathcal{R}f - \mu\mathcal{R}_M f|$$

for arbitrary bounded measurable f.

For any nonnegative bounded measurable f,

$$\begin{aligned}
|\mu\mathcal{R}f - \mu\mathcal{R}_M f| &\le \int \mu(\mathrm{d}x) \int \mathrm{d}t \, |\mathbb{E}_x[f(Y_t)1\{\tau_M > t\}1\{\tau_M^{\mathrm{e}} > t\}] \\
&\qquad - \mathbb{E}_x[f(X_t)1\{\tau_M > t\}]| \\
&\le \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{E}_x\left[f(Y_t)\left(1 - 1\{\tau_M^{\mathrm{e}} > t\}\right)1\{\tau_M > t\}\right] \\
&\le \|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{E}_x[1\{\tau_M^{\mathrm{e}} \le t, \tau_M > t\}] \\
&= \|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t, \tau_M > t).
\end{aligned}$$

Since we are assuming that we have a lower bound $\underline{\kappa}$ on the killing rate, and Assumption 5.24 holds, we can stochastically bound $\tau_M \le \tau'$ where $\tau' \sim \mathrm{Exp}(\underline{\kappa})$ and τ' is independent of everything else. So continuing the chain of inequalities,

$$\begin{aligned}
&\le \|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t, \tau' > t) \\
&= \|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, \mathbb{P}_x(\tau' > t) \\
&= \|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}}.
\end{aligned}$$

A universal upper bound on this quantity is $\|f\|_\infty/\underline{\kappa}$.

For for a given continuous nonnegative bounded f with $\|f\|_\infty \le 1$ we get the following bounds.

$$\begin{aligned}
&|\pi_M f - \pi f| \\
&\le \frac{\mu\mathcal{R}1\|f\|_\infty \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}} + \int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}} |\mu\mathcal{R}f|}{(\mu\mathcal{R}_M 1)(\mu\mathcal{R}1)} \\
&\le \frac{\int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}}}{\mu\mathcal{R}_M 1} + \frac{\int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}}}{\mu\mathcal{R}_M 1} \\
&= \frac{2\int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}}}{\mu\mathcal{R}_M 1} \\
&\le \frac{2\int \mu(\mathrm{d}x) \int \mathrm{d}t \, \mathbb{P}_x(\tau_M^{\mathrm{e}} \le t) \, e^{-t\underline{\kappa}}}{\mu\mathcal{R}1}
\end{aligned}$$

Since this bound is valid for only nonnegative bounded f, in order to bound $\|\pi_M - \pi\|_1$ we pick up an additional factor of 2.

□

Remark 5.5.3. To use this bound we need to further bound

$$\mathbb{P}_x(\tau_M^{\mathrm{e}} \leq t).$$

Intuitively, if κ_M is a reasonable approximation for κ, then κ_M^{e} is low, and hence τ_M^{e} tends to be large, and so this bound is tighter.

Proposition 5.5.4. *Fix a rebirth distribution μ. We have that*

$$\int_0^\infty \mathrm{d}t\, \mathbb{P}_\mu(\tau_M^{\mathrm{e}} \leq t)\, e^{-t\underline{\kappa}} \to 0 \text{ as } M \to \infty. \tag{5.28}$$

Thus by Theorem 5.5.2 as $M \to \infty$,

$$\|\pi_M - \pi\|_1 \to 0.$$

Proof. The event $\left\{\tau_M^{\mathrm{e}} \leq t\right\}$ is contained in the event $\left\{\sup_{s\leq t} \kappa(Y_s) \geq M\right\}$. Thus, for any fixed x

$$\begin{aligned}
\lim_{M\to\infty} \mathbb{P}_x(\tau_M^{\mathrm{e}} \leq t) &= \mathbb{P}_x\left(\bigcap_{M=1}^{\infty} \{\tau_M^{\mathrm{e}} \leq t\}\right) \\
&\leq \mathbb{P}_x\left(\sup_{s\leq t} \kappa(Y_s) = \infty\right) \\
&\leq \mathbb{P}_x\left(\sup_{s\leq t} \|Y_s\| = \infty\right) \text{ since } \kappa \text{ is locally bounded} \\
&= 0 \text{ since } Y \text{ is nonexplosive.}
\end{aligned}$$

By the Dominated Convergence Theorem it follows that

$$\lim_{M\to\infty} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t\underline{\kappa}} \int_E \mathrm{d}\mu(x) \mathbb{P}_x\left(\tau_M^{\mathrm{e}} \leq t\right) = 0,$$

which is precisely (5.28). □

In order for Theorem 5.5.2 to be of practical use, we will further need bounds on

$$\mathbb{P}_x(\tau_M^{\mathrm{e}} \leq t), \tag{5.29}$$

which will vary given the particular situation; given the choice of the underlying diffusion Y, target π and rebirth distribution μ.

The rate at which the probabilities (5.29) decay as a function of M will crucially depend on the rate at which the killing rate κ grows. Thus we define the following,

$$L(M) := \sup\{\ell > 0 : \forall x \in [-\ell, \ell]^d \subset \mathbb{R}^d, \kappa(x) \leq M\},$$

which for a given truncation level M defines the largest hypercube on which no truncation occurs.

The rate at which $L(M)$ grows as $M \to \infty$ will crucially dictate the rate at which the error decays. Then let

$$H(M) := [-L(M), L(M)]^d \subset \mathbb{R}^d,$$

and let

$$T_M := \inf\{t \geq 0 : Y_t \in \mathbb{R}^d \backslash H(M)\}$$

be the first hitting time of the unkilled diffusion Y of the complement of $H(M)$. Clearly we must have

$$T_M \leq \tau_M^{\mathrm{e}}.$$

Thus it follows that

$$\int_0^\infty \mathbb{P}_x(\tau_M^{\mathrm{e}} \leq t)\, e^{-\kappa t}\, \mathrm{d}t \leq \int_0^\infty \mathbb{P}_x(T_M \leq t)\, e^{-\kappa t}\, \mathrm{d}t.$$

To proceed from here we require knowledge of the distribution of the hitting times T_M for the underlying diffusion Y. At this point we will *specialise to the case of Brownian motion*, however a similar analysis can be performed in any situation where we have analogous bounds on the hitting times.

By the reflection principle of one-dimensional Brownian motion, we know that for any $a > 0$,

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} |B_s| > a\right) \leq 2\mathbb{P}(M_t > a) = 4\mathbb{P}(B_t > a) = 4\left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right).$$

Here Φ denotes the standard univariate normal cumulative distribution function. For a multidimensional standard Brownian motion, it follows that

$$\mathbb{P}_0\left(T_M \leq t\right) \leq 4d\left(1 - \Phi\left(\frac{L(M)}{\sqrt{t}}\right)\right).$$

This is because leaving a hypercube is the same as having some component leaving the interval $[-L(M), L(M)]$.

We now make use of the well-known bound for the normal cumulative distribution function: for each $\lambda > 0$,

$$1 - \Phi(\lambda) < \frac{1}{\sqrt{2\pi}\lambda} e^{-\lambda^2/2}.$$

This leads to the bound

$$\int_0^\infty \mathbb{P}_0(T_M \leq t)\, e^{-\underline{\kappa} t}\, \mathrm{d}t \leq \frac{4d}{\sqrt{2\pi}} \int_0^\infty \frac{\sqrt{t}}{L(M)} e^{-L(M)^2/(2t)} e^{-\underline{\kappa} t}\, \mathrm{d}t.$$

This integral can be evaluated analytically[1], to obtain

$$\begin{aligned} &= \frac{4d}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2\underline{\kappa}} \left(\sqrt{2} + \frac{1}{L(M)\sqrt{\underline{\kappa}}} \right) e^{-\sqrt{2\underline{\kappa}} L(M)} \\ &= \frac{2d}{\underline{\kappa}} \left(1 + \frac{1}{L(M)\sqrt{2\underline{\kappa}}} \right) e^{-\sqrt{2\underline{\kappa}} L(M)}. \end{aligned}$$

So for large values of M we have a bound that decays like

$$e^{-\sqrt{2\underline{\kappa}} L(M)}.$$

This can be used to give practical suggestions of how large to choose M in order to balance the bias and variance of the algorithm's output.

Suppose we are able to obtain n i.i.d. draws $X_1, \ldots, X_n \sim \pi^M$, say by running the CFTP algorithm a total of n times.

So for a bounded test function f, we estimate $\pi(f)$ by

$$\sum_{i=1}^n \frac{f(X_i)}{n}.$$

We roughly estimate the error as

$$\begin{aligned} \left| \sum_{i=1}^n \frac{f(X_i)}{n} - \pi(f) \right| &\leq \underbrace{\left| \sum_{i=1}^n \frac{f(X_i)}{n} - \pi^M(f) \right|}_{\approx \frac{1}{\sqrt{n}}} + \underbrace{\left| \pi^M(f) - \pi(f) \right|}_{\leq \|f\|_\infty \|\pi^M - \pi\|_{\mathrm{TV}}} \\ &\approx O\left(\frac{1}{\sqrt{n}} \right) + \exp\left(-\sqrt{2\underline{\kappa}} (L(M))^2 \right). \end{aligned}$$

In order to balance these two terms, it is advisable to choose n and M such that

$$\begin{aligned} \frac{1}{\sqrt{n}} &\sim \exp\left(-\sqrt{2\underline{\kappa}} L(M) \right) \\ \Rightarrow \frac{\log n}{2\sqrt{2\underline{\kappa}}} &\sim L(M). \end{aligned}$$

So this gives some indication of how to choose M, given n. This will achieve an error of order roughly $O(n^{-1/2})$. The cost in n will be roughly $O(n \log n)$.

[1] `https://www.wolframalpha.com/input/?i=int_0%5Einfty+%5Csqrt+(t)+exp(-a%2F(2t))+exp(-t)dt`

5.6 Examples

We now give some examples to illustrate the Restore sampler. Throughout, we write $|\cdot|$ for the Euclidean norm on $\mathbb{R}^d$ and Tr for the matrix trace.

5.6.1 Multivariate Gaussian

We describe the Restore sampler for targeting a multivariate Gaussian on $\mathbb{R}^d$. Namely, when

$$\pi(x) \propto \exp\left(-\frac{1}{2}(x-\nu)^\top \Sigma^{-1}(x-\nu)\right), \quad x \in \mathbb{R}^d.$$

Here $\nu \in \mathbb{R}^d$ and Σ is a positive definite symmetric $d \times d$ covariance matrix.

We take the underlying diffusion to simply be a standard Brownian motion; $A \equiv 0$ in (5.5). This will certainly satisfy the BASSA assumptions, Assumption 10.

It is easily calculated that

$$\tilde{\kappa}(x) = \frac{1}{2}\left((x-\nu)^\top \Sigma^{-2}(x-\nu) - \mathrm{Tr}\left(\Sigma^{-1}\right)\right), \quad x \in \mathbb{R}^d.$$

This is nonpositive on the ellipsoid

$$H = \left\{x \in \mathbb{R}^d : |\Sigma^{-1}(x-\nu)|^2 \leq \mathrm{Tr}\left(\Sigma^{-1}\right)\right\}, \tag{5.30}$$

and grows quadratically to infinity as $|x|$ grows. Since π is bounded above, provided our chosen rebirth density μ is continuous and positive on H, we will be able to find a constant C such that Assumption 12 holds.

For the technical Assumption 13, we can make use of Lemma 5.3.6 to verify (5.8). Since κ is quadratic and we are dealing with a standard Brownian motion, a sufficient condition on μ for (5.9) is that it has at least, say, exponentially decaying tails.

5.6.1.1 Minimal rebirth distribution

One natural choice of rebirth distribution is the *minimal* rebirth distribution μ^* of Section 5.5.1 and corresponding choice of constant C^*. We now derive the limiting shape of the minimal rebirth distribution as the dimension $d \to \infty$. We take

$$\underline{\kappa} = 0.$$

In order to obtain an explicit representation, we will consider the isotropic case, and consider the radial component. Namely, consider a d-dimensional standard Gaussian target.

$$\pi_d(x) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2}, \quad x \in \mathbb{R}^d.$$

As above, we will take the underlying process Y to be a standard Brownian motion on $\mathbb{R}^d$, giving the following partial killing rate,

$$\tilde{\kappa}_d(x) = \frac{1}{2}(|x|^2 - d), \quad x \in \mathbb{R}^d.$$

Recall that the surface area element in d-dimensions is given by

$$S_{d-1}(r) = \frac{d\pi^{d/2}}{\Gamma(d/2+1)} r^{d-1}.$$

The minimal rebirth distribution μ^* is supported on the ellipsoid H as in (5.30), which in this case is simply

$$H_d = \{x \in \mathbb{R}^d : |x|^2 \leq d\}.$$

Let us write $r = |x|$. Consider the following function, the product of π_d, $-\tilde{\kappa}$ and S_{d-1},

$$\mu^*_{u,d}(r) = \frac{1}{2}(d - r^2)\frac{de^{-r^2/2}}{2^{d/2}\Gamma(d/2+1)} r^{d-1}, \quad r \geq 0. \tag{5.31}$$

Written as a function of the radial component, the minimal rebirth density μ^*_d will be proportional on H to $\mu^*_{u,d}$. (Here the subscript u signifies 'unnormalised'.)

Direct differentiation shows that the maximum of $\mu^*_{u,d}$ occurs when

$$r^2 = d + \frac{1}{2} - \sqrt{2d + \frac{1}{4}}.$$

We will use this to set the scaling,

$$r(s) = s + \sqrt{d + \frac{1}{2} - \sqrt{2d + \frac{1}{4}}}. \tag{5.32}$$

This $\sqrt{d}$-shift reflects how the bulk of the probability mass of standard Gaussians 'moves out at rate $\sqrt{d}$'.

Since the radial component must be nonnegative, (5.32) is valid when

$$s \geq -\sqrt{d + \frac{1}{2} - \sqrt{2d + \frac{1}{4}}}.$$

But since we will be considering the limit as $d \to \infty$, (5.32) will eventually be well-defined for all $s \in \mathbb{R}$.

We can now state the limiting shape of the minimal rebirth distribution.

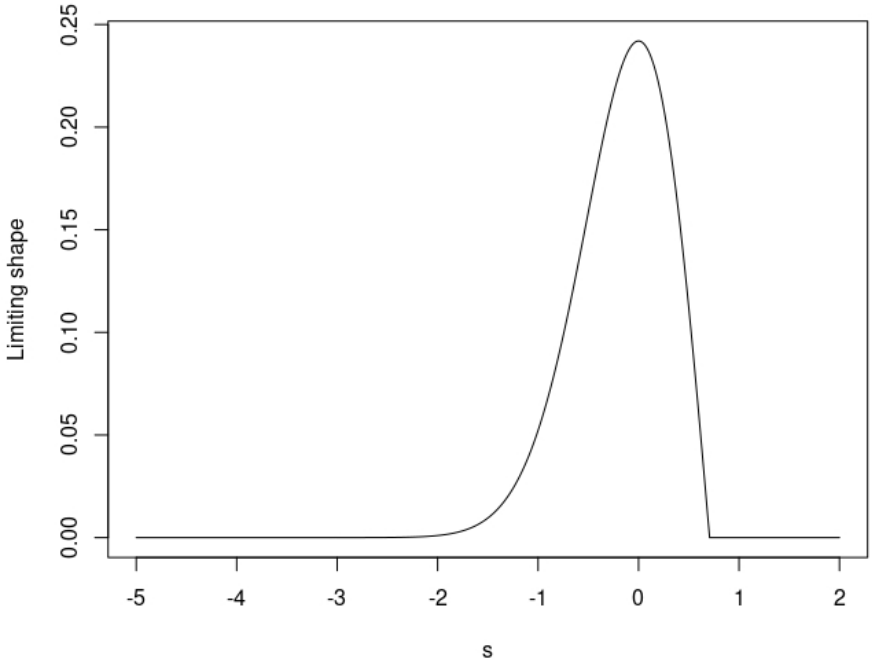


Figure 5.1: A plot of the function $\mu^*_{u,\infty}$ from Theorem 5.6.1, the limiting shape of the radial component of the minimal rebirth distribution for a standard Gaussian target.

Theorem 5.6.1. *For a standard d-dimensional standard Gaussian target, with underlying process Brownian motion and $\underline{\kappa} = 0$, we have the following limit for the radial component of the unnormalised minimal rebirth distribution.*

As $d \to \infty$, for each $s \in \mathbb{R}$,

$$d^{-1/2}\mu^*_{u,d}(r(s)) \to \frac{1}{\sqrt{\pi \mathrm{e}}}\left(\frac{\sqrt{2}}{2} - s\right)\exp(-s^2)\exp(\sqrt{2}s) \vee 0 =: \mu^*_{u,\infty}(s).$$

Proof. Substitute (5.32) into (5.31) and carefully manipulate the resulting expression, using Stirling's formula to expand the Gamma function. □

This limiting shape is plotted in Figure 5.1.

Remark 5.6.2. We can use this to estimate the cost of this Restore sampler. Using the constant C in the killing rate as an estimate of the cost of introducing rebirths, we can calculate here that

$$C^*_d = \int (-\kappa_d(x))\pi_d(x)\,\mathrm{d}x = \int \mu^*_{u,d}(r)\,\mathrm{d}r \approx d^{1/2}\int \mu^*_{u,\infty}(s)\,\mathrm{d}s.$$

So the cost of introducing the rebirths scales roughly like $d^{1/2}$.

5.6.2 Cauchy posterior

We now give a univariate example where π has heavy tails and is multimodal. This simple example is based on Example 3.1 of Murdoch [2000]. We take

$$\bar{\pi}(x) \propto \prod_{i=1}^{n} \frac{1}{1 + (y_i - x)^2},$$

for some observations $(y_1, \ldots, y_n) \in \mathbb{R}^n$, with respect to Lebesgue measure on $\mathbb{R}$. (We use the notation $\bar{\pi}$, since in the notation of Section 5.3.2, the symbol π is reserved for the target density with respect to the measure Γ.)

This can be thought of as the posterior distribution for i.i.d. Cauchy(x) data, with an improper uniform prior on $\mathbb{R}$ for x. In Example 3.1 of Murdoch [2000], the author considered this density with a $N(0, 100^2)$ prior on x. We choose to use a uniform prior since in this case the posterior π has polynomially decaying tails.

We will take the same data as Murdoch [2000], namely $n = 3$ and observations $(1.3, -11.6, 4.4)$. The resulting posterior is plotted in Figure 5.2.

In Murdoch [2000], a method was given to couple MCMC chains started from different initial locations, by using a mixture of random walk Metropolis moves interspersed with occasional independence sampler moves. Our approach with Restore is very similar. However, there are crucial differences. Firstly, our target $\bar{\pi}$ has heavy, polynomially-decaying tails, whereas Murdoch [2000] takes a Gaussian prior, giving exponential decay in the tails. Secondly, for the rebirth distribution we take the minimal rebirth distribution of Section 5.5.1, which is compactly supported. Thus we do not need to tune the proportion of local versus global moves; this is done automatically by the event rate. And most significantly, neither our local nor global dynamics are themselves $\bar{\pi}$-invariant. By choosing local dynamics which rapidly explore the tails we will obtain a straightforward recipe, using CFTP, to obtain exact draws from the target.

We would like to use the diffusion form of Restore described in Section 5.3.2. Since our target has heavy tails, we will need a diffusion which is rapidly able to enter the tails.

As such, for our underlying process, we will take the following diffusion: an *unstable* Ornstein–Uhlenbeck process, described by the SDE

$$\mathrm{d}Y_t = Y_t \, \mathrm{d}t + \mathrm{d}B_t, \tag{5.33}$$

where B is a standard univariate Brownian motion. We showed in Section 5.3.2.1 that this diffusion satisfies the BASSA conditions (Assumption 10). This diffusion,

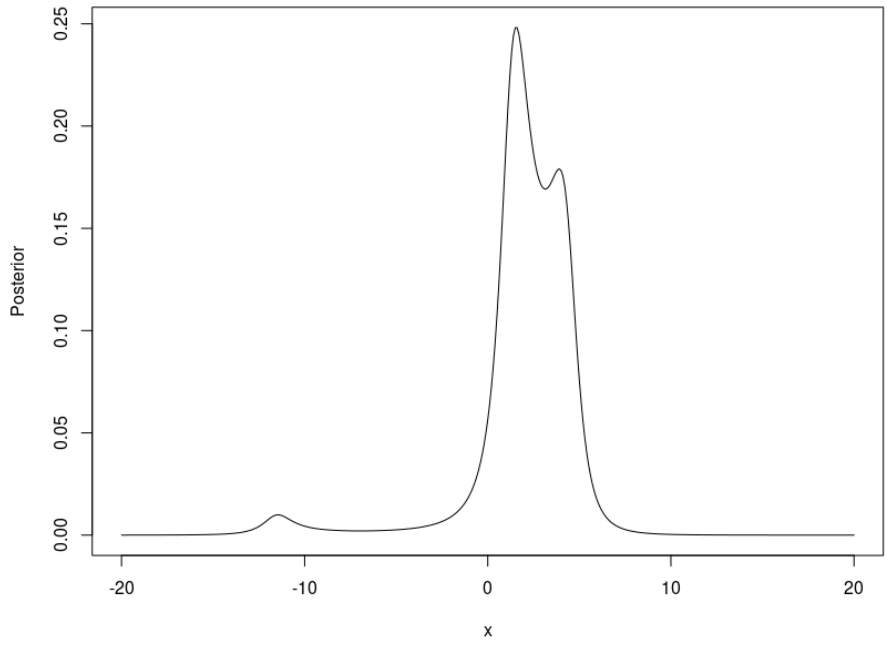


Figure 5.2: The heavy-tailed multi-modal target distribution π of (5.6.2).

like a stable Ornstein–Uhlenbeck process, is also a Gaussian process with known finite-dimensional distributions, and so can be simulated easily without error.

This diffusion accelerates into the tails, and so produces via CFTP an exact draw from heavy-tailed invariant distribution π in finite time.

Through routine, but slightly cumbersome, calculations, we arrive at the partial killing rate,

$$\begin{aligned}\tilde{\kappa}(x) = &\sum_{i=1}^{n} \frac{-1 + (y_i - x)^2}{(1 + (y_i - x)^2)^2} + 2\left[\sum_{i=1}^{n} \frac{(y_i - x)}{1 + (y_i - x)^2}\right]^2 \\ &+ \sum_{i=1}^{n} \frac{2(y_i - x)^2}{1 + (y_i - x)^2} - \sum_{i=1}^{n} \frac{2y_i(y_i - x)}{1 + (y_i - x)^2} - 1.\end{aligned} \tag{5.34}$$

This is plotted in Figure 5.3. As expected, $\tilde{\kappa}$ is negative precisely in the vicinity of the local maxima of π. We can see that in this case the partial killing rate is bounded above; a global upper bound is given by

$$M = 16.$$

Analytically, we know that the limiting value of $\tilde{\kappa}(x)$ as $|x| \to \infty$ is 5.

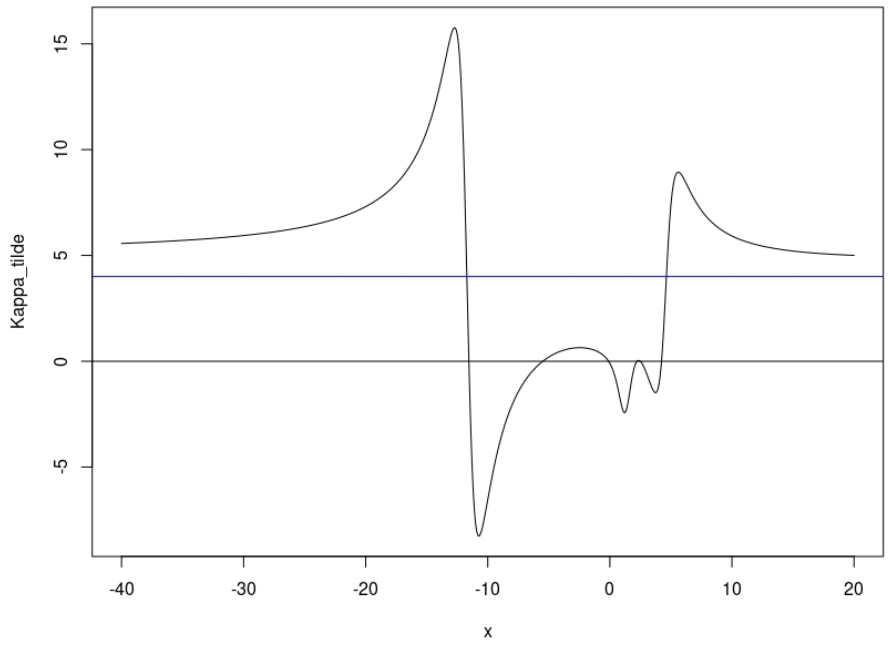


Figure 5.3: $\tilde{\kappa}$ for the Cauchy example as in (5.34). The blue line is the value $\underline{\kappa} = 4$.

For our subsequent CFTP implementation we will take the lower bound

$$\underline{\kappa} = 4.$$

For the rebirth distribution we will take the minimal rebirth distribution μ^* from Section 5.5.1, with $\underline{\kappa} = 4$. This is plotted in Figure 5.4. We take the corresponding value of $C = C^*$ as in (5.21), so that the resulting killing rate is

$$\kappa(x) = \tilde{\kappa}(x) \vee 4,$$

where $\tilde{\kappa}$ is defined in (5.34).

Practically, we need to draw samples from μ^*. For this simple example, this can be done through straightforward rejection sampling from a uniform distribution on the interval $[-11.8, 5]$, since we know this interval contains the support of μ^*.

Assumptions 11 and 12 are easily seen to be satisfied. Assumption 13 is relatively easy to check. We can use the sufficient integral conditions of Lemma 5.3.6 for (5.8), and since κ is bounded (5.9) is immediate.

By construction, we have that the killing rate κ satisfies

$$4 \leq \kappa(x) \leq 16, \quad x \in \mathbb{R},$$

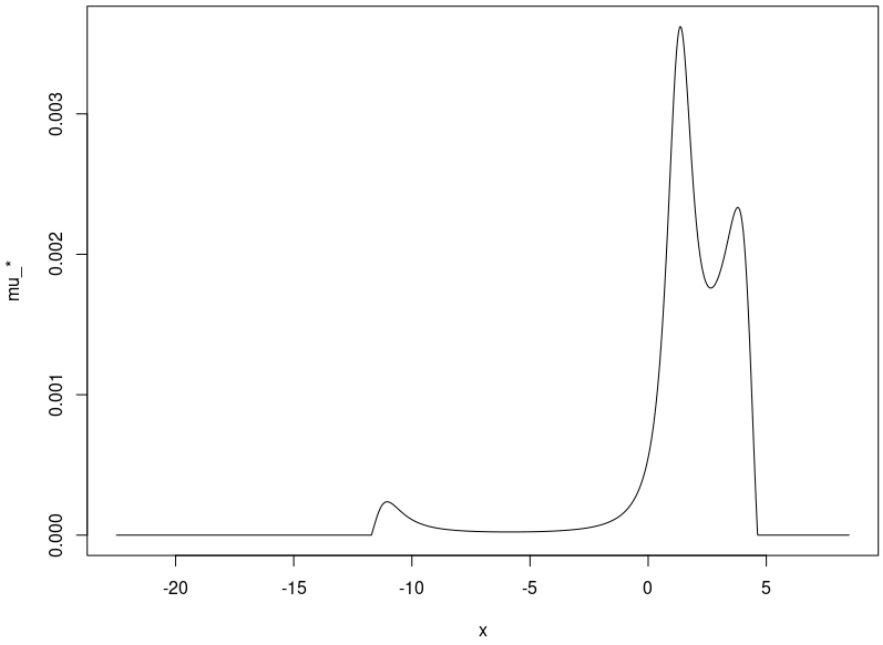


Figure 5.4: The minimal rebirth distribution μ^* for the Cauchy example with lower bound $\underline{\kappa} = 4$. Its precise definition is given in (5.22). In this case draws from μ^* can be obtained easily by rejection sampling from a uniform distribution.

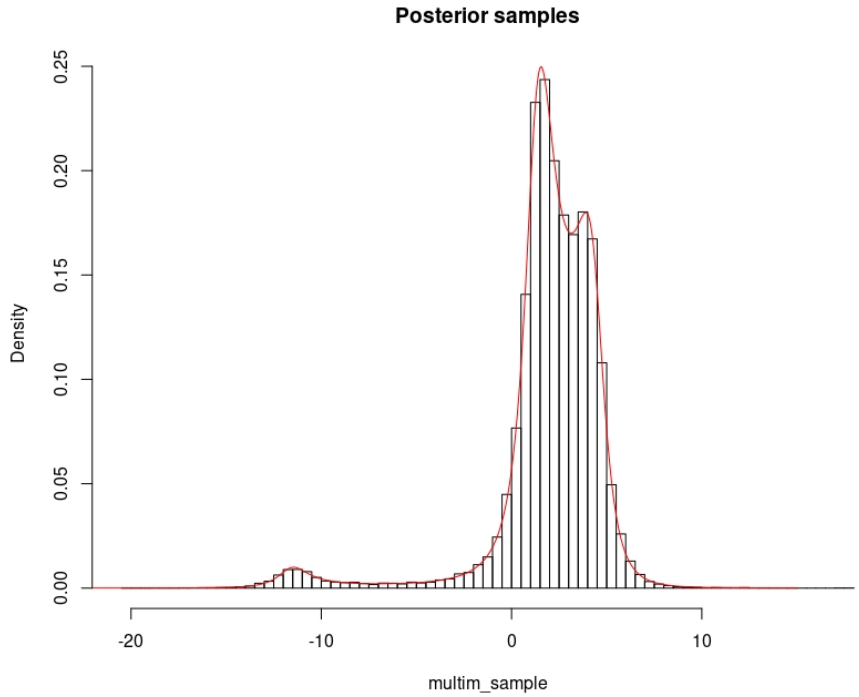


Figure 5.5: 30,000 samples obtained from the CFTP implementation. This an i.i.d. sample from π. The true posterior π is plotted in red.

so Assumption 15 holds.

We now have all the ingredients to apply the CFTP implementation to obtain exact i.i.d. draws from π. Implementation is straightforward since the killing rate is bounded above so we can make use of Poisson thinning; see Algorithm 4.

A histogram consisting of 30,000 draws from the CFTP implementation are plotted in Figure 5.5. This took a matter of seconds to obtain. These were obtained by running the CFTP algorithm 30,000 times (independently), and so the resulting sample is i.i.d.

5.6.2.1 Minimal rebirth distribution

We can perform an analysis for the minimal rebirth distribution of a heavy-tailed multivariate Cauchy distribution, as was done for the multivariate Gaussians in Section 5.6.1.1.

Again we will want a radially symmetric density, so we take

$$\pi_d(x) = \frac{\Gamma((1+d)/2)}{\Gamma(1/2)\pi^{d/2}(1+|x|^2)^{(1+d)/2}}, \quad x \in \mathbb{R}^d.$$

This is a multivariate Cauchy density.

Using an unstable Ornstein–Uhlenbeck process as in (5.33), except this time on $\mathbb{R}^d$, we find that the partial killing is given by

$$\tilde{\kappa}_d(x) = \frac{1+d}{2(1+|x|^2)^2} \cdot (3|x|^2 - d) + \frac{|x|^2 - d}{1+|x|^2}, \quad x \in \mathbb{R}^d.$$

We will take again $\underset{\sim}{\kappa} = 0$. Then as in Section 5.6.1.1, the radial component unnormalised minimal rebirth density is given by

$$\mu^*_{u,d}(r) = (-\tilde{\kappa}_d(r))\pi_d(r)S_{d-1}(r) \vee 0, \quad r \geq 0.$$

In this setting, the relevant scaling is $r\sqrt{d/3}$. We obtain the following.

Theorem 5.6.3. *For the symmetric multivariate Cauchy target, with an unstable OU process* (5.33) *as the underlying process and $\underset{\sim}{\kappa} = 0$, we have the following limit for the minimal rebirth distribution.*

As $d \to \infty$, for each $r \geq 0$,

$$\sqrt{d}\mu^*_{u,d}\left(r\sqrt{\frac{d}{3}}\right) \to -\frac{9\sqrt{2}}{\sqrt{\pi}}\frac{1}{r^6}\exp\left(-\frac{3}{2r^2}\right)\left(\frac{r^4}{3} + \frac{r^2}{2} - \frac{3}{2}\right) \vee 0 =: \mu^*_{u,\infty}(r).$$

Proof. This is just a matter of carefully working through the algebra, and making use of Stirling's formula. □

The limiting shape is plotted in Figure 5.6. By solving the polynomial, we can calculate that the root of $\mu^*_{u,\infty}$ is at $r = \sqrt{6}/2$.

Remark 5.6.4. We can now estimate the value of C^*_d to get a sense of the cost of adding rebirths.

$$\begin{aligned} C^*_d &= \int (-\kappa_d(x))\pi_d(x)\,\mathrm{d}x = \int \mu^*_{u,d}(r)\,\mathrm{d}r = \int \mu^*_{u,d}\left(s\sqrt{\frac{d}{3}}\right)\sqrt{\frac{d}{3}}\,\mathrm{d}s \\ &\approx \frac{1}{\sqrt{3}}\int \mu^*_{u,\infty}(s)\,\mathrm{d}s. \end{aligned}$$

So in this case the cost of adding rebirths is roughly constant as the dimension d increases.

5.6.3 Jump process: correcting a poor approximation

We give an example where Restore is able to correct a poor approximation of the target π. This example is meant to emulate situations where the practitioner has some sense of approximately where the modes of the target lies, but does not possess

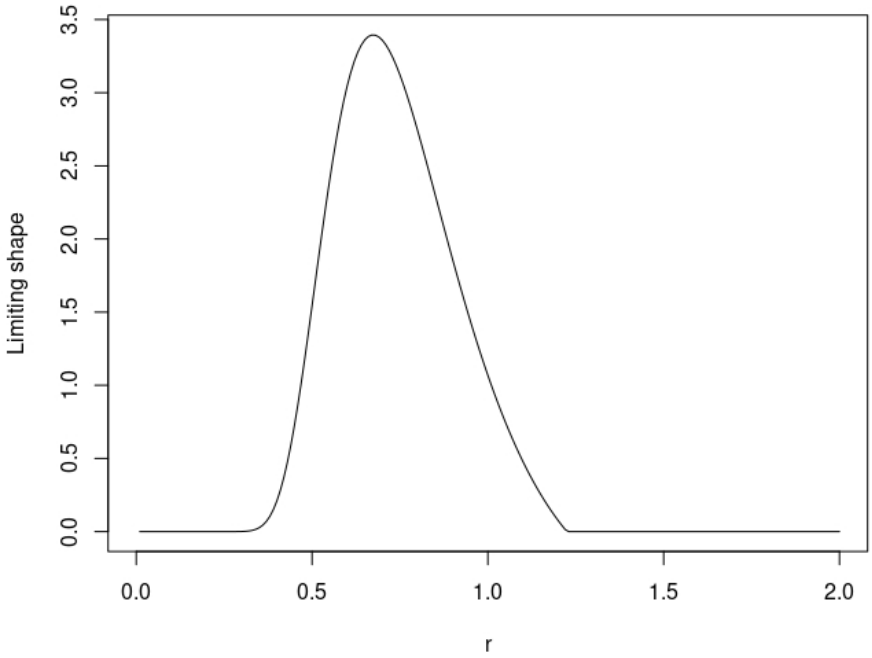


Figure 5.6: The limiting shape $\mu^*_{u,\infty}$ for the radial component of the minimal rebirth distribution for multivariate Cauchy targets. The root is at $r = \sqrt{6}/2$.

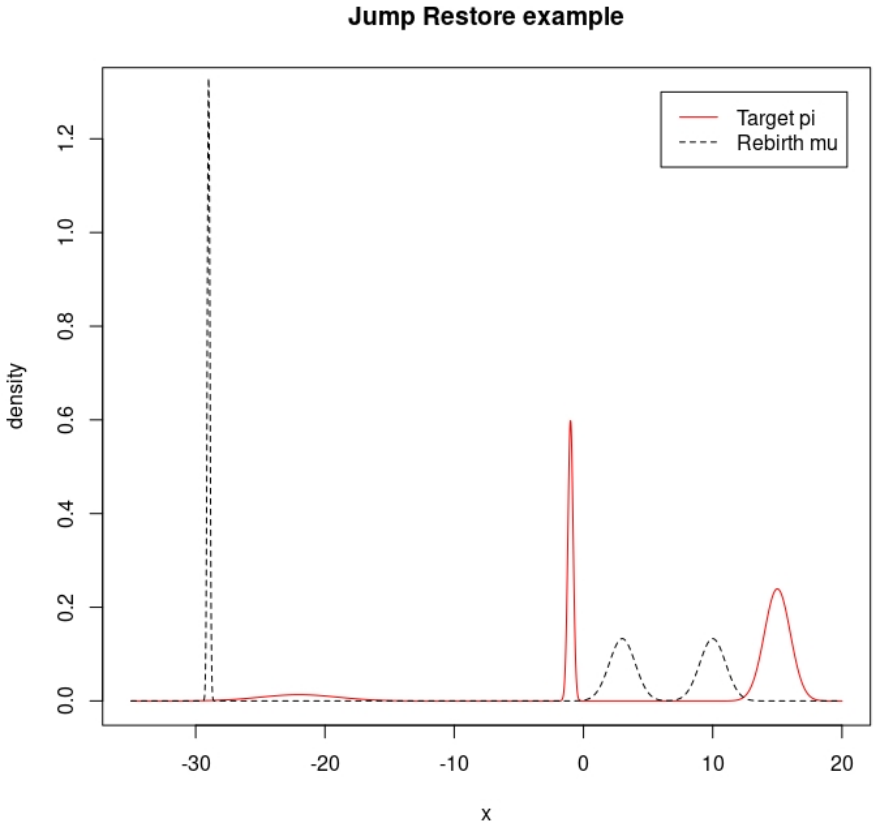


Figure 5.7: The target density π and rebirth density μ for the jump Restore example.

knowledge of their widths or relative weights. We take a univariate example for convenience and clarity of presentation, but this could easily extend to more complicated multivariate settings.

In this section we write $\phi(\cdot; \nu, \sigma^2)$ for the univariate Gaussian density with mean $\nu \in \mathbb{R}$ and variance $\sigma^2 > 0$. We take the following target density π on $\mathbb{R}$:

$$\pi(x) = 0.1\,\phi(x; -22, 3^2) + 0.3\,\phi(x; -1, 0.2^2) + 0.6\,\phi(x; 15, 1^2), \quad x \in \mathbb{R}.$$

For the rebirth density μ, we take

$$\mu(x) = \frac{1}{3}\left(\phi(x; -29, 0.1^2) + \phi(x; 3, 1^2) + \phi(x; 10, 1^2)\right), \quad x \in \mathbb{R}.$$

These are both plotted in Figure 5.7.

This is a situation where simple implementations of both random-walk Metropolis and the independence sampler with μ as a proposal are highly inefficient. RWM is able to explore individual modes, but the large separation of the modes means the sampler is unable to cross from one mode the other. The independence sampler struggles since

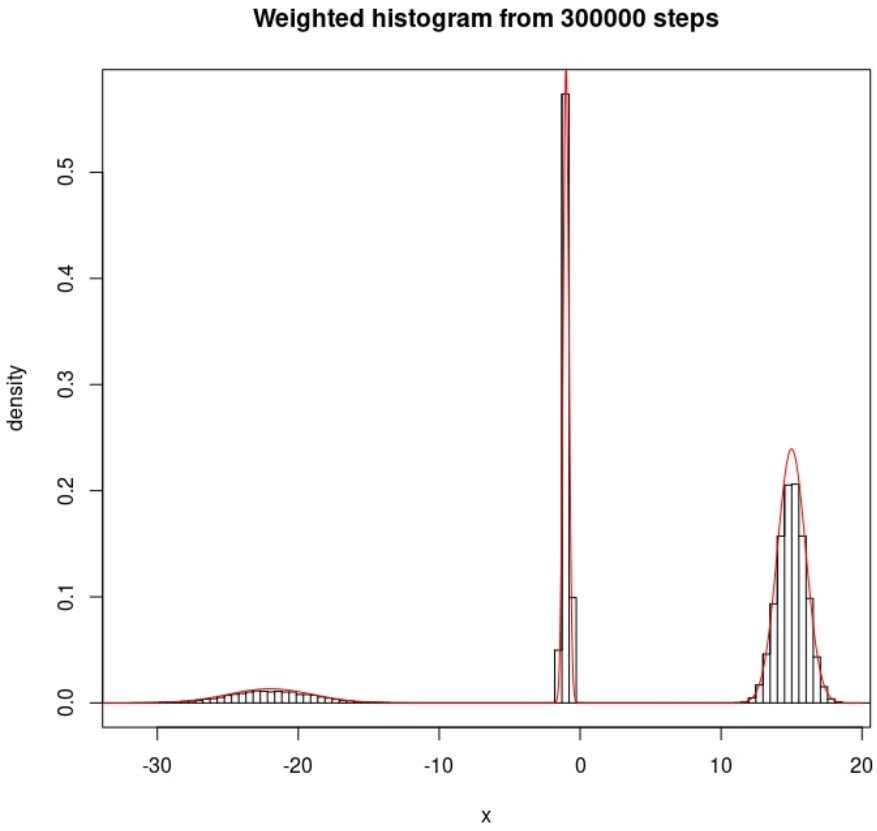


Figure 5.8: A weighted histogram of the Restore run, taking into account holding times.

μ is not aligned well with π: the sampler experiences very high rejection rates and is unable to explore the modes.

We implement the jump process variant of the Restore sampler as described in Section 5.3.3.1. We take the underlying process Y to be RWM with variance 1 embedded in continuous time, with constant holding rate 1. The algorithm is implemented as in Algorithm 3. We took the constant $C = 1$ in the killing rate.

We ran the Restore sampler for 300,000 steps of the jump chain. Restore is able to correct the discrepancy between the rebirth distribution and the target, and explore all three modes. The corresponding weighted histogram is in Figure 5.8. Since the process is continuous-time, we cannot naively plot a histogram of the outputted values of the chain, but must also take into account the holding times.

In Figure 5.9 we have plotted the continuous-time trajectory of the first 50,000 jump steps of this run. This includes teleports, and the stochastic process time elapsed was about 27,000. Among these 50,000 steps, a proportion of roughly 0.46

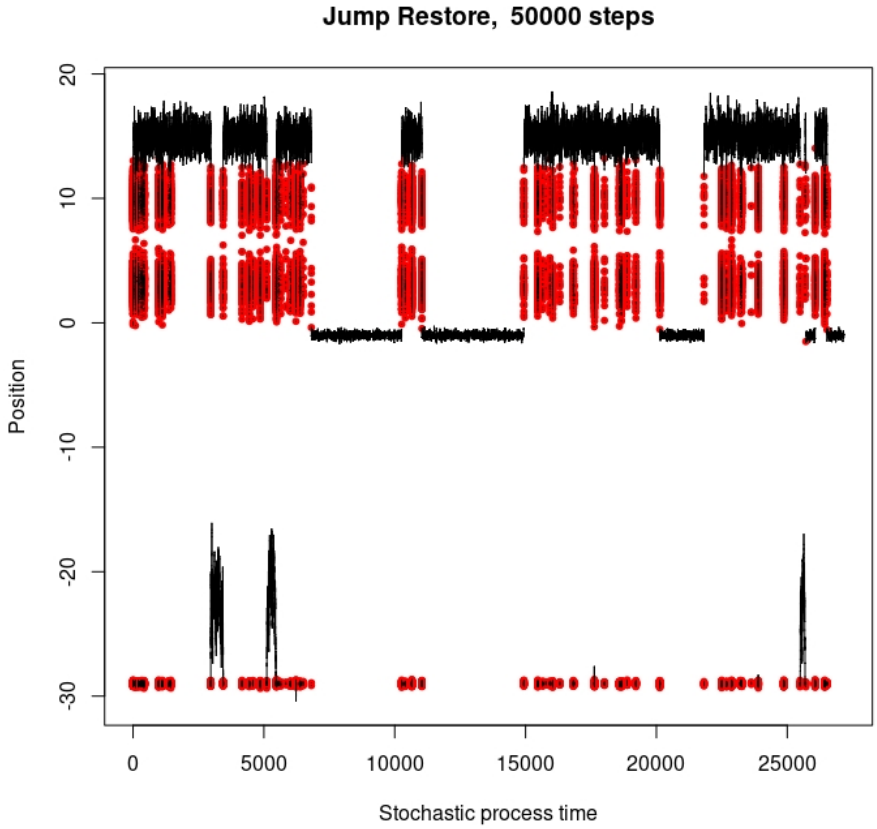


Figure 5.9: The continuous-time trajectory of the first 50,000 steps of the jump Restore sampler. The proportion of teleport moves was roughly 0.46. The red points are the teleport moves.

were teleport moves.

Because μ and π are not quite aligned, the sampler does exhibit rapid sequences of teleports, which correspond to rejections in a discrete-time setting. However, because of the non-teleport RWM moves, it is sometimes then able to move into the neighbouring mode and explore it.

Chapter 6

Conclusions and further work

In this thesis we have studied the theory relating to killing and regeneration in the context of continuous-time Monte Carlo samplers.

In Chapters 3 and 4 we focused on quasi-stationary Monte Carlo methods, where the idea is to construct a killed diffusion whose quasi-stationary distribution coincides with a target distribution of interest. This study was motivated by recent work of Pollock et al. [2016], where such methods were shown to scale well with the data size, while retaining exactness.

In Chapter 3 we provided some theoretical results to shore up the quasi-stationary theory in Pollock et al. [2016]. For general reversible diffusions, we gave conditions under which the quasi-stationary distribution of the killed diffusion coincides with a given target density. This simplifies and extends the theory given in Pollock et al. [2016]. We also gave a result on the spectral gap.

In Chapter 4 we considered an alternative implementation of QSMC, inspired by the sample path method of Aldous et al. [1988], Blanchet et al. [2016], Benaïm et al. [2018]. The resulting QSMC method has been dubbed ReScaLE, whose computational properties have been explored in the doctoral work of Kumar [2019]. In this chapter we provided the first proof of the convergence of such stochastic approximation methods to the quasi-stationary distribution in continuous-time on a continuous state space. We focused on compact spaces, but this topic has subsequently been studied further in Mailler and Villemonais [2018] and Benaïm et al. [2019]. However, the general case of diffusions on noncompact state spaces with potentially unbounded killing still remains an open problem.

QSMC methods have shown promise for tackling tall data problems, but several issues remain. For ReScaLE, the need to sample from the past trajectory poses a practical issue.

In Chapter 5 we pulled these strands together and introduced the Restore sampler. By fixing the rebirth distribution of ReScaLE, we overcome one of its key bottlenecks. We gave conditions under which the Restore process, an instance of a regenerative process, can be constructed to have a given target density as its invariant distribution. We proved various theoretical results related to the resulting sampler and gave some simple examples to demonstrate its potential.

We believe that the Restore sampler shows promise both as a new sampler in its own right, and as a recipe for potentially improving the mixing of existing samplers. This is because of its simplicity, and the desirable properties which follow from the presence of regenerations: the lack of burn-in issues, relatively straightforward variance estimates and its parallelisable nature.

However, of course, the Restore sampler as is has its limitations, presenting challenges and open questions for future research.

- While the global moves introduced are never rejected, in practice we may experience very short lifetimes. When one rebirth event is immediately followed by another, this can be seen as a kind of 'rejection'.

- In practice the selection of the constant C in (5.2) can be difficult. The calculations involved in finding $\tilde{\kappa}$, and then computing C can be onerous, especially as the dimension of the problem increases. There is the practical question of finding an appropriate value when implementing the sampler, and the associated theoretical question of what the optimal value is.

- The selection of μ is also a nontrivial problem. While we have suggested a natural choice of rebirth distribution in Section 5.5, in practice it may be very difficult to sample from. An interesting future development would be to take μ as an *approximation* of π, obtained via, say some preliminary MCMC simulations, or some general approximation such as INLA, Rue et al. [2009], then using Restore to correct the bias.

- We considered in Chapter 5 diffusions and jump processes for the underlying process. Another class of potential processes would be PDMPs or ODE flows.

- Since the event rate has been modified compared to ReScaLE, the Restore sampler is no longer amenable to subsampling, and so it is not necessarily suitable for tall data problems.

- Of course, a thorough analysis of the computational properties of Restore via more detailed examples still needs to be done.

Overall, I believe the study of continuous-time methods in the context of Monte Carlo is very valuable for improving existing Monte Carlo samplers and proposing new ones. Such methods have already proven popular for tackling the difficult data problems of the current age and I believe there are many more interesting and practical results and algorithms remaining to be discovered.

Appendix A

Symmetric diffusions

A.1 Basic Assumptions

For convenience we list here the Basic Assumptions of Stochastic Spectral Analysis (BASSA) from [Demuth and van Casteren, 2000, Chapter 1.B].

E is a second countable locally compact Hausdorff space with Borel σ-algebra $\mathcal{E}$, equipped with a Radon measure m. A function $f : E \to \mathbb{C}$ belongs to $C_0(E)$ if for each $\epsilon > 0$, there is a compact set K such that $|f(x)| \leq \epsilon$ for each $x \notin K$. $C_0(E)$ is always equipped with the supremum norm.

The function $p^0(t, x, y)$ defined on $(0, \infty) \times E \times E$ is a function such that $(t, x, y) \mapsto p^0(t, x, y)$ is continuous, and satisfies the following assumptions:

1. Markov property:

$$\int p^0(s, x, y) p^0(t, z, y) \, \mathrm{d}m(z) = p^0(s + t, x, y), \quad s, t > 0, x, y \in E,$$

   and

$$\int p^0(t, x, y) \, \mathrm{d}m(y) \leq 1, \quad t > 0, x \in E.$$

2. Feller property: for each $f \in C_0(E)$, the function

$$x \mapsto \int p^0(t, x, y) f(y) \, \mathrm{d}m(y)$$

   belongs to $C_0(E)$.

3. Continuity: for each $f \in C_0(E)$ and for each $x \in E$, we have

$$\lim_{t \downarrow 0} \int p^0(t, x, y) f(y) \, \mathrm{d}m(y) = f(x).$$

4. Symmetry: for each $t > 0$, $x, y \in E$, $p^0(t, x, y) = p^0(t, y, x)$.

Under BASSA, $p^0(t,x,y)$ is the kernel of a Feller semigroup, and we denote the $\mathcal{L}^2(E,m)$ generator of the semigroup by $-L^0$. Given $p^0(t,x,y)$ satisfying BASSA, then there exists a strong Markov process $(\Omega, \mathcal{F}, \theta_t, Y_t, \mathbb{P}_x)$ associated with it which has $\mathbb{P}_x$-almost surely càdlàg sample paths.

Then the semigroup $\{\exp(-tL^0) : t \geq 0\}$ is given by

$$[\exp(-tL^0)f](x) = \int p^0(t,x,y)f(y)\,\mathrm{d}m(y) = \mathbb{E}_x[f(Y_t)],$$

whenever this integral makes sense.

A.2 Theorem 2.5 of Demuth and van Casteren [2000]

In this subsection we state the relevant parts of Theorem 2.5 of Demuth and van Casteren [2000].

Let $V : E \to [0,\infty]$ be a Borel measurable function on E. The function V is said to belong to $K(E)$, the Kato class, if

$$\limsup_{t\downarrow 0} \sup_{x\in E} \int_0^t \left(\int p^0(t,x,y)V(y)\,\mathrm{d}m(y) \right) \mathrm{d}s = 0.$$

V is said to belong to $K_{loc}(E)$ if $1_B V$ belongs to $K(E)$ for all compact subsets B of E. Clearly bounded functions V belong to $K(E)$.

Theorem A.2.1 (Theorem 2.5 of Demuth and van Casteren [2000])**.** *We assume that $p^0(t,x,y)$ is a transition density satisfying BASSA, and that $V = V_+ - V_-$ is a Borel measurable function defined on E such that $V_- \geq 0$ belongs to $K(E)$ and such that $V_+ \geq 0$ belongs to $K_{loc}(E)$. Then the following statements hold.*

- *There exists a closed, densely defined linear operator $L^V = L^0 \dot{+} V$ in $C_0(E)$, extending $L^0 + V$, which generates a strongly continuous positivity-preserving semigroup $\left\{\exp\left(-tL^V\right) : t \geq 0\right\}$ in $C_0(E)$. Every operator $\exp(-tL^V)$, $t > 0$ is of the form*

$$\left[\exp\left(-tL^V\right) f\right](x) = \int p^V(t,x,y)f(y)\,\mathrm{d}m(y), \quad f \in C_0(E),$$

where $p^V(t,x,y)$ is a continuous function in (t,x,y) on $(0,\infty)\times E \times E$ which satisfies the Chapman–Kolmogorov identity:

$$p^V(s+t,x,y) = \int p^V(s,x,z)p^V(t,z,y)\,\mathrm{d}m(z),$$

for $t > 0$, $x, y \in E$. Furthermore, $p^V(t,x,y)$ is symmetric ([Demuth and van Casteren, 2000, p52]).

- *The operator* $\exp\left(-tL^V\right)$ *is given via the Feynman–Kac formula:*

$$\left[\exp\left(-tL^V\right)f\right](x) = \mathbb{E}_x\left[\exp\left(-\int_0^t V(Y_u)\,\mathrm{d}u\right)f(Y_t)\right].$$

- *The semigroup* $\left\{\exp\left(-tL^V\right) : t \geq 0\right\}$ *acts as a strongly continuous semigroup in* $\mathcal{L}^p(E,m)$ *for each* $1 \leq p < \infty$.

- *In* $\mathcal{L}^2(E,m)$, *the family* $\left\{\exp\left(-tL^V\right) : t > 0\right\}$ *is a self-adjoint positivity-preserving strongly continuous semigroup with a self-adjoint generator.*

- *The Feynman–Kac semigroup in* $\mathcal{L}^2(E,m)$ *coincides with the semigroup corresponding to the quadratic form* $\mathcal{E}^V$ *with domain*

$$\mathcal{D}\big(\mathcal{E}^V\big) = \mathcal{D}\left((L^0)^{1/2}\right) \cap \mathcal{D}\left(V_+^{1/2}\right)$$

and defined by

$$\mathcal{E}^V(f,g) = \langle (L^0)^{1/2}f, (L^0)^{1/2}g\rangle - \langle V_-^{1/2}f, V_-^{1/2}g\rangle + \langle V_+^{1/2}f, V_+^{1/2}g\rangle,$$

where $f, g \in \mathcal{D}(\mathcal{E}^V)$.

A.3 Proof of Lemma 5.3.6

Recall that we write $W^{2,1}(\mathbb{R}^d)$ for the Sobolev space of measurable functions on $\mathbb{R}^d$ whose first and second derivatives are integrable with respect to Lebesgue measure on $\mathbb{R}^d$, equipped with the corresponding Sobolev norm. Precise definitions can be found in Adams [1975].

First note that the fact that $\pi \in W^{2,1}(\mathbb{R}^d)$ along with the integral assumptions imply that

$$\int_{\mathbb{R}^d} \varphi(x)\kappa(x)\,\mathrm{d}\Gamma(x) < \infty,$$

since we can write

$$\varphi\kappa\gamma = \frac{1}{2}\Delta\pi - \nabla A\cdot\nabla\pi - \Delta A\pi + C\mu\gamma.$$

Now, as in the proof of Theorem 3.18 of Adams [1975], let f be a fixed smooth, nonnegative (take the square if necessary), compactly supported function on $\mathbb{R}^d$, a *mollifier*, with the following properties: for $x \in \mathbb{R}^d$,

- $f(x) = 1$ if $|x| \leq 1$,
- $f(x) = 0$ if $|x| \geq 2$,
- $|D^\alpha f(x)| \leq M$ (constant) for all $0 \leq |\alpha| \leq 2$.

Here we use the standard multi-index notation for derivatives; see Adams [1975] for the details.

We can now define for each $n \in \mathbb{N}$,

$$\varphi_n := f_n \cdot \varphi,$$

here $\cdot$ denotes the pointwise product and $f_n(x) := f(x/n), x \in \mathbb{R}^d$.

Since φ is smooth, the φ_n are a sequence of smooth, compactly supported functions with the following properties: $\pi_n := \gamma \cdot \varphi_n$ converges to $\pi = \gamma \cdot \varphi$ pointwise and in $W^{2,1}(\mathbb{R}^d)$ (see the proof of Theorem 3.18, Adams [1975]), and we have $\varphi_n \leq M\varphi$ pointwise, uniformly over n. This in particular implies that φ_n converges to φ in $\mathcal{L}^1(\Gamma)$.

We can write

$$-\gamma L^0 \varphi_n = \frac{1}{2}\Delta\pi_n - \nabla A \cdot \nabla \pi_n - \Delta A \, \pi_n.$$

From this we see that $-\gamma L^0 \varphi_n$ converges in $\mathcal{L}^1(\mathbb{R}^d)$ – $\mathbb{R}^d$ equipped with Lebesgue measure – to

$$\frac{1}{2}\Delta\pi - \nabla A \cdot \nabla \pi - \Delta A \, \pi.$$

Convergence of the first term is immediate since π_n converges to π in $W^{2,1}(\mathbb{R})$.

Convergence of the third term follows since π converges to π pointwise, with $\pi_n \leq M\pi$, so we can make use of the dominated convergence theorem.

Convergence of the second term $\nabla A \cdot \nabla \pi_n$ is obtained by the following argument. We have

$$\nabla A \cdot \nabla \pi_n = f_n \nabla A \cdot \nabla \pi + \pi \nabla A \cdot \nabla f_n. \tag{A.1}$$

Of the two terms on the right-hand side, the first converges in $\mathcal{L}^1(\mathbb{R}^d)$ to $\nabla A \cdot \nabla \pi$ straightforwardly; for instance by dominated convergence; the mollifiers are uniformly bounded by M.

For the second term of (A.1), first note that $\nabla f_n(x) = n^{-1} \nabla f(x/n)$, which is bounded (in each component) by M/n, and the support of ∇f_n is by construction within the set $B_n := \{y \in \mathbb{R}^d : n \leq |y| \leq 2n\}$. Thus we have

$$|\pi(x) \nabla A \cdot \nabla f_n(x)| \leq \pi(x) K |x| M n^{-1} \, 1_{B_n}(x) \leq 2KM\pi(x).$$

Here we used the bound on the drift $|\nabla A(x)| \leq K|x|$. Thus we can use once more dominated convergence to establish that $\pi \nabla A \cdot \nabla f_n$ is converging in $\mathcal{L}^1(\mathbb{R}^d)$ to the zero function. This establishes the convergence of the second term $\nabla A \cdot \nabla \pi_n$.

Now by rewriting the terms, we have that $-L^0 \varphi_n$ converges to

$$\frac{1}{2}\Delta\varphi + \nabla A \cdot \nabla \varphi$$

in $\mathcal{L}^1(\Gamma)$.

Thus since L^0 is a closed operator, and the sequence of smooth compactly supported function π_n belongs to $\mathcal{D}(L_1^0)$ for each n, we have established that $\varphi \in \mathcal{D}(L_1^0)$.

Finally, since $\int \pi\kappa \, \mathrm{d}\Gamma < \infty$, and we have that $\pi_n \leq M\pi$, we have that $\int \pi_n \kappa \, \mathrm{d}\Gamma \to \int \pi\kappa \, \mathrm{d}\Gamma$.

Thus

$$L^\kappa \pi_n \to L^0 \pi + \kappa\pi = C\mu$$

in $\mathcal{L}^1(\Gamma)$. This shows that $\varphi \in \mathcal{D}(L_1^\kappa)$ and $L_1^\kappa \varphi = C\mu$, concluding the proof of Lemma 5.3.6.

A.4 Proof of Theorem 5.3.7

We want to prove that φ is an invariant distribution for the Restore process X with interarrival dynamics Y, killing rate κ as defined in (5.7) with rebirth distribution μ. We are in the setting $(E, m) = (\mathbb{R}^d, \Gamma)$.

Since we are assuming (Assumption 4) that κ is continuous, hence bounded on compact sets, and nonnegative, it is easy to see that the potential $V = \kappa$ belongs to $K_{loc}(\mathbb{R}^d)$, as defined in Appendix A.2. We are assuming that our underlying process Y satisfies BASSA (Assumption 10), hence we can apply Theorem 2.5 of Demuth and van Casteren [2000], reproduced above as Theorem A.2.1.

Thus we obtain the existence of a semigroup $\{\exp(-tL^\kappa) : t \geq 0\}$ with continuous, symmetric kernel $p^\kappa(t, x, y)$ on $(0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d$. This semigroup is strongly continuous on $C_0(\mathbb{R}^d)$ as well as $\mathcal{L}^p(\Gamma)$ for each $1 \leq p < \infty$.

We know that the Restore process X is a strong Markov process by Proposition 5.2.1. Let $\{P_t^\mu : t \geq 0\}$ denotes its semigroup.

Our goal is to consider for appropriate functions f,

$$t \mapsto \varphi P_t^\mu f := \int \mathrm{d}\Gamma(x)\varphi(x)[P_t^\mu f](x) = \mathbb{E}_\varphi[f(X_t)].$$

We will compute the time derivative and show that is it 0. By time homogeneity, it is sufficient to show that the derivative is 0 at $t = 0$. This was the method used to

prove φ-invariance of the Bouncy Particle Sampler in the supplementary material of Bouchard-Côté et al. [2018].

The Restore process naturally exhibits a renewal-process like behaviour since the individual lifetimes are independent and identically distributed. So we will seek a renewal-like representation of the semigroup P_t^μ by conditioning on the first arrival τ_∂. Since κ is locally bounded, τ_∂ is absolutely continuous on $\mathbb{R}^+$, hence will possess a density with respect to Lebesgue measure on $\mathbb{R}^+$.

Since κ is nonnegative, the semigroup $\exp(-tL^\kappa)$ can also be expressed as

$$[\exp(-tL^\kappa)f](x) = \mathbb{E}_x\left[f(Y_t)1\{\tau_\partial > t\}\right],$$

where τ_∂ is defined as in (4.2) for each f where the integral is well-defined.

Note that we have

$$-L^\kappa\varphi = -L^0\varphi - \kappa\varphi = -C\mu. \tag{A.2}$$

This equation holds formally, where we view L^κ and L^0 as formal differential operators, and rigourously as the $\mathcal{L}^1(\Gamma)$ generator by Assumption 13. Since we additionally assume that in Assumption 13 that both $\varphi, \mu \in \mathcal{L}^2(\Gamma)$, it follows that (A.2) also holds rigourously as the $\mathcal{L}^2(\Gamma)$ generator.

Consider

$$\begin{aligned}\mathbb{P}_\varphi(\tau_\partial > t) &= \int \mathrm{d}\Gamma(x)\varphi(x)\int \mathrm{d}\Gamma(y)\, p^\kappa(t,x,y)\\ &= \int \mathrm{d}\Gamma(y)\int \mathrm{d}\Gamma(x)\,\varphi(x)p^\kappa(t,y,x) = \int \mathrm{d}\Gamma(y)[\exp(-tL^\kappa)\varphi](y).\end{aligned}$$

Here we used Tonelli's theorem to exchange the order of integration and symmetry of $p^\kappa(t,x,y)$. The final integral is well-defined since $\varphi \in \mathcal{L}^1(\Gamma)$ and the semigroup $\exp(-tL^\kappa)$ maps $\mathcal{L}^1(\Gamma)$ to itself by Theorem A.2.1. Thus by strong continuity and the fact that $\varphi \in \mathcal{D}(L_1^\kappa)$ (Assumption 13) we can differentiate this expression to find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_\varphi(\tau_\partial > t)|_{t=s} &= \int \mathrm{d}\Gamma(y)[\exp(-sL^\kappa)(-L^\kappa\varphi)](y)\\ &= -\int \mathrm{d}\Gamma(y)[\exp(-sL^\kappa)(C\mu)](y)\\ &= -C\int \mathrm{d}\Gamma(x)\mu(x)[\exp(-sL^\kappa)1](x).\end{aligned}$$

Here we have used (A.2) and Assumption 13 and symmetry again.

This shows that the density on $\mathbb{R}^+$ with respect to Lebesgue measure of the first arrival time under $\mathbb{P}_\varphi$ is given by

$$h(t) = C\int \mathrm{d}\Gamma(x)\mu(x)[\exp(-tL^\kappa)1](x) = C\,\mathbb{P}_\mu(\tau_\partial > t), \quad t \geq 0.$$

This also shows that $C = 1/\mathbb{E}_\mu[\tau_\partial]$.

From this we can represent the semigroup of the Restore process started in φ as

$$\begin{aligned}\varphi P_t^\mu f &= \int_0^t C\mathbb{P}_\mu(\tau_\partial > s)\, \mu P_{t-s}^\mu f \,\mathrm{d}s + \varphi \exp(-tL^\kappa) f \\ &= C\int_0^t \mathbb{P}_\mu(\tau_\partial > t-s)\, \mu P_s^\mu f \,\mathrm{d}s + \varphi \exp(-tL^\kappa) f.\end{aligned}$$

Our goal is to differentiate this expression with respect to t and show that the derivative at $t = 0$ is zero. Now we take $f \in \mathcal{D}(L_2^\kappa)$ and bounded. From this representation, we can see that $t \mapsto \varphi P_t^\mu f$ is a continuous function.

We wish to compute the time derivative of $t \mapsto \varphi P_t^\mu f$ using Leibniz's rule. Consider

$$\begin{aligned}t \mapsto \mathbb{P}_\mu(\tau_\partial > t) &= \int \mathrm{d}\Gamma(x)\mu(x) \int p^\kappa(t,x,y)\,\mathrm{d}\Gamma(y) \\ &= \mathbb{E}_\mu^0\left[\exp\left(-\int_0^t \kappa(Y_s)\,\mathrm{d}s\right)\right].\end{aligned}$$

We would like to compute the time derivative. Because of our technical assumption (5.9), we can differentiate under the integral sign to obtain

$$g(s) := -\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_\mu(\tau_\partial > t)|_{t=s} = \mathbb{E}_\mu^0\left[\kappa(Y_s)\exp\left(-\int_0^s \kappa(Y_u)\,\mathrm{d}u\right)\right],$$

for each $s \in [0,1]$. g is a continuous function, and will be uniformly bounded over $s \in [0,1]$.

Thus we obtain in a similar manner to above, the representation

$$\mu P_t^\mu f = \int_0^t g(s)\, \mu P_{t-s}^\mu f \,\mathrm{d}s + \varphi \exp(-tL^\kappa) f.$$

So $t \mapsto \mu P_t^\mu f$ is a continuous function.

So now we can compute time derivatives by Leibniz's rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi P_t^\mu f\bigg|_{t=s} = C\int_0^s g(s-u)\, \mu P_u^\mu f \,\mathrm{d}u + C\mu(1)\mu P_s^\mu f + \varphi\exp(-tL^\kappa)(-L^\kappa f).$$

Taking $t = 0$, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi P_t^\mu f\bigg|_{t=0} = C\mu(f) + \varphi(-L^\kappa f).$$

Since we chose $f \in \mathcal{D}(L_2^\kappa)$ and using the fact that $\varphi \in \mathcal{D}(L_2^\kappa)$ (Assumption 11), this final expression is equal to

$$C\mu(f) + \varphi(-L^\kappa f) = \int \mathrm{d}\Gamma(x)\, f(x) \left(C\mu(x) + (-L^0\varphi)(x) - \kappa(x)\varphi(x) \right).$$

This will equal 0 for any such f if

$$\kappa(x) = \frac{-L^0\varphi}{\varphi}(x) + C\frac{\mu(x)}{\varphi(x)}, \quad x \in \mathbb{R}^d,$$

which is exactly our (5.7). This concludes the proof of Theorem 5.3.7.

Bibliography

Robert A. Adams. *Sobolev Spaces*, volume 65 of *Pure and applied mathematics*. Academic Press, New York, 1st edition, 1975. ISBN 0120441500.

David Aldous, Barry Flannery, and Jose Luis Palacios. Two applications of urn processes: the fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. *Probab. Eng. Inf. Sci.*, 2(03):293–307, 1988.

Søren Asmussen. *Applied Probability and Queues*, volume 51 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, 2nd edition, 2003. ISBN 0387002111.

Andrew D. Barbour and Phil K. Pollett. Total Variation Approximation for Quasi-Stationary Distributions. *J. Appl. Probab.*, 47(4):934–946, dec 2010. ISSN 0021-9002. doi: 10.1239/jap/1294170510.

Andrew D. Barbour and Phil K. Pollett. Total variation approximation for quasi-equilibrium distributions, II. *Stochastic Processes Appl.*, 230(1):1–14, 2012. ISSN 03044149. doi: 10.1016/j.spa.2012.07.004.

Rémi Bardenet, Arnaud Doucet, and Chris Holmes. On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.*, 18:1–43, 2017.

M. S. Bartlett. *Stochastic Population Models*. Methuen's Monographs on Applied Probability and Statistics. Methuen& Co. Ltd., London, 1960.

Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pages 1–68. Springer, Berlin, 1999.

Michel Benaïm and Bertrand Cloez. A stochastic approximation approach to quasi-stationary distributions on finite spaces. *Electron. Commun. Probab.*, 20(37):1–14, 2015. doi: 10.1214/ECP.v20-3956.

Michel Benaïm, Michel Ledoux, and Olivier Raimond. Self-interacting diffusions. *Probab. Theory Related Fields*, 122:1–41, 2002. doi: 10.1007/s004400100161.

Michel Benaïm, Bertrand Cloez, and Fabien Panloup. Stochastic approximation of quasi-stationary distributions on compact spaces and applications. *Ann. Appl. Probab.*, 28(4):2370–2416, 2018. ISSN 1050-5164. doi: 10.1214/17-AAP1360.

Michel Benaïm, Nicolas Champagnat, and Denis Villemonais. Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain. *arXiv preprint*, apr 2019. URL `http://arxiv.org/abs/1904.08620`.

Alexandros Beskos, Omiros Papaspiliopoulos, and Gareth O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6): 1077–1098, 12 2006. ISSN 1350-7265. doi: 10.3150/bj/1165269151.

Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv preprint*, 2017. URL `http://arxiv.org/abs/1701.02434`.

Joris Bierkens. Non-reversible Metropolis-Hastings. *Stat. Comput.*, 26(6):1213–1228, 2016. ISSN 0960-3174. doi: 10.1007/s11222-015-9598-x.

Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Stat.*, 47(3):1288–1320, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1715.

Jean-Michel Bismut. *Large deviations and the Malliavin calculus*, volume 45 of *Progress in Mathematics*. Birkhäuser, 1984. ISBN 0817632204.

Jose Blanchet, Peter Glynn, and Shuheng Zheng. Analysis of a stochastic approximation algorithm for computing quasi-stationary distributions. *Adv. Appl. Probab.*, 48(10):792–811, 2016. doi: 10.1017/apr.2016.28.

Alexandre Bouchard-Côté, Sebastian J. Vollmer, and Arnaud Doucet. The Bouncy Particle Sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Am. Stat. Assoc.*, 113(522):855–867, 2018. ISSN 0162-1459. doi: 10.1080/01621459.2017.1294075.

Maxim Braverman, Ognjen Milatovic, and Mikhail Shubin. Essential self-adjointness of Schrödinger-type operators on manifolds. *Russ. Math. Surv.*, 57(4):641–692, 8 2002. ISSN 0036-0279. doi: 10.1070/RM2002v057n04ABEH000532.

Anthony E. Brockwell and Joseph B. Kadane. Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes. *J. Comput. Graphical Stat.*, 14(2):436–458, jun 2005. ISSN 1061-8600. doi: 10.1198/106186005X47453.

Krzysztof Burdzy, Robert Hołyst, and Peter March. A Fleming–Viot Particle Representation of the Dirichlet Laplacian. *Commun. Math. Phys.*, 214(3):679–703, 2000. doi: 10.1007/s002200000294.

Pietro Caputo and Matteo Quattropani. Mixing time of PageRank surfers on sparse random digraphs. *arXiv preprint*, 2019. URL `http://arxiv.org/abs/1905.04993`.

Nicolas Champagnat and Denis Villemonais. Exponential convergence to quasi-stationary distribution and Q-process. *Probab. Theory Related Fields*, 164(1-2): 243–283, 2016. ISSN 0178-8051. doi: 10.1007/s00440-014-0611-7.

Nicolas Champagnat and Denis Villemonais. General criteria for the study of quasi-stationarity. *arXiv preprint*, 2017. URL `https://arxiv.org/pdf/1712.08092.pdf`.

Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications*. Probability and its Applications. Birkhäuser, Basel, 2 edition, 2015. ISBN 9781493928668.

Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*. Probab. Appl. Springer-Verlag, Berlin, Heidelberg, 2013. ISBN 9783642331305.

Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017. ISSN 1369-7412. doi: 10.1111/rssb.12183.

John N. Darroch and Eugene Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probab.*, 2(1):88–100, jun 1965. ISSN 00219002. doi: 10.2307/3211876.

E. Brian Davies. *Spectral Theory and Differential Operators*. Cambridge University Press, Cambridge, 1995. ISBN 9780511623721. doi: 10.1017/CBO9780511623721.

Mark H. A. Davis. Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *J. Roy. Stat. Soc. Ser. B (Stat. Methodol.)*, 46(3): 353–376, 1984. ISSN 00359246. doi: 10.1111/j.2517-6161.1984.tb01308.x.

Marcelo Martins de Oliveira and Ronald Dickman. How to simulate the quasi-stationary state. *Phys. Rev. E*, 71(1):016129, jan 2005. ISSN 1539-3755. doi: 10.1103/PhysRevE.71.016129.

Pierre Del Moral. *Feynman-Kac Formulae*. Probability and its Applications. Springer–Verlag, New York, 1st edition, 2004. ISBN 978-0-387-20268-6. doi: 10.1007/978-1-4684-9393-1.

Pierre Del Moral and Laurent Miclo. Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman–Kac semigroups. *ESAIM: Probab. Stat.*, 7:171–208, mar 2003. ISSN 1292-8100. doi: 10.1051/ps:2003001.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *J. Roy. Stat. Soc. Ser. B (Stat. Methodol.)*, 68(3):411–436, jun 2006. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2006.00553.x.

Michael Demuth and Jan A. van Casteren. *Stochastic Spectral Theory for Selfadjoint Feller Operators: a functional integration approach*. Birkhäuser Basel, 2000. ISBN 9783034884600.

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986. ISBN 1461386438.

Persi Diaconis and Laurent Miclo. On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse*, 4:973–1016, 2014.

Persi Diaconis, Susan Holmes, and Radford M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000. ISSN 1050-5164. doi: 10.1214/aoap/1019487508.

Ronald Dickman and Marcelo Martins de Oliveira. Quasi-stationary simulation of the contact process. *Physica A: Statistical Mech. Appl.*, 357(1):134–141, 2005. ISSN 03784371. doi: 10.1016/j.physa.2005.05.051.

Joseph L. Doob. Markoff Chains–Denumerable Case. *Trans. Am. Math. Soc.*, 58(3): 455, nov 1945. ISSN 00029947. doi: 10.2307/1990339.

Randal Douc and Christian P. Robert. A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *The Annals of Statistics*, 39(1):261–277, 2011. ISSN 0090-5364. doi: 10.1214/10-AOS838.

Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10(3):197–208, 2000. ISSN 09603174. doi: 10.1023/A:1008935410038.

Douglas Down, Sean P. Meyn, and Richard L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4):1671–1691, 1995.

Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987. ISSN 03702693. doi: 10.1016/0370-2693(87)91197-X.

Richard M. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. ISBN 052180972X.

Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017. ISSN 1050-5164. doi: 10.1214/16-AAP1238.

F. Eggenberger and G. Pólya. Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.*, 3(4):279–289, 1923. doi: 10.1002/zamm.19230030407.

K. David Elworthy. *Stochastic Differential Equations on Manifolds*, volume 70 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1982. ISBN 9781107325609. doi: 10.1017/CBO9781107325609.

Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. Wiley, 1986. ISBN 9780470316658.

Warren J. Ewens. The Diffusion Equation and a Pseudo-Distribution in Genetics. *J. Roy. Stat. Soc.: Ser. B (Methodol.)*, 25(2):405–412, jul 1963. ISSN 00359246. doi: 10.1111/j.2517-6161.1963.tb00524.x.

Warren J. Ewens. The pseudo-transient distribution and its uses in genetics. *J. Appl. Probab.*, 1(1):141–156, jun 1964. ISSN 0021-9002. doi: 10.2307/3212065.

Paul Fearnhead, Joris Bierkens, Murray Pollock, and Gareth O. Roberts. Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo. *Stat. Sci.*, 33 (3):386–412, 2018. ISSN 0883-4237. doi: 10.1214/18-STS648.

William Feller. Diffusion processes in one dimension. *Trans. Am. Math. Soc.*, 77(1): 1–31, 1954. ISSN 0002-9947. doi: 10.1090/S0002-9947-1954-0063607-6.

Pablo A. Ferrari, Harry Kesten, Servet Martinez, and Pedro Picco. Existence of Quasi-Stationary Distributions. A Renewal Dynamical Approach. *Ann. Probab.*, 23(2):501–521, 1995. ISSN 0091-1798. doi: 10.1214/AOP/1176988277.

Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, 3rd edition, 2001. ISBN 0198572239.

Pablo Groisman and Matthieu Jonckheere. Simulation of quasi-stationary distributions on countable spaces. *Markov Processes Related Fields*, 19(3):521–542, 2013.

Alice Guionnet and Boguslaw Zegarlinski. Lectures on Logarithmic Sobolev Inequalities. In *Séminaire de probabilités (Strasbourg)*, volume 36, pages 1–134. Springer-Verlag, 2002. doi: 10.2307/2373688.

Wilfred K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. ISSN 1464-3510. doi: 10.1093/biomet/57.1.97.

James P. Hobert, Galin L. Jones, Brett Presnell, and Jeffrey S. Rosenthal. On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89(4):731–743, dec 2002. ISSN 0006-3444. doi: 10.1093/biomet/89.4.731.

Martin Kolb and David Steinsaltz. Quasilimiting behavior for one-dimensional diffusions with killing. *Ann. Probab.*, 40(1):162–212, 2012. ISSN 00911798. doi: 10.1214/10-AOP623.

Divakar Kumar. *On a Stochastic Approximation and Quasi-stationary Approach for the Simulation of Bayesian Posterior Distributions, with Application to Tall Data*. PhD thesis, University of Warwick, 2019.

Aline Kurtzmann. The ODE method for some self-interacting diffusions on $\mathbb{R}^d$. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):618–643, 2010. doi: 10.1214/09-AIHP206.

Harold J. Kushner and George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, volume 35 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, New York, 2 edition, 2003.

Krzysztof Latuszynski. *Regeneration and Fixed-Width Analysis of Markov Chain Monte Carlo Algorithms*. PhD thesis, University of Warsaw, 2009. URL `http://arxiv.org/abs/0907.4716`.

Anthony Lee, Arnaud Doucet, and Krzysztof Latuszyński. Perfect simulation using atomic regeneration with application to Sequential Monte Carlo. *arXiv preprint*, 2014. URL `http://arxiv.org/abs/1407.5770`.

Cécile Mailler and Denis Villemonais. Stochastic approximation on non-compact measure spaces and application to measure-valued Pólya processes. *arXiv preprint*, 2018. URL `http://arxiv.org/abs/1809.01461`.

Petr Mandl. Spectral theory of semi-groups connected with diffusion processes and its application. *Czechoslovak Math. J.*, 11(4):558–569, 1961.

Servet Martinez and Jaime San Martin. Quasi-stationary distributions for a Brownian motion with drift and associated limit. *J. Appl. Probab.*, 31(4):911–920, 1994.

Servet Martinez, Pierre Picco, and Jaime San Martin. Domain of attraction of quasi-stationary distributions for the brownian motion with drift. *Adv. Appl. Probab.*, 30 (2):385–408, jun 1998. ISSN 0001-8678. doi: 10.1239/aap/1035228075.

Sylvie Méléard and Denis Villemonais. Quasi-stationary distributions and population processes. *Probab. Surv.*, 9(0):340–410, 2012. ISSN 1549-5787. doi: 10.1214/11-PS191.

Giorgio Metafune, Diego Pallara, and Enrico Priola. Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures. *J. Funct. Anal.*, 196(1): 40–60, 2002. ISSN 00221236. doi: 10.1006/jfan.2002.3978.

Nicholas Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science*, pages 125–130, 1987. Special Issue dedicated to Stanislaw Ulam.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, jun 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.

Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993. URL `http://probability.ca/MT/`.

Sean P. Meyn and Richard L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, nov 1994. ISSN 1050-5164. doi: 10.1214/aoap/1177004900.

Manon Michel, Sebastian C. Kapfer, and Werner Krauth. Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.*, 140(5):054116, feb 2014. ISSN 0021-9606. doi: 10.1063/1.4863991.

Manon Michel, Johannes Mayer, and Werner Krauth. Event-chain Monte Carlo for classical continuous spin models. *EPL (Europhysics Letters)*, 112(2):20003, oct 2015. ISSN 0295-5075. doi: 10.1209/0295-5075/112/20003.

Do Le (Paul) Minh, David D. L. Minh, and Andrew L. Nguyen. Regenerative Markov chain Monte Carlo for any distribution. *Commun. Stat. - Simulation Computation*, 41(9):1745–1760, oct 2012. ISSN 0361-0918. doi: 10.1080/03610918.2011.615433.

José E. Moyal. Discontinuous Markoff processes. *Acta Mathematica*, 98(0):221–264, 1957. ISSN 0001-5962. doi: 10.1007/BF02404475.

Duncan J. Murdoch. Exact sampling for Bayesian inference: Unbounded state spaces. In *Monte Carlo Methods*, volume 26 of *Fields Institute Communications*, pages 111–121. American Mathematical Society, Providence, RI, 2000. doi: 10.1090/fic/026/08.

Duncan J. Murdoch and Peter J. Green. Exact Sampling from a Continuous State Space. *Scandinavian J. Stat.*, 25(3):483–502, sep 1998. ISSN 0303-6898. doi: 10.1111/1467-9469.00116.

Per Mykland, Luke Tierney, and Bin Yu. Regeneration in Markov chain samples. *J. Am. Stat. Assoc.*, 90(429):233–241, 1995.

Radford M. Neal. MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC, 2011.

Andrew L. Nguyen. Regenerative Markov chain importance sampling. *Commun. Stat. - Simulation Computation*, pages 1–15, jun 2015. ISSN 0361-0918. doi: 10.1080/03610918.2015.1043383.

Esa Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie Verwandte Gebiete*, 43(4):309–318, 1978. doi: 10.1007/BF00534764.

El Maati Ouhabaz. *Analysis of Heat Equations on Domains*. Princeton University Press, 2005. ISBN 9781400826483.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999. URL `http://ilpubs.stanford.edu:8090/422`.

E. A. J. F. (Frank) Peters and Gijsbertus de With. Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E*, 85(2):026703, 2012. ISSN 1539-3755. doi: 10.1103/PhysRevE.85.026703.

Ross G. Pinsky. On the Convergence of Diffusion Processes Conditioned to Remain in a Bounded Region for Large Time to Limiting Positive Recurrent Diffusion Processes. *Ann. Probab.*, 13(2):363–378, 1985. ISSN 0091-1798. doi: 10.1214/aop/1176992996.

Ross G. Pinsky. *Positive harmonic functions and diffusion*. Cambridge University Press, 1995. ISBN 0521470145.

Ross G. Pinsky. Explicit and almost explicit spectral calculations for diffusion operators. *J. Funct. Anal.*, 256(10):3279–3312, 2009. ISSN 00221236. doi: 10.1016/j.jfa.2008.08.012.

Phil K. Pollett. Quasi-stationary Distributions: A Bibliography, 2015. URL `http://www.maths.uq.edu.au/~pkp/papers/qsds/qsds.html`.

Murray Pollock, Paul Fearnhead, Adam M. Johansen, and Gareth O. Roberts. The Scalable Langevin Exact Algorithm: Bayesian inference for big data. *arXiv preprint*, 2016. URL `https://arxiv.org/abs/1609.03436`.

James G. Propp and David B. Wilson. Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Random Struct. Algorithms*, 9(2):223–252, 1996.

Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics IV: Analysis of Operators*. Academic Press, London, 1978. ISBN 978-0125850049.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, Heidelberg, 1 edition, 1991. ISBN 3540521674.

Christian Robert and George Casella. A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statist. Sci.*, 26(1):102–115, 2011. doi: 10.1214/10-STS351.

Christian P. Robert and George. Casella. *Monte Carlo Statistical Methods*. Springer–Verlag, New York, 2 edition, 2004. ISBN 0387212396. doi: 10.1007/978-1-4757-4145-2.

Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Stat. Soc.: S. B (Stat. Methodol.)*, 60(1): 255–268, 1998. ISSN 1369-7412. doi: 10.1111/1467-9868.00123.

Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004. doi: 10.1214/154957804100000024.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341, 12 1996. ISSN 13507265. doi: 10.2307/3318418.

Leonard C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales*. Cambridge University Press, 2000. ISBN 0521775949.

Jeffrey S. Rosenthal. Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *J. Am. Stat. Assoc.*, 90(430):558–566, jun 1995. ISSN 0162-1459. doi: 10.1080/01621459.1995.10476548.

Havard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian Inference for Latent Gaussian Methods by Using INLA. *J. Roy. Stat. Soc.: Ser. B (Methodol.)*, 71(2):319–392, 2009.

Sebastian Schreiber, Jonathan M. Levine, Oscar Godoy, Nathan Kraft, and Simon Hart. Does deterministic coexistence theory matter in a finite world? Insights from serpentine annual plants. *bioRxiv preprint*, 2018. doi: 10.1101/290882.

Michael Sharpe. *General Theory of Markov Processes*. Pure and applied mathematics. Academic Press, 1988. ISBN 0126390606.

Barry Simon. Large time behaviour of the heat kernel: on a theorem of Chavel and Karp. *Proc. Am. Math. Soc.*, 118(2):513–514, 1993.

Walter L. Smith. Regenerative stochastic processes. *Proc. R. Soc. London, Ser. A. Math. Phys. Sci.*, 232(1188):6–31, oct 1955. doi: 10.1098/rspa.1955.0198.

David Steinsaltz and Stephen N Evans. Quasistationary distributions for one-dimensional diffusions with killing. *Trans. Amer. Math. Soc.*, 359(3):1285—-1324, 2007. doi: https://doi.org/10.1090/S0002-9947-06-03980-8.

Daniel W. Stroock. *An Introduction to the Analysis of Paths on a Riemannian Manifold*, volume 74 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000. ISBN 9780821838396.

Luke Tierney. Introduction to general state-space Markov chain theory. In W.R. Gilks, S. Richardson, and David J. Spiegelhalter, editors, *Markov chain Monte Carlo in practice*, chapter 4, pages 59—-74. Chapman & Hall, 1996.

Pekka Tuominen and Richard L. Tweedie. Exponential decay and ergodicity of general Markov processes and their discrete skeletons. *Adv. Appl. Probab.*, 11(4):784–803, 1979.

Richard L. Tweedie. R-theory for Markov Chains on a general state space I: solidarity properties and R-recurrent chains. *Ann. Probab.*, 2(5):840–864, 10 1974. ISSN 0091-1798. doi: 10.1214/aop/1176996552.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. ISBN 9780521784504.

Denis Villemonais. Interacting particle systems and Yaglom limit approximation of diffusions with unbounded drift. *Electron. J. Probab.*, 16:1663–1692, 2011.

Andi Q. Wang, Martin Kolb, Gareth O. Roberts, and David Steinsaltz. Theoretical properties of quasi-stationary Monte Carlo methods. *Ann. Appl. Probab.*, 29(1): 434–457, 2019a. ISSN 1050-5164. doi: 10.1214/18-AAP1422.

Andi Q. Wang, Gareth O. Roberts, and David Steinsaltz. An approximation scheme for quasi-stationary distributions of killed diffusions. *Stochastic Processes Appl.*, 2019b. URL `https://doi.org/10.1016/j.spa.2019.09.010`. To appear.

Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proc. 28th Int. Conf. Mach. Learn.*, pages 681–688, 2011.