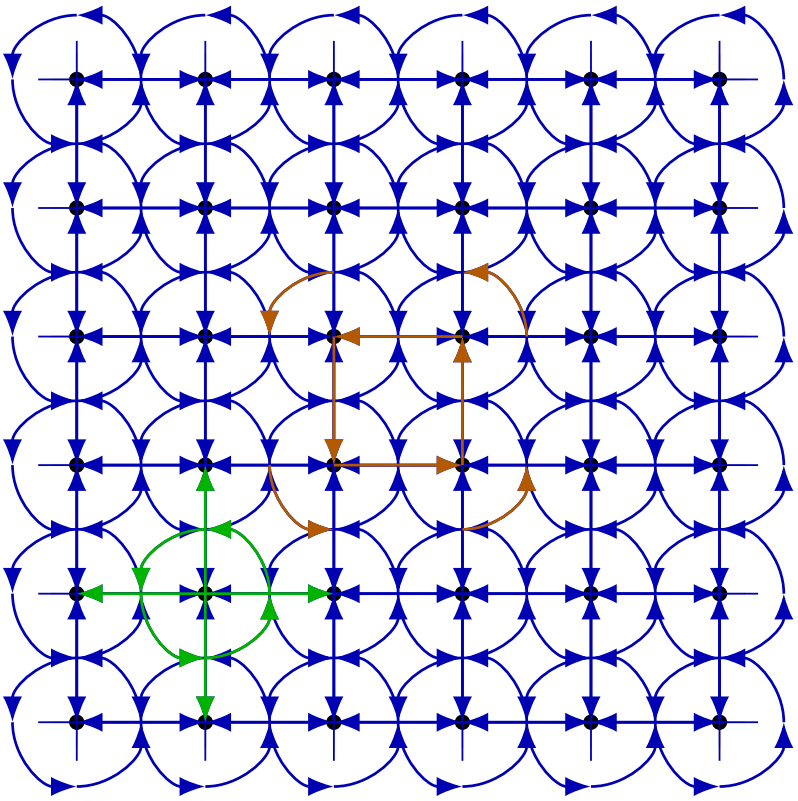

From symmetries to stationarity in Markovian Monte Carlo

Balancing flows on loops

Master's thesis in Engineering mathematics and computational science

VINCENT HARBANDER

DEPARTMENT OF MATHEMATICAL SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

From symmetries to stationarity in Markovian Monte Carlo

Balancing flows on loops

VINCENT HARBANDER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
Gothenburg, Sweden 2026

From symmetries to stationarity in Markovian Monte Carlo
Balancing flows on loops
VINCENT HARBANDER

Supervisor: Moritz Schauer, Department of Mathematical Sciences
Examiner: Moritz Schauer, Department of Mathematical Sciences

Master's Thesis 2026
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Loops of probability flow tiling the base space $\mathbb{Z}^2$. The orange arrows present what a single loop looks like. The green arrows present the possible flows from states with a specific position in the base space. The cover is a scaled version of Figure 3.2, created using the LaTeX package TikZ.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

From symmetries to stationarity in Markovian Monte Carlo
Balancing flows on loops
VINCENT HARBANDER
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Markov chain Monte Carlo (MCMC) algorithms algorithms produce dependent samples of a target distribution. These are used to compute statistics of intractable distributions. A similar theory exists for Markov jump processes called Markovian Monte Carlo (MMC). Research into novel MCMC and MMC algorithms has explored non-reversible alternatives using weaker stationarity conditions than detailed balance. This thesis proves invariance for two new families of kernels belonging to rejection-free, non-reversible MMC algorithms sampling on locally compact Polish groups. The proofs use an unorthodox bottom-up approach. Example implementations of these algorithms are provided in Python.

Keywords: MCMC, MMC, Markov process, Markov chain, Monte Carlo, kernel, group, non-reversible, rejection-free, invariance.

Acknowledgements

First and foremost, I would like to thank my supervisor Moritz Schauer for his enthusiasm, guidance and wisdom. This project would not have been possible without him. He has been very available and ready to answer my messages, even late into the night or during weekends, which I am very appreciative of. I learned something at each one of our weekly meetings. I often entered stuck and less hopeful about my approaches, but exited with increased faith, new ideas and a new perspective on whatever I was working on. Many things Moritz suggested did not make sense to me at first, but I usually found myself arriving at the same conclusions as he did after days of thought. It was commonplace that I would only need to explain what I had done in very little detail because he figured out all the details by himself on the spot. Moritz consistently provided the exact intuition I needed to solve what was ahead of me. His enthusiasm and interest in the subject was contagious. He gave me the freedom to explore what I wanted and provided the right level of guidance to let me accomplish my goals.

I would also like to thank my girlfriend May Ohlsson for her emotional support and for finding the cause of a particularly difficult bug in my code, saving me hours of work.

Large language models, henceforth referred to as the singlar *AI*, were not used for developing the mathematics, writing the original software implementations of the algorithms or writing the thesis. I used AI for fixing simple LaTeX errors and adjusting details of the plotting code, such as aligning the ticks or rotating tick label text for the heatmaps. Moritz used AI when proofreading the thesis and to find a troublesome bug in my code.

Vincent Harbander, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

MCMC	Markov chain Monte Carlo
MMC	Markovian Monte Carlo
PDMP	Piecewise-deterministic Markov process

Notation

Below is the notation of sets, variables, measures, kernels, other functions and general notation that have been used throughout this thesis, listed in order of appearance within categories.

Sets

$\mathbb{R}$	The set of real numbers
$\mathbb{N}$	The set of natural numbers $0, 1, 2, \ldots$
$\mathbb{Z}_+$	The set of positive integers $1, 2, \ldots$
I	Finite index set
$\mathcal{L}^1_\pi(X)$	The space of absolutely integrable functions $X \to \mathbb{R}$ with respect to the measure π
S^1	The circle as a topological space, parametrized by $[-1, 1)$
S^2	The sphere as a topological space
$\emptyset$	The empty set

Variables

X	State space or base space
V	Augmentation space
S	Augmented state space
$x, y, x', x_1, x_2, \ldots$	Element of X
$v, u, v', v'', v_1, v_2, \ldots$	Element of V
$s, s_1, s_2, \ldots$	Element of S
A, B	Measurable set
$(Z_t)_{t\geq 0}$	Continuous-time Markov jump process
$(\tau_n)_{n\in\mathbb{Z}_+}$	Jump times of the Markov process Z_t

τ_∞	The limit of $(\tau_n)_{n\in\mathbb{Z}_+}$
$(Y_n)_{n\in\mathbb{N}}$	Discrete-time Markov chain
$(\xi_n)_{n\in\mathbb{Z}_+}$	Independent exponentially distributed random variables
a_k	Non-negative real number
$\mathcal{X}$	σ-algebra for X
t, t_0	Real numbers representing time
C	Non-negative or positive real number
N	Positive integer
n	Positive integer, usually representing the order of the map R
d	Positive integer representing the dimension of the base space
i, j, k	Index
φ	Non-negative real number representing probability flow
$\hat{\theta}_N$	Estimator for sample size N
z	Real number or element of X
c	Turn correction factor and element of $[0, 1]$
n_k	Positive integer and order of the function R_k
$\mathbf{I}$	Identity matrix
p	Probability
ε	Positive real number
M	Large positive integer

Measures

P	Probability measure
δ_x	The Dirac delta measure with unit mass at x
λ	The left Haar measure
π	Target measure
$\tilde{\pi}$	Reference measure
$\mathcal{U}$	Measure on V
$\breve{\pi}$	The product measure of π and $\mathcal{U}$

Kernels

$\mu, \nu, \iota, \xi, \mu_k, \iota_k$	Kernel
κ_μ	Markov kernel for kernel μ
κ	Markov kernel that is a transition kernel for a time-homogeneous Markov chain
$\tilde{\mu}$	Reference kernel
γ	Velocity refreshment kernel

Functions

id_X	The identity function on X
$\mathbf{1}_A$	Indicator function of a set A
$\mathbb{E}$	The expectation under P
λ_μ	Rate of the kernel μ
t, t_k	Function $X \times V \to \mathbb{R}$
R, R_k	Function $V \to V$
$\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$	The Radon-Nikodym derivative of λ with respect to π
r	The Radon-Nikodym derivative $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ up to a constant
f	Almost always a test function, an indicator of a measurable set
$\lvert \cdot \rvert$	The absolute value function for the real numbers
$\mathfrak{s}$	Involution on the state space
s	Positive function $X \to \mathbb{R}$
f_π	Probability density function for π
h	Positive function $V \to \mathbb{R}$
$\tilde{h}$	The function h up to a constant
a, b	Non-negative function $S \to \mathbb{R}$
W, T	Function $S \to S$
Γ_j	Function $V \to X$

General notation

$X \times V$	The product space of X and V
Exp	The exponential distribution parametrized by its rate
$\mathbb{E}_{Y\sim\pi}[f(Y)]$	The expectation of $f(Y)$ if Y is π distributed, or equivalently f integrated over the measure π

$\pi \ll \tilde{\pi}$	π being absolutely continuous with respect to $\tilde{\pi}$ as measures
$\mathcal{X}\big\vert_A$	The restriction of $\mathcal{X}$ to A
$\pi\big\vert_A$	The restriction of π to A
$\pi \otimes \mathcal{U}$	The product measure of π and $\mathcal{U}$
R^j	The map R composed with itself j times for a non-negative j, defining $R^0 = \mathrm{id}_V$
R^{-1}	The pre-image or inverse of R depending on context
x^{-1}	The inverse of x as a group element
e	The identity element of a group
min	The function producing the minimum element of a finite ordered set
$\hat{\theta}$	Estimator of θ
$\mathcal{O}$	Order of asymptotic algorithmic time complexity
$a \propto b$	The quantity a is proportional to b, meaning they differ only up to a constant multiplicative factor

Contents

1
Introduction

Markov chain Monte Carlo (MCMC) algorithms are used for Bayesian statistics generally, and can therefore be used in many applied fields such as epidemiology [1, p. 383]. They allow for the estimation of expectations of functions of random variables given only the law up to a constant. This generalizes classical Monte Carlo methods, which require independent samples from the distribution of the random variable in question. Markov chain Monte Carlo methods instead use the law to create dependent samples of the distribution through a Markov process. We study a generalization of Markov chain Monte Carlo called Markovian Monte Carlo (MMC) that generalizes the concept to continuous-time pure jump Markov processes.
The reader is expected to have a basic understanding of probability theory, measure theory and stochastic processes. This includes notions such as measures and the Lebesgue integral. To refresh on more advanced measure theoretic topics that will be relevant for the thesis, the reader is instructed to begin by reading Appendix A.

1.1 Problem formulation

The primary objective of this thesis is to create new non-reversible, rejection-free Markovian Monte Carlo in an unorthodox bottom-up approach. The samplers discovered do not require the target measure to be known exactly, only up to a constant, as in the Metropolis-Hastings algorithm [2, p. 167]. The idea behind the bottom-up approach is to combine several invariant sets of flows instead of defining a kernel and proving its invariance in an ad hoc top-down approach. This hopes to simplify the proofs. The non-reversibility increases the efficiency of the samplers by not making the sampler to retread paths just taken. Rejection-freeness is desirable as rejection sampling steps are computationally wasteful.
A secondary objective is implementing the discovered samplers and verifying that they work as expected.

1.2 Limitations

A Markovian Monte Carlo method requires three components: existence of the Markov jump process, stationarity of the process with respect to the target measure and ergodicity of the process, ensuring convergence to the stationary distribution over time. This thesis will present an already known result on existence and primarily focus on proving stationarity. Proofs of ergodicity are not included, but broad strokes on how to prove irreducibility is provided.

1.3 Related work

The formalism, notation, problem formulation and setting of this thesis is inspired by Jansson et al. [3]. Ruben Seyer's licentiate thesis [4] is a well-written and concise introduction to the basic setting of the thesis. Both of the preceding works and this thesis are based on theory from the book *Foundations of Modern Probability* by Olav Kallenberg [5]. Leo Breiman's book *Probability* [6] is a useful resource on ergodic theory and Markov processes in general. The most similar work to this thesis is a part of Sam Power's PhD dissertation where he considers Markov processes on discrete spaces with algebraic structure [7, p. 138–150]. He keeps the setting very general and uses detailed balance and skew-detailed balance theory. This thesis makes more assumptions on the space to be able to construct novel invariant components of probability flow.

1.4 Thesis outline

In Chapter 2, Theory, we cover the necessary prerequisite theory to understand the problem formulation and be able to state the findings. Chapter 3, Theoretical findings, is split up into five sections. The first two sections establish already known theory on constructing kernels and state space augmentation. The third section motivates, defines and proves the invariance of the flow walker kernel. The final section outlines a method for proving ergodicity for such processes and argues for the irreducibility of the processes. In Chapter 4, Method, we cover how to simulate the studied processes, specific implementation details for the implementation provided and construct an error measurement and examples one can use to empirically test convergence. Chapter 5, Results, shows the results of these empirical tests. Chapter 6, Discussion, discusses possible future research based on this. The final Chapter 7 summarizes the methods and findings of the thesis. The appendix contains measure theoretic preliminaries.

2

Theory

Let X be a Polish space, meaning a separable completely metrizable topological space, and let $\mathcal{X}$ be its Borel σ-algebra, meaning it is generated by the topology of the space. Let $\mathbb{N}$ denote the natural numbers (including 0) and let $\mathbb{Z}_+$ denote the positive integers.

2.1 Kernels

The primary object of this thesis is that of the kernel.

Definition 1 (Kernel)**.** *We call a function $\mu : X \times \mathcal{X} \to \mathbb{R}$ a kernel if $\mu(x, \cdot)$ is a finite*[1] *measure for each $x \in X$ and $\mu(\cdot, A)$ is a measurable function*[2] *for each $A \in \mathcal{X}$.*

Definition 2 (Rate of a kernel)**.** *Let*

$$\lambda_\mu(x) = \mu(x, X),$$

be the rate of the kernel μ.

Remark 1. *The rate always exists, since μ is a kernel and thus $\mu(x, \cdot)$ is a finite measure for all $x \in X$.*

Definition 3 (Markov kernel)**.** *A kernel μ is called a Markov kernel if its rate at every point is 1, so $\lambda_\mu(x) = 1$ for all $x \in X$.*

A Markov kernel $\mu(x, A)$ can describe transition probabilities of a time-homogeneous Markov chain from state x to a measurable set A. This generalizes transition matrices for finite state spaces. Thus $\mu(x, \cdot)$ describes the law of the next step, given that the current step is at x.

Lemma 1. *Every kernel μ with $\lambda_\mu(x) > 0$ for all $x \in X$ can be normalized to a Markov kernel by*

$$\kappa_\mu(x, A) = \frac{\mu(x, A)}{\lambda_\mu(x)}.$$

[1] This is a particularly strong restriction we are using in this thesis to make things simpler. Many others, such as Jansson et al. [3] and Seyer [4] don't assume finiteness.

[2] $\mathcal{X}$-$\mathcal{B}(\mathbb{R})$-measurable, that is, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$, the σ-algebra generated by the natural topology on $\mathbb{R}$.

Proof. Let us verify that the rate of κ is 1.

$$\lambda_{\kappa_\mu}(x) = \kappa_\mu(x, X) = \frac{\mu(x, X)}{\lambda_\mu(x)} = \frac{\lambda_\mu(x)}{\lambda_\mu(x)} = 1.$$

□

Remark 2. *Note that the above lemma lets one decompose every kernel with a positive rate everywhere into a rate and a Markov kernel as*

$$\mu(x, A) = \lambda_\mu(x)\kappa_\mu(x, A).$$

2.2 Markov jump processes

The objects of study are Markov jump processes. These are, as defined by Jansson et al. [3, p. 2], continuous-time Markov processes $(Z_t)_{t\geq 0}$ on our state space X with almost surely right-continuous paths that are constant except at isolated jumps. We only consider time-homogeneous processes.
Following Jansson et al. [3, p. 2], one can define the jump times $(\tau_n)_{n\in\mathbb{Z}_+}$ and a Markov kernel $\kappa_\mu(x, A) := \mathrm{P}\left(Z_{\tau_{i+1}} \in A \mid Z_{\tau_i} = x\right)$ for all $A \in \mathcal{X}$ and all $x \in X$. Note that $\tau_{i+1} - \tau_i \mid Z_{\tau_i} = x$ is exponentially distributed because of the Markov property. One can denote the rate of this exponential distribution by $\lambda_\mu(x) := (\mathbb{E}\left[\tau_{i+1} - \tau_i \mid Z_{\tau_i} = x\right])^{-1}$. Jansson et al. further note that $\lambda_\mu(Z_0)\tau_1, \lambda_\mu(Z_{\tau_1})(\tau_2 - \tau_1), \lambda_\mu(Z_{\tau_2})(\tau_3 - \tau_2), \ldots$ form an independent and identically distributed sequence of Exp(1) random variables and that the sequence $Z_0, Z_1, \ldots$ form an induced discrete-time Markov chain. Note that the Markov kernel κ_μ describes transitions of this chain. One can define a rate kernel for the Markov jump process by $\mu(x, A) = \lambda_\mu(x)\kappa_\mu(x, A)$ to describe jump rates. Now one can see that the name "rate" for λ_μ was justified, as it is both a jump rate and the rate of the kernel μ according to Definition 2.
A point $x \in X$ having zero rate means that the process almost surely never leaves the point and is stuck in an absorbing state, which intuitively explains why such a rate kernel cannot be normalized. For this reason, let us, like Seyer, adopt the convention that $\kappa_\mu(x, \mathrm{d}y) = \delta_x(\mathrm{d}y)$ if $\lambda_\mu(x) = 0$, and $\mu(x, \{x\}) = 0$ [4, p. 9]. Kernels will describe the dynamics of Markov jump processes and these conventions ensure that the processes will not jump in place and will stay where they are if they are absorbed. Note that these conventions respect the decomposition $\mu(x, A) = \lambda_\mu(x)\kappa_\mu(x, A)$.

Remark 3. *The trivial kernel $\mu = 0$ describes a Markov jump process that is absorbed in every point. It never moves from where it started and has rate $\lambda_\mu = 0$ everywhere.*

The rate kernel $\mu(x, A)$ describes the rate at which the process leaves x and jumps into A. The rate $\lambda_\mu(x)$ can be thought of as the sum of all these rates over all possible destinations.
Next, let us establish the existence of a Markov jump process with a specified rate kernel.

Theorem 1. *Let μ be a kernel with bounded rate λ_μ. Decompose μ as $\mu(x, \mathrm{d}y) = \lambda_\mu(x)\kappa_\mu(x, \mathrm{d}y)$ and assume that $\mu(x, \{x\}) = 0$. Let $(Y_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel κ_μ and let $(\xi_n)_{n\in\mathbb{Z}_+}$ be independent Exp(1)-distributed random variables. Then there exists a Markov jump process $(Z_t)_{t\geq 0}$ on our measurable space $(X, \mathcal{X})$ with rate kernel μ if and only if*

$$\sum_{n=1}^{\infty} \frac{\xi_n}{\lambda_\mu(Y_{n-1})} = \infty,$$

almost surely for every initial distribution Y_0 [4, p. 9][3].

This condition ensures that a Markov jump process with our jump kernel will exist as long as it has bounded rate, which it always has according to our kernel definition, and does not explode, meaning it doesn't make infinitely many jumps in finite time. It is simple to see that if we, for example, have $\tau_{n+1} - \tau_n = 1/2^n$ we get that $\tau_n \to \tau_\infty < \infty$. Where would the process be at times greater than τ_∞? It is clear that some ambiguity can arise here, which justifies the explosion condition.
Let us elaborate on the connection between explosion and the explosion condition above. Since $\xi_n \sim \mathrm{Exp}(1)$, we have that

$$\frac{\xi_n}{\lambda_\mu(Y_{n-1})} \sim \mathrm{Exp}(\lambda_\mu(Y_{n-1})).$$

This matches the distribution of $\tau_n - \tau_{n-1}$. So conditioning on the path (Y_n),

$$\sum_{n=1}^{N} \frac{\xi_n}{\lambda_\mu(Y_{n-1})}$$

is equal in distribution to τ_N, which are the partial sums of the left-hand side of Equation 1.

2.3 Stationarity and invariance

Definition 4 (Invariance)**.** *A measure π on $(X, \mathcal{X})$ is called invariant to a kernel μ if*

$$\int \mu(x, A)\pi(\mathrm{d}x) = \int_A \mu(x, X)\pi(\mathrm{d}x),$$

for all measurable sets A where the integrals above are finite.

If π is a probability measure, one can interpret $\mu(x, \mathrm{d}y)\,\pi(\mathrm{d}x)$ as the probability flow from a small region $\mathrm{d}x$ to a small region $\mathrm{d}y$. To arrive at this interpretation, one should think of the amount of probability flow being equal to the amount of probability of being in that region, π, times the rate at which it is transported per amount of probability, μ. Then the above can be seen as a flow condition. The flow into A should equal the flow out of A for π to be invariant.

[3] Seyer's licentiate thesis is cited here to credit the stylistic aspects in writing the theorem. Seyer himself cites Kallenberg [5, p. 280] for the theorem.

Lemma 2 (Invariance condition)**.** *A measure π is invariant to a kernel μ if and only if*

$$\int\int (f(y) - f(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) = 0,$$

for all indicator functions f for all sets from Definition 4.

Proof. We begin by proving that invariance yields the invariance condition. Rewriting invariance in terms of indicator functions, we get

$$\begin{aligned}\int_{x\in X}\int_{y\in X} \mathbf{1}_A(y)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) &= \int_{x\in X}\int_{y\in X} \mathbf{1}_A(x)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) \\ &\overset{\text{Fubini by finiteness and positivity}}{=} \int_{y\in X}\int_{x\in X} \mathbf{1}_A(x)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x),\end{aligned}$$

for all such A. Next, we subtract the right-most integral from the left one so

$$\begin{aligned}&\int_{x\in X}\int_{y\in X} \mathbf{1}_A(y)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) - \int_{y\in X}\int_{x\in X} \mathbf{1}_A(x)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) \\ &\qquad = \int_{x\in X}\int_{y\in X} (\mathbf{1}_A(y) - \mathbf{1}_A(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) = 0.\end{aligned}$$

This holds for all such sets A, so the invariance condition is proven.
Next assume the invariane condition and let us prove invariance. We know that

$$\int_{x\in X}\int_{y\in X} (\mathbf{1}_A(y) - \mathbf{1}_A(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) = 0,$$

for all such sets A. By assumption, the integrals

$$\int_{x\in X}\int_{y\in X} \mathbf{1}_A(y)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) \qquad \text{and} \qquad \int_{x\in X}\int_{y\in X} \mathbf{1}_A(x)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x),$$

are finite. Then we can add the second integral to both sides and use additivity.

$$\int_{x\in X}\int_{y\in X} \mathbf{1}_A(y)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) = \int_{x\in X}\int_{y\in X} \mathbf{1}_A(x)\mu(x, \mathrm{d}y)\pi(\mathrm{d}x)$$

Incorporating the indicators into the domains and then integrating out y we get

$$\int \mu(x, A)\pi(\mathrm{d}x) = \int_A \mu(x, X)\pi(\mathrm{d}x).$$

This completes the proof. □

The indicators should be seen as test functions and will henceforth be referred to as such. The difference of test functions should be interpreted as the flow from and to that set.

Definition 5 (Stationarity)**.** *A probability measure π is called stationary with respect to a Markov jump process $(Z_t)_{t\geq 0}$ if*

$$\int_{x\in X} \mathrm{P}\left(Z_t \in A \mid Z_{t_0} = x\right)\pi(\mathrm{d}x) = \pi(A), \quad \textit{for all } A \in \mathcal{X} \textit{ and } 0 < t_0 < t \ \ [3, p.\ 3].$$

This should be interpreted as if you are starting π-distributed, you stay π-distributed over time according to the Markov jump process' dynamics. Note how similar this is to the flow interpretation of invariance. Stationarity and invariance turn out to be directly connected, so we can describe Markov jump processes with kernels. This is characterized in the following results.

Lemma 3. *Let $(Y_n)_{n\in\mathbb{N}}$ be a time-homogeneous Markov chain on $(X, \mathcal{X})$ with transition kernel κ, meaning $\kappa(x, A) = P(Y_{n+1} \in A \mid Y_n = x)$. If the initial distribution is π, then (Y_n) is stationary with respect to π if and only if π is invariant for κ* [4, p. 8].

Proposition 1. *Consider the setting of Theorem 1. Let π be a probability measure such that $\int \lambda_\mu(x)\pi(\mathrm{d}x) < \infty$. Then the Markov jump process $(Z_t)_{t\geq 0}$ is stationary with respect to π if and only if the Markov chain $(Y_n)_{n\in\mathbb{N}}$ is stationary with respect to $\lambda_\mu(x)\pi(\mathrm{d}x)$. Additionally, π is invariant for the rate kernel μ if and only if $\lambda_\mu(x)\pi(\mathrm{d}x)$ is invariant for the Markov kernel κ_μ* [4, p. 9].

Remark 4. *Combining the above results gives us an equivalence between stationarity and invariance under sufficient conditions. This has reduced the construction of our Markov jump processes to analysis of kernels.*

2.4 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) algorithms like the Metropolis-Hastings algorithm have been used to obtain expectations of functions of random variables in situations where classical methods are intractable. Classical Monte Carlo methods require independent samples, $Y_1, Y_2, \ldots$, of our target distribution π and estimates it as

$$\mathbb{E}_{Y\sim\pi}[f(Y)] \approx \frac{1}{N}\sum_{j=1}^{N} f(Y_j).$$

This works because the strong law of large numbers ensures almost sure convergence as $N \to \infty$. This whole scheme fails if one can't produce independent samples of the target distribution.
MCMC algorithms avoid this by constructing an ergodic Markov chain $Y_1, Y_2, \ldots$ with the target distribution π as a stationary distribution. Then the chain will converge to π regardless of initial distribution and one can apply the strong law of large numbers for Markov chains, which states that for $f \in \mathcal{L}^1_\pi(X)$ one has

$$\frac{1}{N}\sum_{j=1}^{N} f(Z_j) \xrightarrow[N\longrightarrow\infty]{} \mathbb{E}_{Z\sim\pi}[f(Z)] \ \ [9, p.\ 416],$$

almost surely. This motivates the usage of the estimator

$$\hat{\theta}_N = \frac{1}{N}\sum_{j=1}^{N} f(Z_j),$$

for large N.
We will use Markovian Monte Carlo, an analogous framework for Markov jump processes in continuous-time. An almost identical theorem to the one above exists for Markov jump processes once the samples are weighted by the wait times in each state. To reduce the variance of the estimator, we will replace the wait times with their expectation. This is captured by the following theorem from Jansson et al.

Theorem 2. *If the Markov jump process Z corresponding to the rate kernel μ is ergodic, has invariant measure π and positive rate λ_μ, it holds that for any π-integrable function f that*

$$\mathbb{E}_{\tilde{Z}\sim\pi}\left[f(\tilde{Z})\right] = \lim_{N\to\infty}\left(\sum_{n=0}^{N}\frac{f(Z_{\tau_n})}{\lambda_\mu(Z_{\tau_n})}\right)\Big/\left(\sum_{n=0}^{N}\frac{1}{\lambda_\mu(Z_{\tau_n})}\right),$$

almost surely, for almost every Z_0, where $\tau_0 = 0$ [3, p. 5].

Ergodicity is important to establish convergence to the stationary distribution. Stationarity does not suffice for an ergodic theorem. A Markov jump process may have many different stationary distributions. Consider the following, for instance.

Proposition 2. *Let X be our state space and $\mathfrak{s} : X \to X$ be an involution, meaning $\mathfrak{s}\circ\mathfrak{s} = \mathrm{id}_X$, that is π-invariant for a finite measure π, meaning*[4] *$\pi\circ\mathfrak{s}^{-1} = \pi$. Then π is invariant to the rate kernel $\mu(x,\mathrm{d}y) = C\delta_{\mathfrak{s}(x)}(\mathrm{d}y)$ for all $C \geq 0$.*

Proof. Let us show the invariance condition.

$$\begin{aligned}
&\int\int(f(y)-f(x))\mu(x,\mathrm{d}y)\pi(\mathrm{d}x) = \int\int(f(y)-f(x))C\delta_{\mathfrak{s}(x)}(\mathrm{d}y)\pi(\mathrm{d}x)\\
&= C\int(f(\mathfrak{s}(x))-f(x))\pi(\mathrm{d}x)\\
&= C\int f(x)\left(\pi\circ\mathfrak{s}^{-1}\right)(\mathrm{d}x) - C\int f(x)\pi(\mathrm{d}x)\\
&= C\int f(x)\pi(\mathrm{d}x) - C\int f(x)\pi(\mathrm{d}x) = 0.
\end{aligned}$$

□

The proposition above lets one construct a Markov jump process that is invariant for a very large class of target measures using Theorem 1. The construction of ergodic Markov jump processes with a specified limiting distribution does motivate our pursuit, but does not provide a complete problem formulation as the following proposition demonstrates.

Proposition 3 (Importance sampling for Markov jump processes)**.** *Suppose a measure $\tilde{\pi}$ is invariant with respect to a kernel $\tilde{\mu}$ on a state space X. Suppose that one instead desires a kernel μ that leaves a measure π invariant. Assume that $\tilde{\pi}$ is absolutely continuous, meaning $\tilde{\pi} \ll \pi$. Then*

$$\mu(x,\mathrm{d}y) = C\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(x)\tilde{\mu}(x,\mathrm{d}y),$$

[4] One should interpret $\mathfrak{s}^{-1}$ as the pre-image.

suffices for any constant $C \geq 0$, where $\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}$ is a Radon-Nikodym derivative of $\tilde{\pi}$ with respect to π.

Proof. Let us verify the invariance using the invariance condition. The integral can easily be transformed to the invariance condition of $\tilde{\pi}$ with respect to $\tilde{\mu}$ using the Radon-Nikodym property.

$$\begin{aligned}
&\int\int (f(y) - f(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) \\
&= \int\int (f(y) - f(x))C\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(x)\tilde{\mu}(x, \mathrm{d}y)\pi(\mathrm{d}x) \\
&= C\int\int (f(y) - f(x))\tilde{\mu}(x, \mathrm{d}y)\tilde{\pi}(\mathrm{d}x) \\
&= C \cdot 0 = 0
\end{aligned}$$

□

This proposition lets one take an ergodic Markov jump process and simply change the wait times to change the limiting distribution almost arbitrarily. This can be seen as a form of importance sampling to reweight the induced Markov chain's walk or alternatively as speeding up and slowing down the wait times of the Markov jump process. This means it is not innovative to find just any ergodic Markov process with the correct limiting distribution. What we'd like is for a process to converge quickly to the limiting distribution while being computationally inexpensive to simulate. Methods such as Metropolis-Hastings have many flaws.

- **Rejections.** It selects where to go by performing random rejections, which requires more computation steps.
- **Choice of proposal distribution.** Metropolis-Hastings requires a proposal distribution to navigate the space. This proposal distribution impacts the probability of rejecting steps and its efficiency may depend a lot on the target measure.
- **Reversibility.** The process is reversible. Intuitively, this means that the probability flow from x to y is always equal to the probability flow from y to x. This is a weakness since it very easily leads to poor mixing.

Ideally, we'd like our process to quickly explore much of the target. This means we want to avoid retreading where it has been. Reversibility goes against this as it can cause random walk behavior. If one has a process on the state space $\mathbb{R}$ that walks by adding a standard independent Gaussian to its position, it will spread out and explore quite slowly. Ideally we would like to base our MCMC algorithm on a non-reversible process like a random walk on a circle that always walks in one direction according to some distribution. Such a process would quickly walk around the entire circle.
Formally, a kernel is called reversible if it fulfills the detailed balance condition.

Definition 6 (Detailed balance condition)**.** *A kernel μ and measure π are said to fulfill the detailed balance condition if*

$$\mu\left(x, \mathrm{d}y\right)\pi\left(\mathrm{d}x\right) = \mu\left(y, \mathrm{d}x\right)\pi\left(\mathrm{d}y\right).$$

Proposition 4 (Detailed balance implies invariance)**.** *Let μ and π be in detailed balance and let π be a finite measure. Then π is invariant with respect to μ.*

Proof. Observe the following.

$$\begin{aligned}
&\int_x \int_y (f(y) - f(x))\mu\left(x, \mathrm{d}y\right) \pi\left(\mathrm{d}x\right) \\
&\overset{\text{detailed balance}}{=} \int_x \int_y (f(y) - f(x))\mu\left(y, \mathrm{d}x\right) \pi\left(\mathrm{d}y\right) \\
&\overset{\text{Fubini by finiteness}}{=} \int_y \int_x (f(y) - f(x))\mu\left(y, \mathrm{d}x\right) \pi\left(\mathrm{d}y\right) \\
&\overset{\text{relabel}}{=} \int_x \int_y (f(x) - f(y))\mu\left(x, \mathrm{d}y\right) \pi\left(\mathrm{d}x\right) \\
&= -\int_x \int_y (f(y) - f(x))\mu\left(x, \mathrm{d}y\right) \pi\left(\mathrm{d}x\right)
\end{aligned}$$

This implies the integral is 0 since 0 is the only solution to the equation $z = -z$. This proves invariance through the invariance condition. □

Remark 5. *Such a kernel may be a Markov kernel and represent a time-homogeneous, discrete-time Markov chain, which is used by Metropolis-Hastings.*

Remark 6. *Intuitively, it is obvious that it must be invariant. The total flow into a point is equal to the total flow out of the point if the flow between any pair of points is equal.*

Remark 7. *Detailed balance is not equivalent to invariance. A simple counterexample consists of constructing a triangle of unidirectional flow, where there is probability flow $x \to y \to z \to x$, but no flow in reverse. This idea of forming loops of flow is key to our later results.*

We will construct non-reversible, rejection-free MMC algorithms. The Markovian Monte Carlo framework will allow an additional degree of freedom as we can design not only where processes move, but also how long they spend on average before moving.

3

Theoretical findings

This chapter will build up to proving that one can construct specific types of kernels that leave any target measure of a very large class invariant. Recalling Remark 4, one can see that this leads to Markovian Monte Carlo samplers, under sufficient conditions.

3.1 Kernel construction

Let us call our measurable space $(X, \mathcal{X})$ and suppose we have a finite measure π on X. Suppose we have an invariant flow within a measurable subset $A \subset X$ to the measure $\pi\big|_A$, which is π restricted to A. Let us represent the rate of this flow as a rate kernel $\tilde{\mu}$ on the measurable space $\left(A, \mathcal{X}\big|_A\right)$, where $\mathcal{X}\big|_A$ denotes the restriction of $\mathcal{X}$ to A, meaning $\{B \cap A \;\big|\; B \in \mathcal{X}\}$. This can then be extended to a kernel on the entire space that is invariant with respect to π in a natural way.

$$\mu(x, B) = \begin{cases} \tilde{\mu}(x, B \cap A) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Proposition 5 (Kernel extension)**.** *Kernels $\tilde{\mu}$ invariant for $\pi\big|_A$ on a measurable subspace $\left(A, \mathcal{X}\big|_A\right)$ can be extended to the entire space $(X, \mathcal{X})$ such that π is invariant with respect to the extension according to the construction above.*

Proof. Since we know that $\pi\big|_A$ is invariant with respect to $\tilde{\mu}$, we can use that

$$\int_A \int_A (f(y) - f(x)) \tilde{\mu}\,(x, \mathrm{d}y)\, \pi\big|_A(\mathrm{d}x) = 0,$$

according to the invariance condition. Likewise, the invariance condition tells us it is sufficient to prove

$$\int_X \int_X (f(y) - f(x)) \mu\,(x, \mathrm{d}y)\, \pi(\mathrm{d}x) = 0,$$

to obtain invariance. The computation is straight-forward.

$$
\begin{aligned}
&\int_{x\in X}\int_{y\in X}(f(y)-f(x))\mu\left(x,\mathrm{d}y\right)\pi(\mathrm{d}x)\\
&=\int_{x\in A}\int_{y\in X}(f(y)-f(x))\mu\left(x,\mathrm{d}y\right)\pi(\mathrm{d}x)+\int_{x\in X\setminus A}\int_{y\in X}(f(y)-f(x))\mu\left(x,\mathrm{d}y\right)\pi(\mathrm{d}x)\\
&\overset{\mu\text{ def.}}{=}\int_{x\in A}\int_{y\in X}(f(y)-f(x))\mathbf{1}_A(y)\tilde{\mu}\left(x,\mathrm{d}y\right)\pi(\mathrm{d}x)+\int_{x\in X\setminus A}\int_{y\in X}(f(y)-f(x))\ 0\ \pi(\mathrm{d}x)\\
&=\int_{x\in A}\int_{y\in A}(f(y)-f(x))\tilde{\mu}\left(x,\mathrm{d}y\right)\pi(\mathrm{d}x)\\
&=\int_{x\in A}\int_{y\in A}(f(y)-f(x))\tilde{\mu}\left(x,\mathrm{d}y\right)\pi\big|_A(\mathrm{d}x)\\
&\overset{\text{Equation 3.1}}{=}0
\end{aligned}
$$

□

This construction puts meaning to studying invariant flows on measurable subsets of the space, without looking at the whole space. Let us define such invariant subsets.

Definition 7 (Invariant component)**.** *If $\pi\big|_A$ is invariant to a kernel on a measurable subspace $(A,\mathcal{X}|_A)$ of $(X,\mathcal{X})$, we call A an invariant component of π with respect to the kernel.*

We may have several of these invariant components and want to combine them. If we are invariant on A with one kernel and invariant on B with another, we can "glue" them together, as long as they are disjoint. If they are not disjoint, we can still preserve invariance by adding their flows together. This is characterized in the following result.

Theorem 3 (Kernel superposition)**.** *Suppose a measure π is invariant under some finite number of kernels, $(\mu_k)_{k\in I}$. Then π is invariant under the kernel*

$$\mu(x,\mathrm{d}y)=\sum_{k\in I}a_k\mu_k(x,\mathrm{d}y),$$

for all non-negative sequences of numbers $(a_k)_{k\in I}$.

Proof. That μ is a kernel is immediately obvious, so let us prove invariance. Since π is invariant under μ_k, we know that

$$\int\int(f(y)-f(x))\mu_k(x,\mathrm{d}y)\pi(\mathrm{d}x)=0,$$

for all $k\in I$ and test functions f. It suffices to show that

$$\int\int(f(y)-f(x))\mu(x,\mathrm{d}y)\pi(\mathrm{d}x)=0,$$

for all test functions f. The result follows immediately.

$$\begin{aligned}0 &= \sum_{k\in I} a_k \int\int (f(y) - f(x))\mu_k(x, \mathrm{d}y)\pi(\mathrm{d}x) \\ &= \int\int (f(y) - f(x)) \left(\sum_{k\in I} a_k \mu_k(x, \mathrm{d}y)\right) \pi(\mathrm{d}x) \\ &= \int\int (f(y) - f(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x)\end{aligned}$$

□

Additionally, there exists a restricted subtraction of kernels. We restrict the subtraction to ensure that the resulting object is a kernel. The property threatened is the non-negativity of the measure in the second argument of the kernel.

Theorem 4 (Kernel subtraction)**.** *Let π be a measure invariant to the kernels μ and ν. Further let $\mu(x, \mathrm{d}y) \geq \nu(x, \mathrm{d}y)$ for all x. Then*

$$\iota(x, \mathrm{d}y) = \mu(x, \mathrm{d}y) - \nu(x, \mathrm{d}y),$$

is a kernel and π is invariant with respect to it. We denote ι by $\mu - \nu$.

Proof. The only non-trivial part about showing that ι is a kernel is showing that $\iota(x, \cdot)$ is a measure, specifically its non-negativity. This follows directly from the assumption that $\mu(x, \mathrm{d}y) \geq \nu(x, \mathrm{d}y)$. Let us prove the invariance next. Once again, we resort to our integral condition for invariance.

$$\begin{aligned}0 &= \int\int (f(y) - f(x))\mu(x, \mathrm{d}y)\pi(\mathrm{d}x) - \int\int (f(y) - f(x))\nu(x, \mathrm{d}y)\pi(\mathrm{d}x) \\ &= \int\int (f(y) - f(x))(\mu(x, \mathrm{d}y) - \nu(x, \mathrm{d}y))\pi(\mathrm{d}x) \\ &= \int\int (f(y) - f(x))\iota(x, \mathrm{d}y)\pi(\mathrm{d}x)\end{aligned}$$

□

3.2 State space augmentation

Recall that our state space X is a Polish space. Since the product of two Polish spaces is Polish [8, p. 378], we can create a new state space S as a product space $S = X \times V$ for some other Polish space V.
If we have a finite target measure π on X, we may create a finite measure $\mathcal{U}$ on V and define the product measure $\breve{\pi} = \pi \otimes \mathcal{U}$ on S. We call this measure the augmented measure. In this context, we call X the *base space*, where the target measure lives, V the *augmentation space* and S the *augmented space*. We may also refer to S simply as the state space. We then want to construct a kernel μ on S such that $\breve{\pi}$ is invariant with respect to it. The idea behind this is that one can have a Markov jump process on S and project dependent samples $(x_1, v_1), (x_2, v_2), \ldots$ from $\breve{\pi}$ down on the first coordinate to obtain dependent samples $x_1, x_2, \ldots$ from the target measure π, since it is a marginal measure.

In particular, we will be interested in the case where $V \subseteq X$ and V is $\mathcal{X}$-measurable and Polish. The elements of V should be thought of as "velocity vectors" in the two samplers covered. This intuition comes from $X = \mathbb{R}^d$. The process can either walk according to the velocity vector or change it. This is intended to provide a richer dynamic in the augmented space, similar to Hamiltonian Monte Carlo. Just like Hamiltonian Monte Carlo, we have our own kind of "momentum refreshment" step, represented by the following definition.

Definition 8 (Velocity refreshment kernel)**.** *Let S and $\breve{\pi}$ be defined as above. Then we call*

$$\mu((x,v), \mathrm{d}(x',v')) = s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v'), \tag{3.1}$$

a velocity refreshment kernel for some positive, sufficiently integrable function $s : X \to \mathbb{R}$.

Lemma 4. *A velocity refreshment kernel leaves its augmented measure invariant.*

Proof. Let us use the invariance condition to show the following integral is 0.

$$\begin{aligned}
&\int_{a\in S}\int_{b\in S}(f(b)-f(a))\mu\left(a,\mathrm{d}b\right)\breve{\pi}\left(\mathrm{d}a\right)\\
=&\int_{(x,v)\in S}\int_{(x',v')\in S}(f(x',v')-f(x,v))\mu\left((x,v),\mathrm{d}(x',v')\right)\breve{\pi}\left(\mathrm{d}(x,v)\right)\\
\overset{\text{def. } 8}{=}&\int_{(x,v)\in S}\int_{(x',v')\in S}(f(x',v')-f(x,v))s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v')\breve{\pi}\left(\mathrm{d}(x,v)\right)\\
\overset{\breve{\pi}=\pi\otimes\mathcal{U}}{=}&\int_{(x,v)\in S}\int_{(x',v')\in S}(f(x',v')-f(x,v))s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v')\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
=&\int_{(x.v)\in S}\int_{v'\in V}(f(x,v')-f(x,v))s(x)\mathcal{U}(\mathrm{d}v')\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)
\end{aligned}$$

The restriction of the test functions that comes from the sets in Definition 4 allows one to split the integral into a difference of two positive integrals. One can then apply Fubini separately on them and recombine them. Thus we can make v and v' change roles, proving the integral is equal to its own negative and is therefore 0. □

Remark 8. *The velocity refreshment kernel trivially does not produce ergodic Markov jump processes on its own as long as the base space has more than one point since it never moves in the base space.*

Remark 9. *This velocity refreshment kernel is a part of both of the families of samplers presented in the following sections.*

A velocity refreshment kernel can be added to a kernel to ensure it does not get stuck on specific subsets of the space, while preserving stationarity. Hamiltonian Monte Carlo preserves the Hamiltonian, so it can never traverse to a state that has a different Hamiltonian than where it started, in theory[1]. Therefore it has a momentum refreshment step to randomize the momentum vector and get a new Hamiltonian. The velocity refreshment can be thought of as analogous to this.

[1]In practice, Hamilton's equations are solved numerically and numerical errors can lead to the Hamiltonian not being fully preserved.

3.3 Flow walker samplers

Consider a discrete base space $X = \mathbb{Z}^d$ for some $d \geq 1$ to begin with. One can imagine a line of equal probability flow $\varphi \geq 0$.

$$\cdots \longrightarrow x - v \longrightarrow x \longrightarrow x + v \longrightarrow x + 2v \longrightarrow \ldots$$

Let us informally call this type of invariant component a *flow line*.

Remark 10. *If the rate to walk from x to the next step is equal to $\varphi/\pi(\{x\})$, then this is invariant to $\check{\pi} = \pi \otimes \mathcal{U}$ for any positive finite measure π and finite measure $\mathcal{U}$.*

Proof. A kernel to describe this would be

$$\mu\left((x, v), \mathrm{d}(x', v')\right) = \frac{1}{\pi(\{x\})} \delta_{(x+v,v)}\left(\mathrm{d}(x', v')\right).$$

Let us proceed with an informal proof as it may bring more insight into what we are doing. If x only walks to $x + v$, then one can only walk to x from $x - v$. For invariance we require that the flow in matches the flow out. The flow in would be $\frac{\varphi}{\pi(\{x-v\})}\pi(\{x - v\}) = \varphi$. The flow out is $\frac{\varphi}{\pi(\{x\})}\pi(\{x\}) = \varphi$. The velocity is invariant since it never changes. Thus we can imagine this flow is invariant. □

Remark 11. *The resulting jump process from walking along a flow line will eventually only accelerate as it walks along the flow line since $\pi(\{x + nv\}) \to 0$ as $n \to \infty$. This is because π is a finite measure. This is not true for flow lines in other spaces, like $\mathbb{Z}_m$, where the flow line will form a loop as there are only finitely many states to visit.*

This kernel will clearly not produce an ergodic jump process as the velocity never changes. Additionally, the process will simply run away to infinity along $x_0 + nv$ after the nth jump when starting at x_0. The key is to combine this with the velocity refreshment kernel, which does the exact opposite, changing the velocity, but leaving the position unchanged.
To develop the flow walker sampler, we generalize our setting from a discrete base space. Assume that our base space X is a locally compact topological group for the topology of the Polish space, also called a locally compact Polish group. Analogizing the flow lines, we turn the addition into the group multiplication.
Since X is a locally compact topological group, a regular left Haar measure λ on X exists. Assume that $\pi \ll \lambda$ and $\lambda \ll \pi$. Further assume λ and π are σ-finite[2]. This ensures that a Radon-Nikodym derivative $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ exists and is positive π-almost everywhere.

Remark 12. *Letting $X = \mathbb{R}^d$ and the group structure be vector addition, a regular left (and right) Haar measure is the Lebesgue measure. Since the probability density*

[2]Above, in Remark 10 and Remark 11 we assumed that π is a finite measure, which immediately yields σ-finiteness. This is not used as we will drop the finiteness assumption in the coming Theorem 5.

function f_π can be defined as

$$f_\pi := \frac{\mathrm{d}\pi}{\mathrm{d}\lambda},$$

we have

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\pi} = \frac{1}{f_\pi}, \quad \pi\text{-almost everywhere.}$$

Since the inverse of the density is the continuous analogue to $1/\pi(\{x\})$, we arrive at how to generalize our flow line kernel.

Theorem 5 (Invariance of a flow line kernel)**.** *Let X be a locally compact Polish group. Let $V \subseteq X$ be measurable*[3] *and Polish. Let $S = X \times V$, π be a σ-finite measure on X and $\mathcal{U}$ be a measure on V. Further let λ be a left Haar measure on X. Then $\check{\pi} = \pi \otimes \mathcal{U}$ is invariant to*

$$\mu\left((x,v), \mathrm{d}(x',v')\right) = h(x,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(vx,v)}\left(\mathrm{d}(x',v')\right), \tag{3.2}$$

for a positive,[4] *sufficiently integrable function $h : X \times V \to \mathbb{R}$ that fulfills $h(x,v) = h(vx,v)$ $\lambda \otimes \mathcal{U}$-almost everywhere.*

Proof. We will use the invariance condition, so we need to show that the following integral is 0.

$$\begin{aligned}
&\int\int (f(x',v') - f(x,v))\mu\left((x,v), \mathrm{d}(x',v')\right)\check{\pi}\left(\mathrm{d}(x,v)\right) \\
&= \int\int (f(x',v') - f(x,v))h(x,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(vx,v)}\left(\mathrm{d}(x',v')\right)\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&\overset{\text{Radon-Nikodym property}}{=} \int\int (f(x',v') - f(x,v))h(x,v)\delta_{(vx,v)}\left(\mathrm{d}(x',v')\right)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \int (f(vx,v) - f(x,v))h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)
\end{aligned}$$

This is 0 because of the translation invariance of the left Haar measure. One can

[3] This thus induces a measurable subspace $(V, \mathcal{X}\big|_V)$.

[4] Positive λ-almost everywhere is required for the resulting sampler to have any chance of being ergodic after being combined with a velocity refreshment kernel. Otherwise there is a measurable set of positive measure that absorbs.

split the integral by assumption of the sets in Definition 4, so we can directly do

$$
\begin{aligned}
&\int (f(vx,v) - f(x,v))h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \int f(vx,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) - \int f(x,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \left[\begin{array}{r} y = vx \implies x = v^{-1}y \\ \text{Measurable bijection. } \lambda \text{ is left Haar, so it's left-translation invariant.} \end{array}\right] \\
&= \int f(y,v)h\left(v^{-1}y, v\right)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}v) - \int f(x,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&\overset{h\left(v^{-1}y,v\right) = h\,(y,v)}{=} \int f(y,v)h(y,v)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}v) - \int f(x,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&\overset{\text{relabel}}{=} \int f(x,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) - \int f(x,v)h(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= 0.
\end{aligned}
$$

□

With this, we can assemble the flow walker kernel by adding the kernels from Equations (3.1) and (3.2).

Corollary 1 (Invariance of the flow walker kernel)**.** *Under the assumptions of Theorem 5, one can add a velocity refreshment kernel to a flow line kernel, meaning a kernel of the form*

$$\mu\left((x,v), \mathrm{d}(x',v')\right) = h(x,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(vx,v)}\left(\mathrm{d}(x',v')\right) + s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v'),$$

leaves $\check{\pi}$ invariant.

Proof. This follows directly from kernel superposition, Theorem 3, since $\check{\pi}$ is invariant to both of the kernels separately. □

The kernel above represents the first sampler developed, the "flow walker". It has this name because it walks according to flow lines before it randomly picks a new direction, and thus a new flow line, to walk in. A possible problem with this sampler can appear in practice on non-compact spaces. If for example $X = V = \mathbb{R}$, we have $\pi \to 0$ as $|x| \to \infty$, so the flow line eventually accelerates as it goes to infinity, given that it does not turn around. Thus, to save ergodicity, some conditions must likely be imposed on s. One natural choice is $s = C\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ for some positive constant C since that would preserve the probability of performing a velocity refreshment at each jump time, across all points $x \in X$, for some fixed $v \in V$. One should also be concerned with the sampler's tendency to jump faster and faster as it goes further out as this may create numerical issues with floating point numbers. For $X = \mathbb{R}^d$, a Radon-Nikodym derivative $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ is the inverse of the probability density of π. Very far out, this will invert a very small number. The density may even be rounded to 0 as the true density is too small to fit in the floating point representation. This can create division by zero errors. It is therefore recommended that this sampler should only be used on compact spaces unless ergodicity can be definitively proven. In the later Section 3.5, I will mention a type of argument one can use to prove ergodicity of samplers like this under appropriate assumptions.

Remark 13. *One only needs to know* $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ *up to a constant since the unknown constant can be baked into* h*.*

If one has a function h and a Radon-Nikodym derivative up to the unknown constant $C > 0$ along the lines of $r(x) = C\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$, one can let $\tilde{h}(x, v) = Ch(x, v)$ and apply the previous corollary to $\tilde{h}$ as the h since $h(x, v)r(x) = \tilde{h}(x, v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$. This constant cannot harm the integrability condition of h. Being able to use the sampler despite only knowing the density up to a constant is a desirable property of Metropolis-Hastings that this sampler thus also exhibits [2, p. 167].

The function h describes the probability flow from (x, v) to (vx, v). This is apparent if one considers that $h(x, v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$ is the walk rate and $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$ converts flows to rates since it's the continuous analog to $1/\pi(\{x\})$. This trick will be repeated in the following section. We use flows against a left Haar measure instead of rates. We can then exploit the left-translate invariance of our left Haar measure.

One could even look at it the other way. We are constructing a process invariant to our left Haar measure and then reweighting it using the importannce sampling, Proposition 3, to be invariant to the target instead. Then our rates are proportional to the flows.

3.4 Loop samplers

Let us again consider some discrete base space X for simplicity. As inspiration for our next sampler, let us study another kind of invariant component. Choose some integer $N \geq 2$ and consider flow between states $s_j \in S$ such that

$$s_N \longrightarrow s_1 \longrightarrow s_2 \longrightarrow \cdots \longrightarrow s_j \longrightarrow s_{j+1} \longrightarrow \cdots \longrightarrow s_{N-1} \longrightarrow s_N,$$

which forms a closed loop of probability flow $\varphi \geq 0$. We call these flow loops. Let us specifically consider $X = \mathbb{Z}^2$ under the discrete topology and $V = \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$. Suppose V represents a velocity again, where the process on the augmented space S either takes a step in X according to the direction in V whilst leaving V unchanged or it changes V whilst leaving X unchanged. Suppose the V step involves a counter-clockwise rotation by 90 degrees, meaning it applies the map $R : V \to V$ that maps $(x, y) \mapsto (-y, x)$ to the V entry. This would correspond to a kernel of the form

$$\mu\left((x, v), \mathrm{d}(x', v')\right) = a(x, v)\delta_{(x+v,v)}\left(\mathrm{d}(x', v')\right) + b(x, v)\delta_{(x,Rv)}\left(\mathrm{d}(x', v')\right).$$

A flow loop can appear as alternating steps of walking and turning, as illustrated in Figure 3.1.

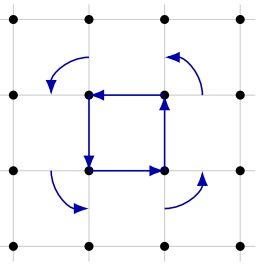

Figure 3.1: Illustration of steps one can take in $\mathbb{Z}^2$ to end up back where one started, facing the same direction. Having equal probability flows along these transitions would form a flow loop.

Note that the full state space S includes both where you are in the base space X and which way you are intending to walk, V. One can tile $\mathbb{Z}^2 \times V$ with these squares. Figure 3.2 demonstrates exactly how to do this.

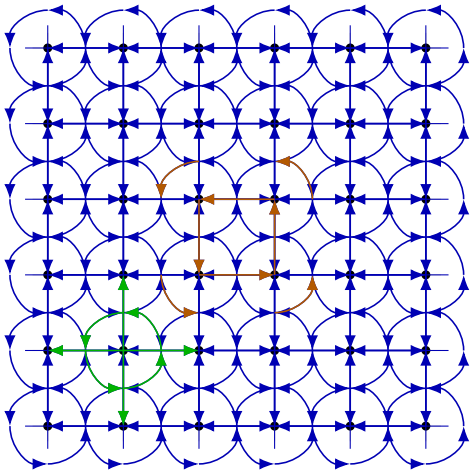

Figure 3.2: Though very cluttered, one can see how the space is tiled by square components, like the one in orange, while every possible transition is included for every node. The latter can be confirmed by looking at the arrows highlighted in green. Every turn and a walk in every direction from that node is included.

This concept can be generalized to continuous spaces. Let us choose $X = V = \mathbb{R}^2$ and define $W : S \to S$ as $W(x, v) := (x + v, v)$ and $T : S \to S$ as $T(x, v) := (x, Rv)$. For any state $s \in S$ one can consider the loop generated by alternating W and T, eventually completing the walk around the square.

$$s \longrightarrow W(s) \longrightarrow T(W(s)) \longrightarrow \cdots \longrightarrow W\left((T \circ W)^3(s)\right) \longrightarrow (T \circ W)^4(s) = s.$$

One can form such a loop from any starting state $s \in S$.
A natural next step is to add a velocity refreshment step for a measure $\mathcal{U}$ on V that is invariant under 90 degree rotations[5]. This refreshment step ensures that the

[5]This invariance of $\mathcal{U}$ will help us in the final proof as this imbues the measure $\mathcal{U}$ with the symmetry of R.

process won't get stuck in an axis-aligned lattice with fixed step size, so the process can hopefully become ergodic. Thus our current proposed kernel would be of the form

$$\begin{aligned}\mu\left((x,v),\mathrm{d}(x',v')\right) = a(x,v)\delta_{(x+v,v)}\left(\mathrm{d}(x',v')\right)\\ + b(x,v)\delta_{(x,Rv)}\left(\mathrm{d}(x',v')\right)\\ + s(x)\delta_x\left(\mathrm{d}x'\right)\mathcal{U}(\mathrm{d}v').\end{aligned}$$

Our next generalization will be that of the map R. We began with a rotation map, but it turns out that all we need is a map of finite order n, meaning $R^n = \mathrm{id}_V$, that is invariant to $\mathcal{U}$, meaning[6] $\mathcal{U} \circ R^{-1} = \mathcal{U}$, with a final loop condition

$$\sum_{j=1}^{n} R^j v = 0.$$

This corresponds to the vectors of all the edges of the loops cancelling and thus putting them end to end forms a closed loop. Note that this generalizes beyond the need for tiling the plane with loops as in Figure 3.2. Our loops can therefore form arbitrary regular n-gons in the base space. Thus the total number of steps, counting both turn steps and walk steps, is $2n$ around an n-gon.

The final generalization of our setting is generalizing the spaces from $X = V = \mathbb{R}^2$. The essential element we need from our space is a translation invariant measure on the space and a natural way for V to represent displacements in X. Therefore, we will be using a locally compact topological group X so it has a regular left Haar measure λ. We will consider subspaces $V \subseteq X$ and let the group operation represent this displacement relation, meaning $x + v$ gets replaced by vx.

What do we demand of a and b to ensure alternating between walking and turning results in a loop map? Firstly, one must ensure that the turn flow before each turn step is the same. To simplify, we can change from the turn rate b to a turn probability flow t by letting $b(x,v) = \frac{\mathrm{d}\lambda}{\mathrm{d}\pi}t(x,v)$. A Radon-Nikodym derivative $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ corresponds to $\frac{1}{\pi(\{x\})}$ from the discrete case in Remark 10. With this notation, the turn flow being the same after turning and walking corresponds to

$$t(x,v) = t((Rv)x, Rv).$$

Next, we need that the flow of a step followed by a turn matches. This means that $a(x,v) = \frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)t(vx,v)$. This can be seen by considering a walk from $v^{-1}x$ to x, followed by a turn. The flow of the walk will be correspond to the rate $a(v^{-1}x,v)$ and thus the flow $t(v(v^{-1}x),v) = t(x,v)$, which is exactly the turn flow in (x,v).

Alternating these two properties successively n times each shows that they form a flow loop.

We are finally prepared to face the statement of the following theorem, defining the kernel foundational for the second sampler of this thesis.

[6]The symbol R^{-1} is to be understood as the pre-image of R, but its inverse does exist as R has finite order. Therefore $R^{-1} = R^{n-1}$.

Theorem 6 (Invariance of a loop sampler on a locally compact Polish group)**.** *Let X be a locally compact Polish group with a left Haar measure λ. Assume $\pi \ll \lambda$ and $\lambda \ll \pi$ for a probability measure π on X. Let $V \subseteq X$ be measurable and Polish, let $\mathcal{U}$ be a finite measure on V and let*

$$R : V \to V,$$

be measurable with finite order $n \geq 2$, that is,

$$R^n = \mathrm{id}_V.$$

Assume that $\mathcal{U}$ is R-invariant, meaning

$$\mathcal{U} \circ R^{-1} = \mathcal{U}.$$

Define

$$\Gamma_0(v) := e, \qquad \Gamma_j(v) := R^j(v)R^{j-1}(v)\dots R(v) \quad \text{for } j \geq 1,$$

where e is the identity of X. Assume that R is a loop map in the sense that

$$\Gamma_n(v) = e \quad \text{for every } v \in V.$$

Let $t : X \times V \to \mathbb{R}$ be non-negative $\lambda \otimes \mathcal{U}$-almost everywhere, sufficiently integrable and satisfy

$$t(x, v) = t((Rv)x, Rv) \quad \text{for all } (x, v) \in X \times V.$$

Define the measure $\breve{\pi} = \pi \otimes \mathcal{U}$ on $X \times V$ and the jump kernel

$$\mu((x, v), \mathrm{d}(x', v')) := \frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x) \left(t(vx, v)\delta_{(vx,v)} + t(x, v)\delta_{(x,Rv)} \right) (\mathrm{d}(x', v')).$$

Then for all test functions f we have that

$$\int_{X\times V} \int_{X\times V} (f(x', v') - f(x, v))\mu((x, v), \mathrm{d}(x', v'))\breve{\pi}(\mathrm{d}(x, v)) = 0,$$

meaning $\breve{\pi}$ is invariant to μ.

Proof. We begin by simplifying the integral.

$$
\begin{aligned}
&\int_{X\times V}\int_{X\times V}(f(x',v')-f(x,v))\mu((x,v),\mathrm{d}(x',v'))\check{\pi}(\mathrm{d}(x,v))\\
=&\int_{X\times V}\int_{X\times V}(f(x',v')-f(x,v))\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\left(t(vx,v)\delta_{(vx,v)}+t(x,v)\delta_{(x,Rv)}\right)(\mathrm{d}(x',v'))\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
=&\int_{X\times V}\int_{X\times V}(f(x',v')-f(x,v))\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)t(vx,v)\delta_{(vx,v)}(\mathrm{d}(x',v'))\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
+&\int_{X\times V}\int_{X\times V}(f(x',v')-f(x,v))\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)t(x,v)\delta_{(x,Rv)}(\mathrm{d}(x',v'))\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
=&\int_{X\times V}(f(vx,v)-f(x,v))\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)t(vx,v)\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
+&\int_{X\times V}(f(x,Rv)-f(x,v))\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)t(x,v)\pi(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
\overset{\text{Radon-Nikodym}}{=}&\int_{X\times V}(f(vx,v)-f(x,v))t(vx,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
+&\int_{X\times V}(f(x,Rv)-f(x,v))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
=&\left[\begin{array}{c}u=R^{-1}v\implies v=Ru\\ \text{Measurable bijection. }\mathcal{U}\text{ is }R\text{-invariant.}\end{array}\right]\\
=&\int_{X\times V}(f((Ru)x,Ru)-f(x,Ru))t((Ru)x,Ru)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}u)\\
+&\int_{X\times V}(f(x,Rv)-f(x,v))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
\overset{\text{relabel}}{=}&\int_{X\times V}(f((Rv)x,Rv)-f(x,Rv))t((Rv)x,Rv)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
+&\int_{X\times V}(f(x,Rv)-f(x,v))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
\overset{t((Rv)x,Rv)=t(x,v)}{=}&\int_{X\times V}(f((Rv)x,Rv)-f(x,v))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)=:I_0
\end{aligned}
$$

Now let us define I_1 using

$$
\begin{aligned}
I_0=&\int_{X\times V}(f((Rv)x,Rv)-f(x,v))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)\\
=&\left[\begin{array}{c}u=R^{-1}v,\ \ y=v^{-1}x\implies v=Ru,\ \ x=vy=(Ru)y\\ \text{Both are measurable bijections. }\mathcal{U}\text{ is }R\text{-invariant.}\\ \lambda\text{ is left Haar, so left-translation invariant.}\end{array}\right]\\
=&\int_{X\times V}(f((R^2u)(Ru)y,R^2u)-f((Ru)y,Ru))t((Ru)y,Ru)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}u)\\
\overset{\text{t((Ru)y,Ru)=t(y,u)}}{=}&\int_{X\times V}(f(\Gamma_2(u)y,R^2u)-f(\Gamma_1(u)y,Ru))t(y,u)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}u)\\
\overset{\text{relabel}}{=}&\int_{X\times V}(f(\Gamma_2(v)x,R^2v)-f(\Gamma_1(v)x,Rv))t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v)=:I_1.
\end{aligned}
$$

In the same manner, we can recursively define I_{j+1} from I_j for $j\geq 1$ as

$$
\begin{aligned}
I_j &= \int_{X\times V} \left(f\left(\Gamma_{j+1}(v)x, R^{j+1}v\right) - f\left(\Gamma_j(v)x, R^j v\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \left[\begin{array}{c} u = R^{-1}v,\ \ y = v^{-1}x \implies v = Ru,\ \ x = vy = (Ru)y \\ \text{Both are measurable bijections. } \mathcal{U} \text{ is } R\text{-invariant.} \\ \lambda \text{ is left Haar, so left-translation invariant.} \end{array}\right] \\
&= \int_{X\times V} \left(f\left(\Gamma_{j+1}(Ru)(Ru)y, R^{j+2}u\right) - f\left(\Gamma_j(Ru)(Ru)y, R^{j+1}u\right)\right) t((Ru)y, Ru)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}u) \\
&= \left[\Gamma_m(Ru)(Ru) = \Gamma_{m+1}(u),\ t((Ru)y, Ru) = t(y,u)\right] \\
&= \int_{X\times V} \left(f\left(\Gamma_{j+2}(u)y), R^{j+2}u\right) - f\left(\Gamma_{j+1}(u)y, R^{j+1}u\right)\right) t(y,u)\lambda(\mathrm{d}y)\mathcal{U}(\mathrm{d}u) \\
&\overset{\text{relabel}}{=} \int_{X\times V} \left(f\left(\Gamma_{j+2}(v)x, R^{j+2}v\right) - f\left(\Gamma_{j+1}(v)x, R^{j+1}v\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) =: I_{j+1}.
\end{aligned}
$$

Next, since all of these are equal to I_0, which is finite since it's a sum of two finite integrals, we get

$$
\begin{aligned}
nI_0 &= \sum_{j=0}^{n-1} I_j = \sum_{j=0}^{n-1} \int_{X\times V} \left(f\left(\Gamma_{j+1}(v)x, R^{j+1}v\right) - f\left(\Gamma_j(v)x, R^j v\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \int_{X\times V} \left(\sum_{j=0}^{n-1} \left(f\left(\Gamma_{j+1}(v)x, R^{j+1}v\right) - f\left(\Gamma_j(v)x, R^j v\right)\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&\overset{\text{telescopic}}{=} \int_{X\times V} \left(f\left(\Gamma_n(v)x, R^n v\right) - f\left(\Gamma_0(v)x, R^0 v\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) \\
&= \left[\begin{array}{c} \Gamma_0(v) = e = \Gamma_n(v) \\ R^n = \mathrm{id}_V = R^0 \end{array}\right] \\
&= \int_{X\times V} \left(f\left(x,v\right) - f\left(x,v\right)\right) t(x,v)\lambda(\mathrm{d}x)\mathcal{U}(\mathrm{d}v) = 0.
\end{aligned}
$$

But if $nI_0 = 0$, then $I_0 = 0$ as $n \geq 2$ and we are done.

□

Previously, we generalized R from a rotation map, so it is clear that one can choose R to be a rotation map and choose $\mathcal{U}$ to be some positive probability measure which is invariant under R. What is not clear is what t should be. A trivial valid choice is a constant function.

A concern with this sampler is that it may not mix quickly, despite being non-reversible. It could exhibit random walk behavior on large scales due to uncreative choices of t or, more likely, due to redundant turning. Looking back at Figure 3.2 we can see that the green turning arrows highlight a full rotation at that point, demonstrating how such turning occurs around every point. These arrows do not form a single flow loop, as they belong to different square loops, which may have different probability flows. If they are all positive, they could have redundant flow "spinning around" in place. We want to minimize our turning as it does not explore

the X space, what we are really interested in. Remember that V only exists to create an exploratory dynamic in X. Therefore we'd like to subtract a kernel of the following form.

Definition 9 (Spin kernel)**.** *In the context above, a spin kernel is a kernel of the form*

$$\nu\left((x,v),\mathrm{d}(x',v')\right) := \min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\delta_{(x,Rv)}\left(\mathrm{d}(x',v')\right),$$

for a sufficiently integrable function $a : X \times V \to \mathbb{R}$.

Theorem 7 (Invariance of a spin kernel)**.** *Given that $\mathcal{U}$ is R-invariant, all measures $\check{\pi} = \pi \otimes \mathcal{U}$ are invariant to all spin kernels ν.*

Proof. Take some measure $\check{\pi}$ and a spin kernel ν. Let us use the invariance condition. Let us define I_0 as

$$\begin{aligned}
&\int\int (f(x',v') - f(x,v))\nu\left((x,v),\mathrm{d}(x',v')\right)\check{\pi}\left(\mathrm{d}(x,v)\right)\\
=&\int\int (f(x',v') - f(x,v))\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\delta_{(x,Rv)}\left(\mathrm{d}(x',v')\right)\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
=&\int (f(x,Rv) - f(x,v))\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v) =: I_0.
\end{aligned}$$

We will proceed similar to the proof of Theorem 6. We recursively define I_{k+1} from I_k as

$$\begin{aligned}
I_k =& \int (f(x,R^{k+1}v) - f(x,R^k v))\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
=& \begin{bmatrix} u = R^{-1}v \implies v = Ru \\ \text{Measurable bijection. } \mathcal{U} \text{ is } R\text{-invariant. The minimum is unchanged as} \\ \{Rv, R^2v, \dots, R^{n-1}v, R^n v\} = \{R^2 v, \dots, R^n v, Rv\} = \{R^2 v, R^3 v, \dots, R^n v, R^{n+1}v\}. \end{bmatrix}\\
=& \int (f(x,R^{k+2}u) - f(x,R^{k+1}u))\min_{j\in\{1,\dots,n\}}\{a(x,R^j u)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}u) =: I_{k+1}.
\end{aligned}$$

To show $I_0 = 0$, we show that $nI_0 = 0$. Take

$$\begin{aligned}
nI_0 =& \sum_{k=0}^{n-1} I_k = \sum_{k=0}^{n-1}\int (f(x,R^{k+1}v) - f(x,R^k v))\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
=& \int \sum_{k=0}^{n-1}\Big((f(x,R^{k+1}v) - f(x,R^k v))\Big)\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
\overset{\text{telescopic}}{=}& \int (f(x,R^n v) - f(x,v))\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\}\pi\left(\mathrm{d}x\right)\mathcal{U}(\mathrm{d}v)\\
=&0,
\end{aligned}$$

where we note that the last integral is 0 because $R^n v = v$ as n is the order of R by definition. □

The following theorem shows us how to use this to get rid of flow wasted on just turning around in place.

Theorem 8 (Invariance of a loop sampler with turn correction)**.** *Let μ be the loop sampler kernel from Theorem 6 and ν be the spin kernel from Definition 9 where $a(x,v) = t(x,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$. Then*

$$\iota := \mu - c\nu,$$

is a kernel and $\check{\pi}$ is invariant with respect to it if $0 \leq c \leq 1$.

Proof. We need to verify that the kernel subtraction is well-defined. For this, we need to know that $\mu\left((x,v), \mathrm{d}(x',v')\right) \geq c\nu\left((x,v), \mathrm{d}(x',v')\right)$. Note that

$$\min_{j\in\{1,\dots,n\}}\{a(x,R^j v)\} = \min_{j\in\{1,\dots,n\}}\left\{t(x,R^j v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\right\} = \min_{j\in\{1,\dots,n\}}\left\{t(x,R^j v)\right\}\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x).$$

We are now ready for the estimate.

$$\begin{aligned}
\mu\left((x,v), \mathrm{d}(x',v')\right) &= \frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\left(t(vx,v)\delta_{(vx,v)} + t(x,v)\delta_{(x,Rv)}\right)\left(\mathrm{d}(x',v')\right) \\
&\geq t(x,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(x,Rv)}(\mathrm{d}(x',v')) \\
&\geq \min_{j\in\{1,\dots,n\}}\left\{t(x,R^j v)\right\}\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(x,Rv)}(\mathrm{d}(x',v')) \\
&= \nu\left((x,v), \mathrm{d}(x',v')\right) \\
&\geq c\nu\left((x,v), \mathrm{d}(x',v')\right)
\end{aligned}$$

Since multiplying a kernel by a constant non-negative factor produces another kernel, invariance follows directly from Theorem 4. □

Remark 14. *One can inspect the chosen a in the theorem above and realize that this is the maximal possible a in general for $c = 1$. The subtraction takes away exactly the minimum turn flow around a point, meaning more can't be subtracted. This means that for every state $(x,v) \in S$, at least one of the states $(x,v), (x,Rv), (x,R^2v), \dots, (x,R^{n-1}v)$ has 0 rate to turn.*

Remark 15. *Since the minimum turn flow around a point is subtracted when $c = 1$, if they form a flow loop, the process has 0 rate to turn around that state.*

We call the factor c the turn correction factor as it describes how much turn correction is performed. One can superimpose multiple loop samplers with different t and R and still preserve invariance. Let us also add a velocity refreshment kernel so it has a chance of being ergodic.

Theorem 9 (Invariance of the full loop sampler)**.** *Let $N \geq 1$. Take $\iota_1, \dots, \iota_N$ kernels that fulfill Theorem 8 for possibly different t, n and R, so let us index these by $t_1, \dots, t_N$, $n_1, \dots, n_N$ and $R_1, \dots R_N$. Let γ be a velocity refreshment kernel. Then*

$$\xi := \gamma + \sum_{k=1}^{N} \iota_k,$$

or more verbosely

$$\begin{aligned}\xi((x,v),\mathrm{d}(x',v')) =& s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v') \\ &+ \sum_{k=1}^{N}\Bigg(\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\left(t_k(vx,v)\delta_{(vx,v)} + t_k(x,v)\delta_{(x,R_kv)}\right)(\mathrm{d}(x',v')) \\ &\qquad - c\min_{j\in\{1,\dots,n_k\}}\left\{t_k(x,R_k^jv)\right\}\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(x,R_kv)}(\mathrm{d}(x',v'))\Bigg),\end{aligned}$$

is a kernel and $\check{\pi}$ is invariant with respect to it.

Proof. The result follows directly from kernel superposition, Theorem 3. □

Remark 16 (Rewriting the kernel ξ)**.** *The kernel ξ from above can easily be rewritten as*

$$\begin{aligned}\xi((x,v),\mathrm{d}(x',v')) =& s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v') \\ &+ \sum_{k=1}^{N}\Bigg(t_k(vx,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(vx,v)}(\mathrm{d}(x',v')) \\ &\qquad + \left(t_k(x,v) - c\min_{j\in\{1,\dots,n_k\}}\left\{t_k(x,R_k^jv)\right\}\right)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(x,R_kv)}(\mathrm{d}(x',v'))\Bigg),\end{aligned}$$

by grouping terms based on type of step. If all $t_k(x,v) > 0$ for all states $(x,v) \in S$, then one can further rewrite it as

$$\begin{aligned}\xi((x,v),\mathrm{d}(x',v')) =& s(x)\delta_x(\mathrm{d}x')\mathcal{U}(\mathrm{d}v') \\ &+ \sum_{k=1}^{N}\Bigg(t_k(vx,v)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(vx,v)}(\mathrm{d}(x',v')) \\ &\qquad + t_k(x,v)\left(1 - c\min_{j\in\{1,\dots,n_k\}}\left\{\frac{t_k(x,R_k^jv)}{t_k(x,v)}\right\}\right)\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)\delta_{(x,R_kv)}(\mathrm{d}(x',v'))\Bigg).\end{aligned}$$

Note that the last element of the minimum is 1 as $R_k^{n_k}v = v$. The kernel ξ is the kernel of a fully general loop sampler.

Remark 17. *One only needs to know $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ up to a constant as the unknown constant can be baked into t.*

The argument supporting the remark above is completely analogous to the one for Remark 13. Thus the loop samplers also have the desired property Metropolis-Hastings has [2, p. 167].

3.5 On ergodicity

Ergodicity results regarding the flow walker sampler and the loop sampler are outside the scope of this thesis, but some guidance towards proving such results will be provided below.
In Jansson et al. they use the Meyn-Tweedie approach to prove a stronger condition, geometric ergodicity, using Lyapunov functions [3, p. 20]. Using geometric ergodicity and Harris recurrence one can prove a central limit theorem for the MCMC estimator. This establishes asymptotic normality of the estimator under sufficient conditions [4, p. 12–13][7].
The intuitive idea is that within a timespan $[0, t]$ there can be a positive density to perform jumps to go from an arbitrary state (x, v) to (x', v'). Firstly, a velocity refreshment step occurs that sets the new velocity to the element v'' fulfilling $v''x = x'$, meaning $v'' = x'x^{-1}$ so the state ends up being (x, v''). Secondly, it takes a walk step, ending up at $(v''x, v'') = (x'x^{-1}x, v'') = (x', v'')$. Lastly, it performs another velocity refreshment step setting the velocity to v' and ending up at (x', v'). Since the states are arbitrary, this exact argument requires that $V = X$ and that every velocity has a positive density to be chosen by $\mathcal{U}$. One can imagine weaker assumptions letting $V \subset X$ where one takes more steps to end up at the desired destination. This intuitive argument establishes that the process can walk from anywhere to anywhere, meaning the Markov process is irreducible. Note that periodicity is not a problem for Markov jump processes like they are for Markov chains as jump times are random. Based on this, one can see how it is plausible that the samplers constructed are ergodic under sufficient conditions.

[7] This citation credits Seyer as this is where it was taken from, but Seyer cites Jones [10, p. 10–11] for the resulting theorem.

4

Method

In Chapter 2 we established the theoretical setting. Then in Chapter 3 we proved the invariance of rate kernels corresponding to the new samplers. Next we will cover how to use the rate kernel to simulate the Markov jump process in Section 4.1. Section 4.2 covers implementation details specific to the example implementation developed for the project. Section 4.3 defines relative error and motivates why it is the appropriate performance metric. In the last sections we introduce an example space and example samplers on that space for empirical tests. The results of these can be found in Chapter 5.

4.1 Simulation of time-homogeneous Markov jump processes using rate kernels

Using the notation from Section 2.2, the problem of simulating a time-homogeneous Markov jump process from its rate kernel can be reduced to finding $Z_{\tau_{i+1}}$, the new state, and $\tau_{i+1}-\tau_i$, the wait time before jumping, based on Z_{τ_i}, the current state. Let us call the rate kernel μ. The wait time can be generated by taking an independent sample of the $\mathrm{Exp}(\lambda_\mu(Z_{\tau_i}))$ distribution. Recall that using the Rao-Blackwellization from Theorem 2 makes generating the wait time unnecessary in practice. One can sample the next step $Z_{\tau_{i+1}}$ by taking a sample from a random variable with law $\kappa_\mu(Z_{\tau_i}, \cdot)$.
If the rate kernel is written as a superposition of multiple kernels, additional computational tricks can be used. Suppose that

$$\mu = \sum_{k\in I} a_k \mu_k,$$

for a non-negative sequence of reals $(a_k)_{k\in I}$. Then one can compute the total rate as

$$\lambda_\mu(x) = \mu(x, X) = \sum_{k\in I} a_k \mu_k(x, X) = \sum_{k\in I} a_k \lambda_{\mu_k}(x).$$

Recall that X is the state space according to the notation from Section 2.2. To calculate where to jump, one can take a sample from a random variable with law $\kappa_{\mu_k}(Z_{\tau_i}, \cdot)$ with probability

$$\frac{a_k \lambda_{\mu_k}(Z_{\tau_i})}{\sum_{j\in I} a_j \lambda_{\mu_j}(Z_{\tau_i})}.$$

4.2 Implementation details

Python is used for the example implementations of the samplers. The topological space is represented using $\mathbb{R}^2$ and a kind of quotient map where the elements $[-1, 1)$ represent their equivalence classes. Both samplers support quotient spaces of $\mathbb{R}^d$ for arbitrary dimensions d. As termination conditions, they support both steps and a time limit. They also optionally support burn-in, seeded runs through a supplied numpy random number generator and the Rao-Blackwellization from Theorem 2. The latter is enabled by default and used in the examples. It is worth noting that the loop sampler implementation is limited to $N = 1$ from Theorem 8. The implementation can be found at
https://github.com/VincentHarbander/FromSymmetriesToStationarityInMMC.

4.3 Relative error

Suppose we are trying to estimate the probability of a measurable set A under a measure π, meaning $p := \pi(A)$. After constructing an estimator $\hat{p}$, we would like to assess its quality. Since probabilities of measurable sets can potentially be very small, we will use relative error as our performance metric. If p is the target probability and $\hat{p}$ is our sampler's empirical estimate of p, then we define the relative error of $\hat{p}$ to p as

$$\text{Relative error} = \frac{\hat{p} - p}{p} = \frac{\hat{p}}{p} - 1.$$

The relative error being small in absolute value indicates that our sampler functions well. An estimate $\hat{p}$ having a relative error of -0.01 means that it is $0.99p$. A positive relative error indicates that the estimate is larger than the estimand, while a negative relative error indicates that the estimate is smaller than the estimand.

4.4 Example space

To show that the samplers function in practice, we need an example space and target measure on that space.

For our examples, we choose $X = S^1 \times S^1$, a torus. Specifically let us parametrize the circles S^1 using $[-1, 1)$. Let us choose a Gaussian with mean $(0, 0)$ and identity covariance matrix $\mathbf{I}$ truncated to the torus as our target π. Let $V = X$ and $\mathcal{U}$ be a wrapped Gaussian with mean $(0, 0)$ and covariance matrix $\frac{1}{10}\mathbf{I}$. The group operation on X is coordinate-wise addition up to wrapping. This forms a Polish group. It is also locally compact as it is compact.

Such a space can naturally arise from a robot with two joints that each can make a full rotation.

4.5 Flow walker example

For our flow walker example, let $h(x, v) = 1$. We let $s(x) = \frac{1}{5}\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}(x)$. Since our density $\frac{\mathrm{d}\pi}{\mathrm{d}\lambda}$ exists, $\frac{\mathrm{d}\lambda}{\mathrm{d}\pi}$ and s is positive and therefore a valid choice. This particular s was chosen so it is proportional to the walk rate, while being smaller than it.

4.6 Loop sampler example

Next we construct a loop sampler example. We let $N = 1$ from Theorem 8. We let R be a counter clockwise rotation of order n. For a starting state $s_1 = (x_1, v_1)$ we define $(s_k)_{k=1}^n$ recursively as $x_{k+1} = (Rv_k)x_k$ and $v_{k+1} = Rv_k$ where $s_k = (x_k, v_k)$. These form the corners of a regular n-gon on our torus. The property t must fulfill is $t(x, v) = t((Rv)x, Rv)$, so it is the same for the states around the n-gon. We can therefore choose

$$t(x_1, v_1) = \prod_{k=1}^{n} \frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(x_k).$$

It is notable that the natural algorithm for computing a value of $t(x, v)$ has asymptotic time complexity $\mathcal{O}(n)$. This leads to the complexity for a single step of the loop sampler using the spin reduction to be $\mathcal{O}(n^2)$ since one takes the minimum of n evaluations of t. This should not be a problem for our practical usage as I recommend one chooses a small n. This can also be addressed with different choices of t or a better way of calculating these quantities for the specific problem at hand. We choose $n = 3$ for our experiment.
Let $s(x) = \frac{1}{5}\left(\frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(x)\right)^n$. This choice makes sense if one supposes that $\frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(x_k) \approx \varepsilon$ for all x_k in the n-gon. Then one can see that $s(x_k)$ is almost proportional to and smaller than $t(x_k, v_k)$ as

$$s(x_k) = \frac{1}{5}\left(\frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(x_k)\right)^n \approx \frac{1}{5}\varepsilon^n \propto \varepsilon^n \approx \prod_{k=1}^{n} \frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(x_k) = t(x_k, v_k).$$

5

Results

Recall the example space and target from Section 4.4. We can partition it into a 20 by 20 uniform grid of measurable squares. Let us plot a heatmap of our target measure π on this space. The plotted value of a square A is $\pi(A)$. This can be found in Figure 5.1.

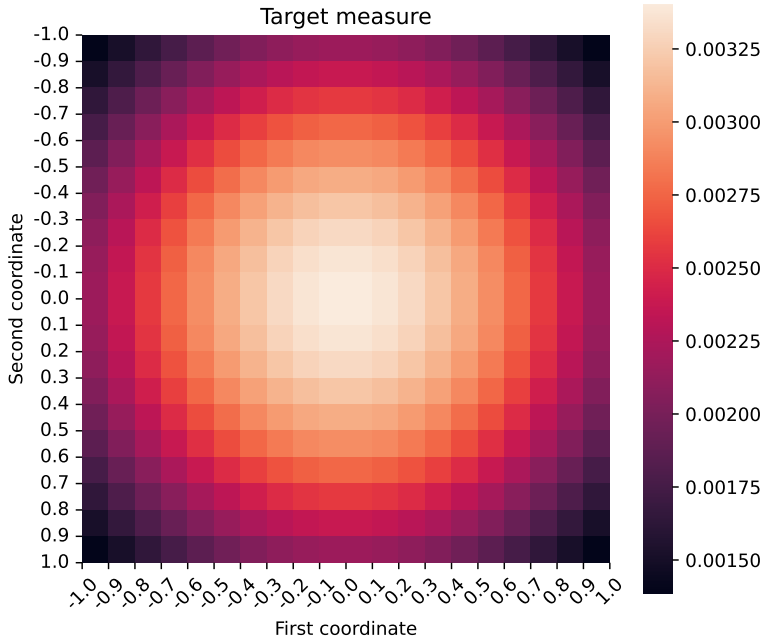

Figure 5.1: Heatmap of the target measure from Section 4.4.

Running the example flow walker sampler and loop sampler specified in Sections 4.5 and 4.6 with 1 million samples, we can plot the estimated probabilities for the squares. This can be found in Figures 5.2 and 5.3.

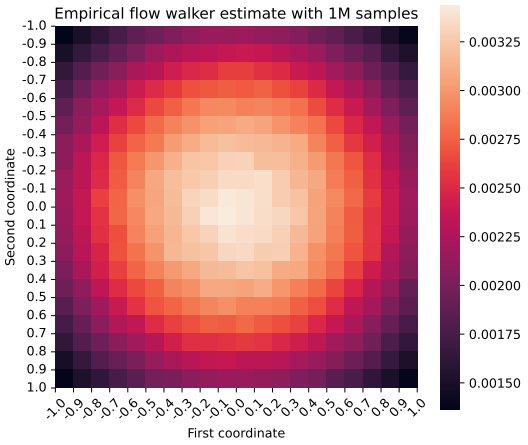

Figure 5.2: Heatmap of the empirical target using the flow walker sampler with 1 million samples.

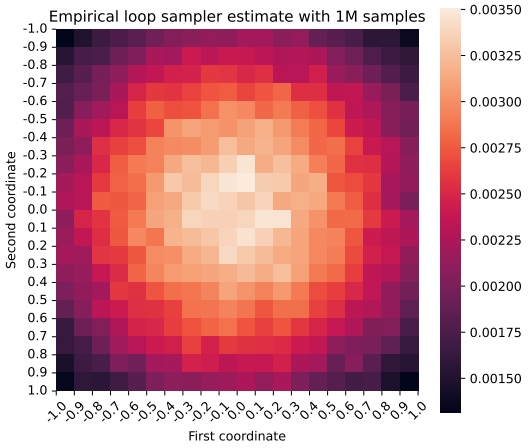

Figure 5.3: Heatmap of the empirical target using the loop sampler with 1 million samples.

Next we plot the relative error from 4.3 of the estimated probabilities using the samplers to the true measure of the squares. See Figures 5.4 and 5.5.

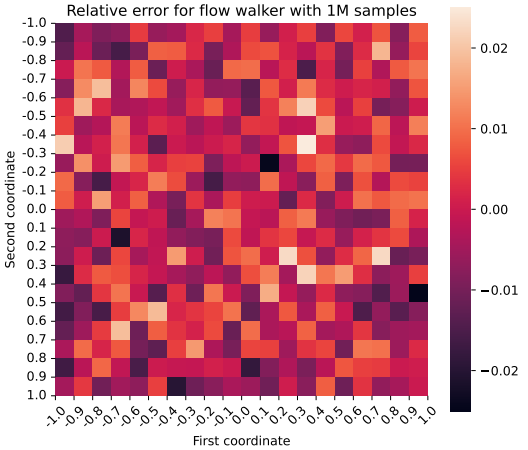

Figure 5.4: Relative error of the flow walker estimate to the true probabilities with 1 million samples.

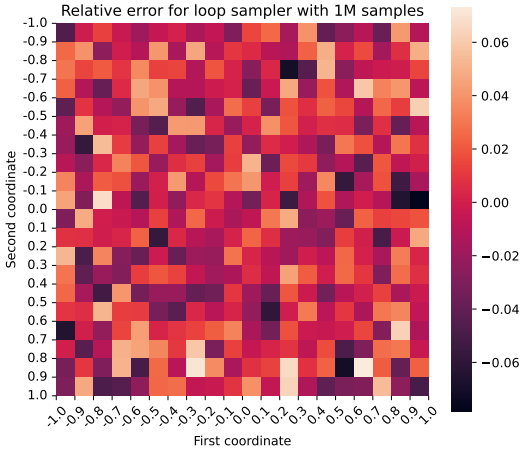

Figure 5.5: Relative error of the loop sampler estimate to the true probabilities with 1 million samples.

One can detect some variation in the empirical estimates in Figures 5.2 and 5.3 compared to the target measure in Figure 5.1. The relative errors appear quite small in absolute value in Figures 5.4 and 5.5, however. It is also a good sign that the relative error does not seem to have a biased structure, which could be the result of a bug. Rerunning the samplers with fewer samples produces larger relative errors

in absolute value, which also makes sense. The implementations seem to work and converge to the target distribution.
The flow walker can be suspected to outperform the loop sampler because the loop map doesn't provide any meaningful exploratory advantage on the torus, but turn steps resample the same point in the base space when turning. Ignoring turn correction and assuming $t(vx, v) \approx t(x, v)$, the loop sampler should redundantly sample around half the time it performs a non-refreshment step. The flow walker does not have this redundancy. According to this argument, one should expect the loop sampler with a large number, $2M$, steps to perform around as well as the flow walker with M steps. At best, the loop sampler almost never turns and then has the same efficiency as the flow walker.

6
Discussion

This thesis is by no means an exhaustive exploration of the possibilities of non-reversible continuous-time Markov chain Monte Carlo algorithms. The reader is recommended to read Jansson et al. [3] or Samuel Power's PhD dissertation "Exploring Probability Measures with Markov Processes" [7] for a coverage of the skew-detailed balance theory. Next we will discuss possible further developments of the findings of this thesis.

The question of ergodicity is not formally resolved. One can find very restrictive conditions for ergodicity to hold, but a non-trivial result for non-compact spaces would be of practical use.

When making MMC samplers one generally constructs the kernel in some way driven by intuition and proves the invariance of the full kernel in the end using some invariance condition. This thesis takes an unusual approach by composing invariant components. Separating the kernel into invariant components decreases the complexity of each proof in exchange for creating more of them. Hopefully this improved readability.

One can take this approach further and generalize Theorem 3 on kernel superposition to continuous superpositions of a family of kernels. These can represent variable step lengths, for example, making the process step by the step length times the velocity. Another option is to have continuous superpositions of a family of different types of samplers. A simple option would be a mixture of loop samplers with different turn correction factors and loop maps.

The setting of the samplers should be able to be usefully generalized. The proofs are premised on the translation invariance of the left Haar measure. This is a natural translation invariant setting for a Lie group, but one may want to sample on state spaces that can't be equipped with any natural topological group structure, such as the S^2 sphere [11, p. 75]. Future research could generalize the setting of the samplers. One suggestion is to study group actions on the base space instead of assuming the base space is a group by itself. Studying group actions on the base space is actually something Power did [7, p. 139].

Another possible generalization of the loop sampler is that turning only moves in the augmentation space, taking another sample of the same point in the base space. This is wasteful. One could explore a type of loop sampler that performs a step after every time it performs a turn. Alternating between stepping and turn-stepping forms regular polygons of double the side length. Turn correction in a point is now obsolete, but a variant of it can still be performed in subtracting flow loops formed by sequential turn-steps, forming regular polygons of the original side length. This approach would require a new property for t. I propose $t(x, v) = t((Rv)^2 x, (Rv)v)$

as it preserves the same symmetry for the alternating step flow loop as the current t property does for its alternating step flow loop.
The flow sampler can be further developed by combining it with the piecewise-deterministic Markov process (PDMP) framework. One can imagine how decreasing the step length while increasing the rates can yield a PDMP in the limit. This should produce lines as invariant components. A similar approach can be imagined for the loop sampler, but one needs to increase the order of the R map in the limit. Doing so appropriately should cause circles to become invariant components. One can read about PDMPs in Power's PhD dissertation [7, p. 84].
No meaningful ethical considerations specific to this project were encountered as this is a theoretical mathematical pursuit that may only outperform existing algorithms in some settings. Further ethical considerations would be required if the project unlocked new qualitative capabilities or yielded significant quantitative advantages over other existing methods.

7
Conclusion

This work accomplished its primary objective as specified in the problem formulation, Section 1.1, while respecting the limitations in Section 1.2. Two non-reversible, rejection-free Markovian Monte Carlo samplers were developed in the bottom-up approach specified. Their stationarity was proven under sufficient conditions by showing the invariance of the target measure to their rate kernels.
The secondary objective was also achieved as example implementations of both of the families of samplers are now available online for everyone.
This thesis hopes to inspire future work to pursue this bottom-up approach in the proofs of invariance.

Bibliography

[1] C. Robert and G. Casella, *Monte Carlo Statistical Models,* 2nd ed. New York, US: Springer, 2004.

[2] A. Gelman, and X.-L. Meng, "Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling," Statistical Science, vol. 13, no. 2, pp. 163–185, 1998.

[3] E. Jansson, M. Schauer, R. Seyer, and A. Sharma, "Rebalancing Markov jump processes for non-reversible continuous-time sampling," arXiv preprint arXiv:2504.12190, 2026.

[4] R. Seyer, "Coupled Timelines and Rebalanced Frogs: Bayesian Computation with Markov Processes", Licentiate thesis, Chalmers University of Technology, Gothenburg, Sweden, 2026.

[5] O. Kallenberg, *Foundations of Modern Probability,* 3rd ed. United States of America: Springer, 2021.

[6] L. Breiman, *Probability.* United States of America: Addison-Wesley, 1968.

[7] S. Power, "Exploring Probability Measures with Markov Processes," PhD dissertation, University of Cambridge, Cambridge, England, 2020.

[8] J. Elstrodt, *Maß- und Integrationstheorie,* 1st ed. Berlin and Heidelberg, Germany: Springer, 1996.

[9] S. Meyn, and R. Tweedie, *Markov Chains and Stochastic Stability,* 1st ed. London, UK: Springer-Verlag, 1993.

[10] G. Jones, "On the Markov Chain Central Limit Theorem," arXiv preprint arXiv:math/0409112, 2004.

[11] I. Santa-María Megía, "Which spheres admit a topological group structure?," Revista de la Real Academia de Ciencias Exactas, Físicas, Químicas y Naturales de Zaragoza, vol. 62, pp. 75–79, 2007.

[12] P. Halmos, *Measure theory.* United States of America: Springer, 1950.

A

Measure theoretic preliminaries

The following appendix is a brief summary of necessary prerequisite knowledge the reader needs to understand the thesis and may need a quick introduction to or refresher in. The construction of the Lebesgue integral is notably absent. This is because any sufficient explanations, definitions, theorems or proofs would be too long and have already been presented well by others. If the reader wants a refresher on that material, I can recommend *Measure theory* by Halmos.
The sections below are mostly not original work and simply adapted presentations of familiar theorems according to the notation of the thesis. The theorems are mostly taken from *Foundations of Modern Probability* by Kallenberg, *Probability* by Breiman and *Measure theory* by Halmos and attributed according to the citations.

A.1 Change of variables for the Lebesgue integral

Let μ be a measure on a measure space $(X, \mathcal{X})$. Suppose we also have a second measurable space $(Y, \mathcal{Y})$.

Theorem 10 (Change of variables)**.** *Let $f : X \to Y$ and $g : Y \to \mathbb{R}$. Then*

$$\int (g \circ f)\, \mu = \int g \, \left(\mu \circ f^{-1}\right),$$

if either side exists [5, p. 23]. Here, f^{-1} denotes the pre-image under f.

The following theorem can be rewritten if one lets $f(x) = \mathbf{1}_A(x)\tilde{f}(x)$ for some measurable A.

Corollary 2 (Change of variables with domain of integration)**.** *Let $f : X \to Y$ and $g : Y \to \mathbb{R}$. Then*

$$\int_{f^{-1}(A)} (g \circ f)\, \mu = \int_A g \, \left(\mu \circ f^{-1}\right),$$

if either side exists.

A.2 Fubini's theorem

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. Let g be $\mathcal{X} \otimes \mathcal{Y}$-measurable, where $\mathcal{X} \otimes \mathcal{Y}$ is the smallest σ-algebra containing the Cartesian product of $\mathcal{X}$ and $\mathcal{Y}$. Such a σ-algebra exists as it is the intersection of all such σ-algebras. Given μ is a measure

on $(X, \mathcal{X})$ and ν is a measure on $(Y, \mathcal{Y})$, one can define a unique product measure $\mu \otimes \nu$ that fulfills $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$ for every $A \in \mathcal{X}$ and $B \in \mathcal{Y}$. This measure is called the cross-product measure in Breiman [6, p. 399].

Theorem 11 (Fubini's theorem)**.** *Let g be $\mathcal{X} \otimes \mathcal{Y}$-measurable, and*

$$\int |g| \, \mathrm{d}(\mu \otimes \nu) < \infty.$$

Then

$$\int g \, \mathrm{d}(\mu \otimes \nu) = \int \left(\int g \, \mathrm{d}\mu \right) \, \mathrm{d}\nu = \int \left(\int g \, \mathrm{d}\nu \right) \, \mathrm{d}\mu \ \ [6, p.\ 399].$$

Remark 18. *Note that if g is non-negative $\mathcal{X} \otimes \mathcal{Y}$-almost everywhere then one only requires that the integral is finite to be able to change integration order.*

A.3 Variational theorem

The following section develops a theorem used to prove Theorem 15. It can be skipped by a reader not interested in reviewing the details of that proof.

Lemma 5. *If f is an integrable, function that is positive almost everywhere on a measurable set A and $\int f \, \mathrm{d}\pi = 0$, then $\pi(A) = 0$* [12, p. 104].

The following proofs and formulations are original, but the result is already known.

Lemma 6 (Variational lemma)**.** *Let $(X, \mathcal{X}, \pi)$ be a measure space. If for all measurable sets of finite measure A,*

$$\int_A g \, \mathrm{d}\pi = 0,$$

where g is a measurable function so the above integrals all exist, then $g = 0$ π-almost everywhere.

Proof. We proceed using a proof by contraposition. Therefore we assume that there exists some measurable set B of positive finite measure $\pi(B) > 0$ where g is non-zero and now aim to prove that there exists a set measurable set A of finite measure such that

$$\int_A g \, \mathrm{d}\pi \neq 0.$$

As g is either positive or negative for each point in B and g is measurable we can construct measurable sets

$$B_+ = B \cap g^{-1}((0, \infty)) \quad \text{and} \quad B_- = B \cap g^{-1}((-\infty, 0)),$$

and note that $B_+ \cup B_- = B$ and $B_+ \cap B_- = \emptyset$. From additivity we have

$$\pi(B_+) + \pi(B_-) = \pi(B) = 0.$$

As measures cannot be negative, at least one of $\pi(B_+)$ and $\pi(B_-)$ is positive. Assume first that B_+ has positive measure.

But then if $\int_{B_+} g \, \mathrm{d}\pi = 0$, Lemma 5 tells us that $\pi(B_+) = 0$, which contradicts that it has positive measure. Therefore $\int_{B_+} g \, \mathrm{d}\pi \neq 0$.
If B_+ had measure zero then B_- has to have positive measure. But then if $\int_{B_-} -g \, \mathrm{d}\pi = 0$, noting that $-g$ is a non-negative function on B_-, Lemma 5 tells us that $\pi(B_-) = 0$, which contradicts that it has positive measure. Therefore $\int_{B_-} -g \, \mathrm{d}\pi \neq 0$ and $\int_{B_-} g \, \mathrm{d}\pi \neq 0$.
The sets B_+ and B_- are sets of finite measure as they were produced by taking an intersection with the set B of finite measure. Therefore $A = B_+$ or $A = B_-$ depending on cases serve as a counter-example showing that there exists a measurable A of finite measure such that

$$\int_A g \, \mathrm{d}\pi \neq 0.$$

This completes the proof by contraposition. □

Theorem 12 (Variational theorem)**.** *Let $(X, \mathcal{X}, \pi)$ be a measure space. If for all measurable sets of finite measure A,*

$$\int_A f \, \mathrm{d}\pi = \int_A g \, \mathrm{d}\pi,$$

where f and g are measurable functions so the above integrals exist, then $f = g$ π-almost everywhere.

Proof. Subtract the right-hand side on both sides of Equation 12 and use the additivity of Lebesgue integrals.

$$\int_A f \, \mathrm{d}\pi - \int_A g \, \mathrm{d}\pi = \int_A f - g \, \mathrm{d}\pi = 0$$

Now apply Theorem 6 to the function $f - g$ to conclude that $f - g = 0$ π-almost everywhere, from which $f = g$ π-almost everywhere immediately follows. □

A.4 The Radon-Nikodym derivative

Let $(X, \mathcal{X}, \pi)$ be a measure space.

Definition 10 (σ-finite set)**.** *A set $A \in \mathcal{X}$ is called σ-finite with respect to π if it is a countable union of measurable sets of finite measure.*

Definition 11 (σ-finite measure)**.** *The measure π is called σ-finite if X is σ-finite with respect to π.*

Suppose π and $\tilde{\pi}$ are σ-finite measures on a measurable space $(X, \mathcal{X})$.

Definition 12 (Absolute continuity of measures)**.** *We say a measure π is absolutely continuous with respect to another measure $\tilde{\pi}$ if*

$$\tilde{\pi}(A) = 0 \implies \pi(A) = 0, \quad \textit{for all } A \in \mathcal{X} \ \ [6, p.\ 398].$$

We denote this by $\pi \ll \tilde{\pi}$.

Theorem 13 (Radon-Nikodym)**.** *If $\pi \ll \tilde{\pi}$ and both measures are σ-finite, then there exists a non-negative $\mathcal{X}$-measurable function we denote by $\frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}}$, that is unique up to $\tilde{\pi}$-null sets, such that for any $A \in \mathcal{X}$,*

$$\pi(A) = \int_A \frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}} \, \mathrm{d}\tilde{\pi} \quad [6, p.\ 398]. \tag{A.1}$$

We call the above property the Radon-Nikodym property.

Remark 19. *Notice how the notation $\frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}}$ is motivated by the Radon-Nikodym property. One can imagine informally cancelling the $\mathrm{d}\tilde{\pi}$ for each other.*

Remark 20. *It is unique only up to $\tilde{\pi}$-null sets as one can change the Radon-Nikodym derivative at $\tilde{\pi}$-null sets without changing the right-hand side of Equation* (A.1)*.*

Theorem 14 (Change of variables using the Radon-Nikodym derivative)**.** *Under the assumptions of Theorem 13, it is the case that*

$$\int_A f \, \mathrm{d}\pi = \int_A f \frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}} \, \mathrm{d}\tilde{\pi},$$

if the left integral exist [12, p. 134]. This is also called the Radon-Nikodym property as it is a generalization of it.

The reader who is unfamiliar with measure theory, but has knowledge of basic probability theory can reassure themselves by recognizing that the probability density function is simply the Radon-Nikodym derivative of the law of the random variable with respect to the Lebesgue measure. Likewise, the probability mass function is the Radon-Nikodym derivative of the law with respect to the counting measure.
The following theorem is well-known, but the formulation and proof is original.

Theorem 15 (Radon-Nikodym inverse)**.** *Let $\pi \ll \tilde{\pi}$ and $\tilde{\pi} \ll \pi$ for two σ-finite measures π and $\tilde{\pi}$. Then*

$$\frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}} = \left(\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}\right)^{-1}$$

$\tilde{\pi}$-almost everywhere. Any representative Radon-Nikodym derivatives fulfill this.

Proof. Let us choose any two representatives $\frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}}$ and $\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}$ and verify that the right-hand side of Equation 15 fulfills the Radon-Nikodym property. If it is a valid Radon-Nikodym derivative, then the left-hand side and right-hand side must agree $\tilde{\pi}$-almost everywhere by Remark 20.
Take an arbitrary $A \in \mathcal{X}$.

$$\begin{aligned}
\int_A \frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}} \, \mathrm{d}\tilde{\pi} &\overset{\text{Radon-Nikodym}}{=} \pi(A) = \int_A \mathrm{d}\pi = \int_A 1 \, \mathrm{d}\pi \\
= \int_A \left(\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}\right)^{-1} \frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi} \, \mathrm{d}\pi &\overset{\text{Radon-Nikodym}}{=} \int_A \left(\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}\right)^{-1} \mathrm{d}\tilde{\pi}
\end{aligned}$$

Since A is arbitrary, it holds for all $A \in \mathcal{X}$. The result follows directly from applying the variational theorem, Theorem 12. □

A.5 Haar measures

Let G be a topological group. This means that it is a group that is equipped with a σ-algebra such that the group inverse is a continuous map and that the group multiplication is a continuous map, as seen from the product space $G \times G$ to G. The topological space can be naturally extended to a measurable space using its Borel σ-algebra, the minimal σ-algebra that contains the topology.

Definition 13 (Hausdorff space)**.** *A topological space X is called Hausdorff if for all distinct points $x, y \in X$ there exists open sets A and B such that $x \in A$, $y \in B$ and $A \cap B = \emptyset$.*

Remark 21. *A Hausdorff space can be interpreted as being able to distinguish points using open sets.*

Definition 14 (Locally compact space)**.** *A topological space X is called locally compact if for all $x \in X$ there exists an open set A and a compact set K such that $x \in A \subseteq K$.*

Remark 22. *Compact spaces are locally compact as one can choose $A = K = X$.*

Definition 15 (Left-translate)**.** *Let G be a group. If $g \in G$ and $S \subseteq G$, we define the left-translate gS as*

$$gS := \{gh \mid h \in S\}.$$

Definition 16 (Regular measure)**.** *A measure is called regular if it can be approximated from above by measures of open sets and from below by measures of compact sets. Formally,*

$$\mu(B) = \inf\{\mu(A) \mid B \subseteq A, A \text{ open}\}$$

and

$$\mu(B) = \sup\{\mu(K) \mid K \subseteq B, K \text{ compact}\}.$$

Definition 17 (Left Haar measure)**.** *Suppose G is a locally compact Hausdorff topological group. Let μ be a measure on G equipped with the Borel σ-algebra such that $\mu(A) > 0$ for all non-empty open sets A. If $\mu(xB) = \mu(B)$ for all measurable B, then we call μ a left Haar measure* [12, p. 251].

Remark 23. *There exist analogous definitions for right-translates and right Haar measures. Everything works exactly analogously for them. The left Haar measures and right Haar measures coincide if the group is abelian, that is $xy = yx$ for all $x, y \in G$. Which one to use is merely convention and therefore all results and proofs regarding left Haar measures can be restated with minor modification to apply to right Haar measures.*

Remark 24. *The left-translate invariance should be compared with the translation invariance of the Lebesgue measure. It is immediately clear that the Lebesgue measure is both a left and a right Haar measure.*

Theorem 16 (Existence of the regular left Haar measure)**.** *For every locally compact Hausdorff topological group, there exists at least one regular left Haar measure* [12, p. 254].

Theorem 17 (Uniqueness up to constant of the regular left Haar measure)**.** *Let G be a locally compact Hausdorff topological group. Let μ and ν be two left Haar measures. Then there exists a $c > 0$ such that $\mu(B) = c\nu(B)$ for all measurable B.*

Remark 25. *The existence and uniqueness results should be compared to those of the Lebesgue measure. If one had a base set of known measure for general Haar measures, like the Lebesgue measure has with the unit cube, one could define it uniquely.*

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY