

In the mood for food: Markov-modulated models
for animal foraging

Lachlan Bridges

January 2019

Thesis submitted for the degree of

Master of Philosophy

in

Applied Mathematics

at The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Mathematical Sciences



THE UNIVERSITY
of ADELAIDE

Signed statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: Date:

Abstract

Early theoretical models of animal foraging determined that Lévy flights were an optimal search strategy in a number of different scenarios. However, a new family of strategies known as *intermittent* or *regime-switching* strategies have been found to provide a higher search efficiency. In this thesis, we investigate regime-switching strategies using Markov-modulated random walks. Our model allows a forager to have any number of different search strategies that it switches between according to some Markov chain. We derive an expression for the efficiency of a Markov-modulated random walk, and develop discrete approximations in order to solve our model numerically. We are able to show that many of the existing strategies investigated throughout the literature, such as giving-up time strategies, can be seen as a special case of a Markov-modulated random walk strategy. We are also able to approximate a search model with hidden-targets, where a forager can only locate targets while in a certain state. Using our new expression for the efficiency we recover some existing results from the literature as well as find some new results for the optimal search strategy under various circumstances. Finally, we outline a very simple two-dimensional model, in which food patches are distributed according to a homogeneous spatial Poisson process. We make some simplifying assumptions about the chance of finding food when backtracking, and find an upper bound on the efficiency of a search. We show that taking into account some backtracking makes the model too difficult to solve, and use simulations to investigate the accuracy of our model.

Acknowledgements

I must sincerely thank my supervisors, Giang Nguyen and Nigel Bean, for all of their help and guidance throughout the last two years of research, and for the numerous drafts they have read. Not only are they both excellent mathematicians, but patient teachers, and warm and friendly people. I always found great value in our meetings, as well as great enjoyment.

I would also like to give an extra thanks to Giang, for her encouragement and assistance over the last three and a half years. She has always gone above and beyond when assisting me, whether it be helping with applications, summer research, or encouraging me to present my work at conferences. Without her help I would have never began my masters, and would certainly not be where I am today.

It would be remiss of me not to thank my family and friends that have all supported me in various ways. In particular, I am grateful to my wife Fern, my sister Emily, and my brother Nick, as well as Ludovica, Cameron, Michael, and Ben.

Last and certainly not least, I would like to express the utmost gratitude to my parents, Heather and Geoff, for all of their love and support throughout my entire academic journey.

Contents

| | |
|--|------------|
| Signed statement | iii |
| Abstract | v |
| Acknowledgements | vii |
| List of Figures | xi |
| 1 Introduction | 1 |
| 2 Background | 7 |
| 2.1 Technical concepts | 7 |
| 2.1.1 Probability Theory | 7 |
| 2.1.2 Stochastic processes | 10 |
| 2.1.3 Point processes | 13 |
| 2.1.4 Linear operators and norms | 14 |
| 2.1.5 Phase-type distributions | 19 |
| 2.1.6 Technical terms in foraging | 20 |
| 2.2 Literature Review | 23 |
| 2.2.1 Early evidence for Lévy flight foraging strategies | 23 |
| 2.2.2 Extensions to the basic foraging model | 28 |
| 2.2.3 Composite search strategies for improved efficiency | 33 |
| 2.2.4 Statistical methods for analysing animal movement data | 36 |
| 3 One-dimensional foraging model | 41 |
| 3.1 General assumptions | 41 |
| 3.2 Random walk strategies | 44 |
| 3.2.1 Total cost of a random walk strategy | 44 |
| 3.2.2 Average total distance travelled | 54 |
| 3.2.3 Average number of steps | 56 |
| | ix |

| | | |
|----------|---|------------|
| 3.3 | Markov-modulated random walk strategies | 57 |
| 3.3.1 | Total cost of a Markov-modulated strategy | 58 |
| 3.3.2 | Average total distance travelled | 69 |
| 3.3.3 | Average number of steps | 71 |
| 3.4 | Analysis of the one-dimensional model assumptions | 71 |
| 3.4.1 | Effects of the deterministically-spaced targets assumption | 72 |
| 3.4.2 | Comparing the two different measures of efficiency | 74 |
| 4 | Discretisation of the one-dimensional search space | 79 |
| 4.1 | Notation and assumptions | 80 |
| 4.2 | Discretised expressions for a random walk strategy | 82 |
| 4.3 | Discretised expressions for a Markov-modulated random walk strategy | 87 |
| 5 | Results for the one-dimensional model | 95 |
| 5.1 | Special cases of the Markov-modulated random walk strategy . . | 95 |
| 5.1.1 | Unmodulated random walk | 95 |
| 5.1.2 | Giving-up-time forager | 101 |
| 5.1.3 | Vision switching forager | 109 |
| 5.2 | Optimal Markov-modulated search strategies for two-states, three states, and higher | 111 |
| 6 | Two-dimensional Foraging Models | 115 |
| 6.1 | General assumptions | 115 |
| 6.2 | Solving the model analytically | 117 |
| 6.2.1 | Zeroth-order approximation | 119 |
| 6.2.2 | First-order approximation | 123 |
| 6.3 | Simulations | 127 |
| 6.3.1 | Monte Carlo simulations for the first-order area | 127 |
| 6.3.2 | Simulating our two-dimensional model | 130 |
| 7 | Conclusion | 133 |
| 7.1 | Future Work | 137 |
| A | Calculating $h(x)$ and A for some step-length distributions | 141 |
| A.1 | Average step length | 141 |
| A.1.1 | Unbounded power-law distribution | 144 |
| A.1.2 | Bounded power-law distribution | 148 |
| A.1.3 | Unbounded exponential distribution | 150 |
| A.1.4 | Bounded exponential distribution | 151 |

| | | |
|----------|---|------------|
| A.2 | Discretized probability distribution matrix | 153 |
| A.2.1 | Unbounded power-law distribution | 153 |
| A.2.2 | Bounded power-law distribution | 153 |
| A.2.3 | Unbounded exponential distribution | 154 |
| A.2.4 | Bounded exponential distribution | 154 |
| B | Area covered in the first-order model | 155 |
| | Bibliography | 165 |

List of Figures

| | | |
|-------|--|-----|
| 2.1.1 | Comparison of different choices of parameter for a random walk with power-law distributed steps | 22 |
| 3.1.1 | The full search space for the one-dimensional model | 41 |
| 3.1.2 | The reduced search space for the one-dimensional model | 43 |
| 3.4.1 | Mean efficiency of an unbounded power-law strategy for three different target distributions for a different numbers of repetitions and targets to be found | 73 |
| 3.4.2 | Two different definitions of the efficiency, η and η_N , compared to each other for non-destructive foraging | 76 |
| 3.4.3 | Two different definitions of the efficiency, η and η_N , compared to each other for non-destructive foraging | 77 |
| 5.1.1 | Search efficiency of an unbounded power-law strategy across a range of different discretisation sizes | 97 |
| 5.1.2 | Search efficiency of a bounded power-law strategy across a range of different discretisation sizes | 98 |
| 5.1.3 | Search efficiency of an unbounded exponential strategy across a range of different discretisation sizes | 98 |
| 5.1.4 | Search efficiency of a bounded exponential strategy across a range of different discretisation sizes | 99 |
| 5.1.5 | Comparison of the search efficiency for the four different step-length distributions | 100 |
| 5.1.6 | Effect of bound size on the search efficiency for a bounded power-law distribution for destructive foraging | 100 |
| 5.1.7 | Effect of bound size on the search efficiency for a bounded exponential distribution for destructive foraging | 101 |
| 5.1.8 | Efficiency vs geometric giving-up time parameter for a Brownian search that gives up and follows a ballistic path | 104 |

| | | |
|--------|---|-----|
| 5.1.9 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.01$, for non-destructive foraging . . | 106 |
| 5.1.10 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.1$, for non-destructive foraging . . | 106 |
| 5.1.11 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.5$, for non-destructive foraging . . | 107 |
| 5.1.12 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.9$, for non-destructive foraging . . | 107 |
| 5.1.13 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.99$, for non-destructive foraging . | 108 |
| 5.1.14 | Efficiency of two-state Markov-modulated power-law strategy, with giving-up parameter $p = 0.01$, for destructive foraging . . . | 108 |
| 5.1.15 | Efficiency of two-state Markov-modulated power-law strategy, with switching radius of vision | 111 |
| 6.2.1 | Area covered by a step in the zeroth-order model | 120 |
| 6.2.2 | Area covered by a step in the zeroth-order model, simplified into a rectangle | 120 |
| 6.2.3 | The four different cases for the turning-angle that we consider when determining the area of overlap | 124 |
| 6.3.1 | An example step where the overlapping area is approximated using a Monte Carlo simulation, and compared with our analytic expressions | 128 |
| 6.3.2 | The relative error in our analytic expressions for the overlap plotted against the step-length and the turning-angle | 129 |
| 6.3.3 | The relative error in our analytic expressions for the overlap plotted against the turning-angle | 129 |
| 6.3.4 | An example simulation of our two-dimensional | 131 |
| 6.3.5 | Comparison of zeroth-order, first-order, and infinite-order model, and the analytic results for the zeroth-order | 132 |
| B.0.1 | The four different cases for the turning-angle that we consider when determining the area of overlap | 155 |
| B.0.2 | Case 1: A large turning angle and large step | 156 |
| B.0.3 | There is no clear threshold between case 1 and case 2, although we assume that there is since the introduced error is only small . | 157 |
| B.0.4 | A simplification that must be made in order to find the area for case 1 | 158 |
| B.0.5 | Case 2: A large turning-angle and a small step | 159 |

| | | |
|-------|--|-----|
| B.0.6 | Case 3: A medium turning-angle | 161 |
| B.0.7 | Case 4: A small turning-angle | 162 |

Dedicated to Fern.

A loving heart is the beginning of all knowledge.

Thomas Carlyle

Chapter 1

Introduction

The study of animal movement is an incredibly broad topic, spanning multiple disciplines, with many unanswered questions. Results from studying the movement behaviour of animals have found applications in a number of different areas, and not always in fields related to animals.

Animals move for a seemingly endless number of reasons. Usually, they move in order to find food to eat, or a partner to mate with, but other reasons exist. Sheep begin running when they see other sheep running. They may also move in order to escape the clutches of a pursuing predator. Some animals move for fun, like dolphins playing games together. Regardless of the true reasons for why an animal moves the way it does, which we may never know, the reproductive fitness of an animal is certainly affected by how it decides to move.

Natural selection should imply that animals evolve to move in a way that improves their reproductive fitness, whether they know it or not. This is the key reasoning that optimal foraging theory is built upon. Since the ability to forage for food is a good proxy for an animal's overall reproductive fitness, it is hypothesised that the optimal foraging strategy, in theory, should be the strategy used in practice.

Traditionally, animal foraging theory has been a branch of ecology, and has investigated the effect of qualitative decision rules, such as the choice of diet [65]. More recently, animal foraging theory has been investigated using a more quantitative approach, generally employing stochastic processes to model animal movement. Stochastic optimal foraging theory, as it is sometimes referred to, is the field of study that investigates the optimal search strategy for a general forager which lacks complete information about its current environment, and does not have any memory capabilities. Stochastic optimal foraging theory complements the study of ordinary optimal foraging theory, in which the forager makes decisions

based on complete information about its environment [6]. In reality, animals fall in between the two, having some, though not complete, information about the environment.

In its most general sense, the study of animal foraging is interested in how different movement patterns can affect the frequency of biological encounters. An animal searching for food should aim to maximise the number of biological encounters, whereas an animal fleeing a predator should aim to minimise the number of encounters. For this reason, many of the models built with animals in mind are still valid for other things such as micro-organisms [12], genetic material [40], and even swarm robots [75]. Because of this, results originally found when investigating animals have been used to investigate how pollen diffuses [79], to program swarm robots to move efficiently [75], to design fisheries [74], and have even been used in medicine [23].

One of the earliest papers in stochastic optimal foraging theory was published in 1995 [24], and showed that the distance of relocations for a type of fruitfly had a heavy-tailed distribution. There is some discrepancy over the exact definition of a heavy-tailed distribution, although in the animal foraging literature it is consistently used to mean not having finite variance, and we will use this definition going forward. This means that there is a reasonable probability of very large values occurring. In the context of the distance of animals relocating, this means that there are some instances of animals making very large relocations. Search strategies of this kind are known as Lévy flights, and have been an important area of research throughout the topic of animal foraging.

There is a common analogy often used [88] to give an intuitive explanation as to why Lévy flight searches may be more efficient. Anyone that has ever lost something in their house has likely performed a search that resembles a Lévy flight. Imagine you are about to leave the house but you cannot find your car keys. You would search the room you are in, making small relocations around the room, while intensively searching in various places. After some time, you may realise that you left your keys in the kitchen. You run to the kitchen, which is a large relocation, before beginning your intensive search again. This scenario somewhat resembles a Lévy flight, with large steps occurring each time you change rooms, making the distribution of steps heavy-tailed. A random walk with light-tailed distribution of steps would instead take small steps throughout the whole search, including as you change rooms, meaning you would spend as much time searching the hallway between rooms as the rooms themselves. This type of search would correspond to something known as Brownian motion.

Considering the above example, the efficiency of a Lévy flight over a Brownian motion seems fairly intuitive, not just for finding car keys, but for searching in general. This leads into one of the big advantages of stochastic optimal foraging theory over traditional foraging theory. By considering a general forager that is fully unaware of its surrounding, we are able to find results that may apply to any animal, or even any non-animal searcher, rather than finding results that only apply to a specific environment. For example, the Lévy flight may be an inherently better search strategy than a Brownian motion in many different scenarios.

A paper in 1999 [86] devised a theoretical model to investigate the optimal search strategy for a forager. The strategies considered were random walks with steps drawn from a power-law distribution. Depending on the parameter μ of the power-law distribution, the steps may or may not be heavy-tailed, meaning qualitatively different strategies may occur with simply a change in the value of μ . Thanks to this property, the power-law distribution has been used very frequently throughout the stochastic optimal foraging theory literature. The theoretical model allowed for two different scenarios: destructive and non-destructive foraging, under which the food once located is, or is not, destroyed. The optimal strategy for destructive foraging was found to occur when $\mu \rightarrow 1$, which corresponds to the forager selecting a very large step-size, meaning it will travel in a straight line, with no changes in direction until it reaches food. For non-destructive foraging, the optimal strategy was to use $\mu \approx 2$, which corresponds to a Lévy flight search strategy.

After the first two papers that found empirical evidence for Lévy flights in animal movements [24, 85], many other papers were published, mostly using the same methodology, but investigating different animals. Evidence of Lévy flights was found for almost all of the animals investigated, included spider-monkeys, fish, and even human hunter-gatherers. However, various papers ([11, 30], etc) have disputed the empirical evidence of Lévy flights, criticizing some of the methods used in other papers, as well as revisiting earlier results using larger datasets.

Many other theoretical papers also built upon the original theoretical model [86]. Various extensions looked at including foraging for targets that are moving [3], foraging for targets in higher dimensions [78], foraging for targets that are hidden while the forager is moving too fast [14], and targets with a delay before they can be revisited again [63].

More recently, a new type of strategy has been investigated, referred to as a giving-up time strategy, which involves switching between two different search

modes. Initially a forager begins performing an intensive search according to a Brownian motion before giving up and performing ballistic motion, which corresponds to walking in a straight line, until food is located. These search strategies have been shown to outperform any kind of non-switching strategy, including the Lévy flight strategy, which was once thought to be the theoretical optimal. Various results have been found for these strategies, such as the optimal giving up time [61]. Others have also investigated a fixed giving-up time and found the optimal strategy to use after giving up, rather than just a ballistic motion. There is, as yet, no paper that has optimised over both the extensive strategy and the giving-up time.

For a long period of time, most of the stochastic foraging literature revolved around Lévy flights, since these were thought to be optimal. Once giving-up time and other switching strategies were considered, the focus then turned to investigating these. However, currently only strategies that switch between two different strategies have been looked into, with no investigation into switching between three or more modes. Generalising the switching strategies to more than two modes is a primary aim of this thesis.

We generalise the idea of a switching search strategy by considering Markov-modulated random walk strategies. A Markov-modulated random walk is a random walk, with the distribution of the steps of the random walk changing depending on the current state of some Markov chain. This effectively means that the forager switches between multiple different random walk search strategies, according to a Markov chain. Determining an analytic expression for the efficiency of a Markov-modulated random walk strategy, solving it numerically, and investigating the results forms the bulk of this thesis, from Chapter 3, Chapter 4, Chapter 5.

In Chapter 2, we discuss some of the background required to understand the rest of the thesis. We outline some useful theorems that we will need, along with the definitions and results required to understand them. We outline some of the key terms used in the literature, and discuss other variations on them, in order to keep consistency when discussing the literature review. Chapter 2 concludes with a literature review, in which we discuss some of the important papers in stochastic foraging, as well as papers that devise extensions to the basic model. We discuss and interpret some of the key results found, including the optimal strategies for various models, as well as the conclusions of the composite strategies.

Chapter 3 begins with an investigation into a somewhat recent model for animal foraging in one dimension by Bartumeus *et al.* [6]. The authors were able to

derive an analytic expression for the efficiency of a forager searching using a random walk with any choice of step-length distribution. We prove some key results from this paper for which proofs were not included. We then extend the model to allow for Markov-modulated random walk search strategies, with any number of states. This means we allow for a forager to switch between any number of different search strategies according to a discrete-time Markov chain. We derive an analytic expression for the efficiency of any Markov-modulated random walk strategy, as well as the expected number of steps in each state of a strategy. The results of this chapter may be considered the most important results of this thesis.

In Chapter 4 we take our model from Chapter 3 and discretise the search space, allowing us to find numerical solutions. We derive discretised approximations of our results from Chapter 4. This chapter also helps to further demonstrate some of the difficulties found throughout our derivation of the Markov-modulated random walk results in Chapter 3.

In Chapter 5 we make use of the results found throughout Chapter 3, Chapter 4 to recover some results from the literature as well as find some new results. We discuss how various strategies throughout the literature can be considered as a specific case of a Markov-modulated random walk, with certain parameters. We show specifically how we can recover unmodulated random walk strategies, giving-up time strategies, some vision-switching strategies, and show that our results match existing results. Finally, we optimise over all of the parameters of our Markov-modulated random walk, finding new results for optimal search strategies.

Chapter 6 is mostly independent of the previous chapters, in which we devise a simple two-dimensional model in order to obtain some analytic results for the optimal search strategy. We consider food targets that are distributed according to a spatial Poisson process, and thus the probability of finding food is proportional to the amount of space that a forager searches. We initially consider a model in which targets may be found upon reexploring a previously explored area, which we refer to as the zeroth-order approximation. We discuss the first-order model, in which area that was searched more than one step ago may contain targets. Analytic results for the first-order model require a discussion of the geometry of the overlap between consecutive steps, and this eventually results in a model that is too hard to solve. We perform numerical simulations of the zeroth-order, first-order, and infinite-order model, which is the exact model, and are able to show that the first-order model does not provide a very accurate approximation

for the infinite-order model anyway. Thus, this chapter provides some basic results for two-dimensional models, while also showing some of the difficulties that arise when trying to obtain analytic results in two dimensions.

Finally, in Chapter 7, we discuss the significance of our results, and examine which results match and which contradict the existing results. We also discuss some possible avenues of future research that have been opened up thanks to this research, as well as possible extensions to our models.

Chapter 2

Background

■ 2.1 Technical concepts

In this section we review the technical concepts needed for this thesis. We do not attempt to give detailed summaries of these topics, but rather present some of the key theorems that will be used along with some basic definitions and results required to understand these theorems.

■ 2.1.1 Probability Theory

Although this thesis, for the most part, does not require a measure-theoretic formalisation of probability theory, we require knowledge of what both a measure space and a σ -finite measure space is, in order to understand some of the key theorems we shall use. This motivates us to discuss some of the basic definitions from measure-theoretic probability.

Definition 2.1.1. If Ω is a given set, then a σ -field \mathcal{F} on Ω is a collection of subsets satisfying the following conditions:

1. $\Omega \in \mathcal{F}$,
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,
3. If a countable sequence $A_1, A_2, \dots \in \mathcal{F}$ then

$$\bigcup_{i \geq 1} A_i \in \mathcal{F}.$$

We call the pair (Ω, \mathcal{F}) a *measurable space*.

Definition 2.1.2. Given two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, a function $X : \Omega_1 \rightarrow \Omega_2$ is said to be *measurable* if for every $A \in \mathcal{F}_2$ we have $X^{-1}(A) \in \mathcal{F}_1$.

Theorem 2.1.3. *If f and g are both measurable functions, then the product of these functions, fg , is also measurable.*

Definition 2.1.4. Let Ω be a sample space and \mathcal{F} be a σ -algebra on Ω . A function $\mu : \mathcal{F} \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ is called a *measure* if it satisfies the following properties

1. $\mu(A) \geq 0$, for all A in \mathcal{F} ,
2. $\mu(\emptyset) = 0$,
3. For all countable collections $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in \mathcal{F} ,

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

A natural extension to a measure is the probability measure.

Definition 2.1.5. A measure on a sample space Ω is called a *probability measure* if the total measure is one, that is $\mu(\Omega) = 1$.

Definition 2.1.6. Let μ be a measure on the measurable space (Ω, \mathcal{F}) . Then we say that $(\Omega, \mathcal{F}, \mu)$ is a *measure space*.

Definition 2.1.7. A measure μ on a measurable space (Ω, \mathcal{F}) is called a *finite measure* if it satisfies

$$\mu(A) < \infty,$$

for all $A \in \mathcal{F}$.

Definition 2.1.8. Let μ be a measure on the measurable space (Ω, \mathcal{F}) . We say that μ is a *σ -finite measure* if the set Ω is the countable union of sets with finite measure.

For example, the Lebesgue measure on the real numbers is a σ -finite measure but not a finite measure. To see this, consider an interval $[k, k+1)$ which has a measure of 1, since the Lebesgue measure for an interval is defined as its width, $\mu([a, b]) = b - a$. The real numbers can be constructed by a countable union of intervals of this form, and so by Definition 2.1.8 is σ -finite. However, the Lebesgue measure of the real numbers, $\mu((-\infty, \infty))$ is not finite.

Another example we can consider is the counting measure, which for any finite set is the number of elements in that set. Clearly, this measure is neither finite nor σ -finite for \mathbb{R} . However, the set \mathbb{N} is countable and the counting measure is σ -finite.

Definition 2.1.9. If μ is a σ -finite measure, then we say that $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space.

Finally, we present a theorem that will be used frequently throughout the thesis, which uses the definitions presented above.

Theorem 2.1.10 (Tonelli's Theorem). *If (X, \mathcal{A}, μ) and (Y, \mathcal{B}, ν) are σ -finite measure spaces and $f : X \times Y \rightarrow [0, \infty)$ is non-negative and measurable, where $X \times Y$ is given the product measure (which is unique since X and Y are σ -finite), then*

$$\int_X \left(\int_Y f(x, y) \nu(dy) \right) \mu(dx) = \int_Y \left(\int_X f(x, y) \mu(dx) \right) \nu(dy) = \int_{X \times Y} f d\mu \times \nu.$$

Theorem 2.1.10 allows us to rearrange the order of integration or summation under certain circumstances. Using some of the above definitions, we can show some specific circumstances in which Tonelli's theorem will hold, in order to avoid discussion of measure-theory throughout the thesis.

Corollary 2.1.11. *If $g(x, y)$ and $h(x, y)$ are two non-negative and measurable functions on $X \times Y$ where $X = [a, b]$ and $Y = [c, d]$ for some $a, b, c, d \in \mathbb{R}$, then*

$$\int_a^b \int_c^d g(x, y) h(x, y) dy dx = \int_c^d \int_a^b g(x, y) h(x, y) dx dy.$$

Proof. Both $g(x, y)$ and $h(x, y)$ are non-negative and measurable and so if we define $f(x, y) = g(x, y)h(x, y)$, then $f(x, y)$ will be a non-negative and measurable function, according to Theorem 2.1.3. Both (X, \mathcal{B}_X) and (Y, \mathcal{B}_Y) are measurable spaces, where \mathcal{B}_X and \mathcal{B}_Y are the Borel σ -algebras generated by $[a, b]$ and $[c, d]$ respectively. Then, we can construct (X, \mathcal{B}_X, μ) and (Y, \mathcal{B}_Y, ν) where μ is the Lebesgue measure, which are both σ -finite measure spaces. Therefore, we can apply Theorem 2.1.10 to arrive at the desired result. \square

Note that a probability density function (PDF) by definition will be a non-negative and measurable function, so if $g(x, y)$ and $h(x, y)$ are both PDFs they must both satisfy the conditions of Corollary 2.1.11.

Corollary 2.1.12. *If $g(x, y)$ and $h(x, y)$ are two non-negative and measurable functions on $X \times Y$ where X and Y are both subsets of the natural numbers, then*

$$\sum_X \sum_Y g(x, y)h(x, y) = \sum_Y \sum_X g(x, y)h(x, y).$$

Proof. The proof is exactly the same as that of Corollary 2.1.11, but we instead define μ as the counting measure. \square

For example, using Corollary 2.1.12 we could get the following

$$\sum_{x=0}^{\infty} \sum_{y=0}^n g(x, y)h(x, y) = \sum_{y=0}^n \sum_{x=0}^{\infty} g(x, y)h(x, y).$$

Corollary 2.1.13. *If $g(x, y)$ and $h(x, y)$ are two non-negative and measurable functions on $X \times Y$ where $X = [a, b]$ for $a, b \in \mathbb{R}$ and Y is a subset of the natural numbers, then*

$$\sum_Y \int_a^b g(x, y)h(x, y)dx = \int_a^b \sum_Y g(x, y)h(x, y)dx.$$

This corollary is simply a combination of Corollary 2.1.11, Corollary 2.1.12, allowing us to switch the the ordering of integrals and summations.

Note that Theorem 2.1.10 can be applied multiple times in order to rearrange an expression on more than two measure spaces, as long as the same conditions are met such as the functions all being non-negative and measurable. This also means that Corollary 2.1.11, Corollary 2.1.12, Corollary 2.1.13 can be applied to switch the order of integration and summation of an arbitrary number of non-negative and measurable functions.

■ 2.1.2 Stochastic processes

Now, we discuss some basic stochastic processes that will be used throughout this thesis, as well as some of their properties. Most important is the random walk.

Definition 2.1.14. Let $\{X_k\}_{k=1}^{\infty}$ be a sequence of independent, identically distributed random variables. For every positive n , we denote

$$S_n = \sum_{k=1}^n X_k.$$

The sequence $\{S_n\}_{n=1}^{\infty}$ is called a *random walk*.

In other words, a random walk is a stochastic process describing a path of successive random steps on some mathematical space. In Section 3.2 we use random walks to represent the movement of an animal, with the distribution of X_k affecting the strategy that is used. Although the random walk is a discrete-time process, it has a strong relationship to some continuous-time processes, such as the Wiener process, which is also referred to as Brownian motion.

Definition 2.1.15. A Brownian motion $\mathcal{B} = \{B_t\}_{t \geq 0}$ with variance σ^2 is a stochastic process that satisfies the following properties:

1. $B_0 = 0$ almost surely.
2. \mathcal{B} has independent increments: for $0 \leq t_1 < \dots < t_k < \infty$ and $x_1, x_2, \dots, x_k \in \mathbb{R}$,

$$\begin{aligned} & \Pr(B_{t_2} - B_{t_1} \leq x_1, B_{t_3} - B_{t_2} \leq x_2, \dots, B_{t_k} - B_{t_{k-1}} \leq x_{k-1}) \\ &= \prod_{2 \leq i \leq k} \Pr(B_{t_i} - B_{t_{i-1}} \leq x_{i-1}). \end{aligned}$$

3. For every $0 \leq s \leq t$, $B_t - B_s \sim N(0, \sigma^2(t - s))$ — stationary, Gaussian increments with variance $\sigma^2(t - s)$.
4. \mathcal{B} has continuous sample paths, almost surely.

We use $\mathcal{B}(\mu, \sigma^2)$ to denote a Brownian motion with drift μ and variance σ^2 , where the drift represents the expected increase or decrease per unit time. A Brownian motion with a drift of 0 and a variance of 1, denoted $\mathcal{B}(0, 1)$, is known as a *standard Brownian motion*.

Recall the Central Limit Theorem (CLT), which gives conditions under which a sequence of properly normalised random variables will converge to a normal distribution.

Theorem 2.1.16 (Central Limit Theorem). *If $\mathbb{E}[X_1] = \mu$, $\text{Var}[X_1] = \sigma^2 \in (0, \infty)$, and $S_n = X_1 + \dots + X_n$, with X_i i.i.d., then*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow N(0, 1) \text{ as } n \rightarrow \infty$$

in distribution.

Note that the CLT requires a finite variance, and so will not apply to heavy-tailed distributions. We now introduce a functional extension of the CLT, which can be used to investigate the continuous-time limit of a random walk.

Theorem 2.1.17 (Donsker's Theorem). *Consider a random walk, S_n , with i.i.d. steps X_i such that $\mathbb{E}[X_i] = 0$ and $\text{Var}[X_i] = 1$. We define*

$$W^{(n)}(t) := \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}. \quad (2.1.1)$$

Then, Donsker's invariance principle [26] states that $W^{(n)}(1) \rightarrow \mathcal{B}$ in distribution as $n \rightarrow \infty$, where \mathcal{B} denotes a standard Brownian motion.

Now, although an expectation of 0 and variance of 1 is required for this to apply, in general we can rescale our random variable as long as these are not infinite. Thus, just like the CLT, Donsker's Theorem is not valid for heavy-tailed distributions. In the following example, we consider an unbounded power-law distribution and show at what values Donsker's Theorem will apply, and how we can rescale to make the variance 1.

Example 2.1.1. Let X be a random variable with unbounded power-law distribution centred about 0, which has density function

$$p(x) = \frac{\mu - 1}{2x_{\min}} \left(\frac{|x|}{x_{\min}} \right)^{-\mu}, \quad |x| \in [x_{\min}, \infty), \quad x_{\min} > 0, \quad \mu > 1.$$

The m th moment of X is given by

$$\begin{aligned} \mathbb{E}[X^m] &= \int_{-\infty}^{-x_{\min}} x^m p(x) dx + \int_{x_{\min}}^{\infty} x^m p(x) dx \\ &= \frac{\mu - 1}{2} x_{\min}^{\mu-1} \left(\int_{-\infty}^{-x_{\min}} x^m (-x)^{-\mu} dx + \int_{x_{\min}}^{\infty} x^{-\mu+m} dx \right). \end{aligned}$$

Then, if $m < \mu - 1$, we get

$$\begin{aligned} \mathbb{E}[X^m] &= \frac{\mu - 1}{2} x_{\min}^{\mu-1} \left(\left[\frac{x^{m+1}(-x)^{-\mu}}{-\mu + m + 1} \right]_{-\infty}^{-x_{\min}} + \left[\frac{x^{-\mu+m+1}}{-\mu + m + 1} \right]_{x_{\min}}^{\infty} \right) \\ &= \frac{\mu - 1}{2(\mu - 1 - m)} (x_{\min}^m + (-x_{\min})^m). \end{aligned}$$

For $m \geq \mu - 1$, the m th moment is undefined. Based on the above, the first moment is $\mathbb{E}[X] = 0$ for all $\mu \geq 2$. The second moment, and hence the variance, will be

$$\text{Var}[X] = \mathbb{E}[X^2] = \begin{cases} \frac{\mu - 1}{\mu - 3} x_{\min}^2 & \text{for } \mu > 3, \\ \text{undefined} & \text{for } \mu \leq 3. \end{cases}$$

Since $\mathbb{E}[X] = 0$, we do not need to shift X , although we do need to scale X to account for $\text{Var}[X] \neq 1$. For $\mu > 3$, we define

$$Y := \sqrt{\frac{\mu - 3}{\mu - 1}} x_{\min}^{-2} X,$$

so $\mathbb{E}[Y] = 0$ and $\text{Var}[Y] = 1$. Therefore, using Donsker's Theorem, a random walk $S_n^{(y)}$ with i.i.d steps, Y_i , will converge in distribution to a standard Brownian motion.

Thus, we can also say that the random walk $S_n^{(x)}$ with i.i.d steps, X_i , will converge in distribution to the Brownian motion $\mathcal{B}\left(0, \frac{\mu-1}{\mu-3}x_{\min}^2\right)$ for $\mu > 3$. Similarly, when $\mu = 3$, $S_n^{(x)}$ will converge in distribution to the Brownian motion $\mathcal{B}\left(0, 2\log(x_{\min})x_{\min}^2\right)$.

As we have just shown in Example 2.1.1, the unbounded power-law distribution converges to a Brownian motion for $\mu \geq 3$, but does not for $1 < \mu < 3$ since the distribution will be heavy-tailed.

■ 2.1.3 Point processes

In Chapter 6, we investigate a two-dimensional model where food targets are randomly distributed over a two-dimensional space. Point processes are collections of points on some mathematical space, and are often used to model physical objects that can be represented as a point in space. Thus, point processes are an obvious choice to represent our food targets. Further, we use the simplest and most ubiquitous point process, which is the Poisson point processes.

Definition 2.1.18. A *Poisson point process* is a collection of points randomly located on some underlying mathematical space, where the number of points in one subregion is independent of the number of points in any other disjoint subregion. We call this property *complete randomness*.

For a Poisson point process, the number of points, N , in a bounded region is given by a Poisson distribution:

$$\Pr(N = n) = \frac{\Lambda^n}{n!} e^{-\Lambda},$$

where the parameter Λ is determined by the region. The parameter Λ is also the expected value of N .

A further simplification we can make is to consider homogeneous Poisson point processes.

Definition 2.1.19. If a Poisson point process has a parameter of the form $\Lambda = \nu\lambda$, where ν is Lebesgue measure, and λ is a constant, then we say that the Poisson point process is *homogeneous* or *stationary*.

Since Lebesgue measure assigns length, area, or volume to a set, then λ can be considered the mean number of points per some unit, usually length, area, or volume, depending on the underlying mathematical space.

For our purposes, we only consider Poisson point processes that are two-dimensional.

Definition 2.1.20. A *spatial Poisson point process*, \mathcal{N} , is a Poisson point process defined in the plane \mathbb{R}^2 .

■ 2.1.4 Linear operators and norms

Throughout Chapter 3, we use linear operators to deal with each step that a forager takes. We also need to prove some properties about operators, such as their norm. This motivates us to give a brief summary of linear operators and norms.

Definition 2.1.21. Given a vector space V , a *norm* on V is a non-negative scalar function $f : V \rightarrow [0, \infty)$ with the properties: For all $a \in \mathbb{R}$ and all $\mathbf{u}, \mathbf{v} \in V$,

1. $f(\mathbf{u} + \mathbf{v}) \leq f(\mathbf{u}) + f(\mathbf{v})$,
2. $f(a\mathbf{v}) = |a|f(\mathbf{v})$,
3. If $f(\mathbf{v}) = 0$ then $\mathbf{v} = \mathbf{0}$.

A norm assigns some positive length or size to vectors that are in a vector space, except for the zero vector which is assigned zero. Some of the most widely used norms are the 1-norm, 2-norm, and ∞ -norm, which are all cases of the p -norm.

Definition 2.1.22. The p -norm, denoted $\|\cdot\|_p$ of a vector $\mathbf{v} \in V$ is defined as

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

for $p \geq 1$, $\mathbf{x} \in \mathbb{R}^n$, $V = \mathbb{R}^p$.

Then, the 1-norm is

$$\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|,$$

the 2-norm is

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2},$$

and the ∞ -norm is

$$\|\mathbf{x}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = \max_{1 \leq i \leq n} |x_i|.$$

We can also talk about the norm of a matrix rather than a vector.

Definition 2.1.23. Given a real matrix A , a matrix norm $\|A\|$ is a non-negative number associated with A having the properties

1. $\|A\| \geq 0$ with equality iff $A = 0$,
2. $\|kA\| = |k|\|A\|$ for any scalar k ,
3. $\|A + B\| \leq \|A\| + \|B\|$,

As with the vector norm, there exists a matrix p -norm.

Definition 2.1.24. The p -norm of a matrix, denoted $\|\cdot\|_p$, is defined as

$$\|\mathbf{x}\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p,$$

for $p \geq 1$, and where $\|\mathbf{x}\|_p$ is the vector p -norm.

The 1-norm is the maximum absolute column sum,

$$\|A\|_1 = \max_{1 \leq j \leq m} \sum_{i=1}^n |a_{ij}|,$$

where A has dimensions $n \times m$, and the ∞ -norm is the maximum absolute row sum,

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^m |a_{ij}|.$$

A natural generalisation of the p -norm for finite-dimensional vector spaces is the L^p -norm for function spaces.

Definition 2.1.25. On a measure space (X, \mathcal{F}, μ) , the L^p -norm of a function $f : X \rightarrow \mathbb{R}$ is

$$\|f\|_{L^p} = \left(\int_X |f|^p d\mu \right)^{1/p}.$$

Definition 2.1.26. Let V and W be vector spaces. A function $f : V \rightarrow W$ is said to be a *linear map* (or linear transformation) if for any two vectors $\mathbf{u}, \mathbf{v} \in V$ and any scalar $c \in \mathbb{R}$ the following conditions are satisfied:

1. $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$,
2. $f(c\mathbf{u}) = cf(\mathbf{u})$.

For example, an integral over some interval I is a linear map from the space of all real-valued integrable functions on I to \mathbb{R} . The expectation of a random variable is also a linear map, but the variance of a random variable is not since it is not linear.

Definition 2.1.27. A *linear operator* is a linear map as defined in Definition 2.1.26, with $V = W$.

Definition 2.1.28. A *bounded linear operator* is a linear operator $T : X \rightarrow Y$ for which there exists an $M \geq 0$ such that

$$\frac{\|Tv\|_Y}{\|v\|_X} \leq M < \infty. \quad (2.1.2)$$

Definition 2.1.29. A Neumann series is a series of the form

$$\sum_{k=0}^{\infty} T^k,$$

where T is an operator and T^k represents k consecutive operations of T . Under certain conditions, the convergence of a Neumann series is guaranteed. We discuss this further in Theorem 2.1.31

We can also consider the norm of an operator.

Definition 2.1.30. The operator norm of a linear operator $T : V \rightarrow W$ is given by

$$\|T\|_{op} = \sup \left\{ \frac{\|Tv\|}{\|v\|} : v \in V \text{ with } v \neq 0 \right\},$$

where the numerator's norm is on W and the denominator's norm is on V .

There are other equivalent definitions of the operator norm, such as

$$\|T\|_{op} = \inf \{c \geq 0 : \|Tv\| \leq c\|v\| \text{ for all } v \in V\},$$

although we will use the Definition 2.1.30 throughout this thesis. Note that each pair of norms on V and W will induce an operator norm.

Intuitively, the operator norm tells us the “size” of T by how much it “stretches” a vector in the “biggest” case. When V and W are function spaces case we generally use L^p -norms in Definition 2.1.30. If the operator norm is less than 1, it will always shrink the vector it is acting on. Thus, it makes sense that the operator norm will tell us about the convergence of a Neumann series.

Theorem 2.1.31. *If V is a Banach space, and the linear operator $T : V \rightarrow W$ has $\|T\|_{op} < 1$, then convergence of the Neumann series $\sum_{k=0}^{\infty} T^k$ is guaranteed [83].*

To make sense of this theorem, we first need to know what a Banach space is.

Definition 2.1.32. A Banach space is a complete normed vector space.

That is, a Banach space is a vector space with a norm defined on it, and is complete in the sense that a Cauchy sequence of vectors always converges to a well-defined limit in the space. For example, \mathbb{R}^n together with the p -norm is a Banach space. Similarly, the Riesz–Fischer theorem [28] tells us that L^p spaces are complete. For the entirety of our thesis we will be considering spaces that are known to be Banach spaces, and so only need to show that the operator norm is less than 1 to establish convergence.

Establishing the convergence of a Neumann series is important throughout this thesis, since it will allow us to write our expressions without an infinite summation, as the following theorem demonstrates.

Theorem 2.1.33. *Suppose that T is a bounded linear operator on the normed vector space V . If the Neumann series for T converges in the operator norm, then $\text{Id} - T$ is invertible and its inverse is the series*

$$(\text{Id} - T)^{-1} = \sum_{k=0}^{\infty} T^k,$$

where Id is the identity operator in V .

For a simple, and non-rigorous explanation of why this theorem is true, we can consider the partial sums given by

$$S_n := \sum_{k=0}^n T^k.$$

If T converges in the operator norm, then

$$\begin{aligned} \lim_{n \rightarrow \infty} (\text{Id} - T)S_n &= \lim_{n \rightarrow \infty} \left(\sum_{k=0}^n T^k - \sum_{k=0}^n T^{k+1} \right) \\ &= \lim_{n \rightarrow \infty} (\text{Id} - T^{n+1}) \\ &= \text{Id}, \end{aligned}$$

where we used the fact that $\lim_{n \rightarrow \infty} T^{n+1} = 0$ if $\|T\|_{op} < 1$. Rearranging gives us

$$\lim_{n \rightarrow \infty} S_n = (\text{Id} - T)^{-1}.$$

We now present a theorem about the operator norm of a composition of operators.

Theorem 2.1.34. *If V , W , and X are three normed spaces and $A : V \rightarrow W$ and $B : W \rightarrow X$ are bounded operators, then*

$$\|BA\|_{op} \leq \|B\|_{op}\|A\|_{op}.$$

A concept that we also use is the adjoint of an operator.

Definition 2.1.35. For a linear operator $A : V \rightarrow W$, the *adjoint* operator, denoted $A^* : W \rightarrow V$, is an operator that satisfies

$$\langle Av, w \rangle = \langle v, A^*w \rangle,$$

for all $v \in V$ and $w \in W$, where V and W are vector spaces with inner products, and $\langle \cdot, \cdot \rangle$ denotes the inner product.

The following theorem follows from the uniqueness of the adjoint operator.

Theorem 2.1.36. For a linear operator $A : V \rightarrow W$ with adjoint $A^* : W \rightarrow V$ and $B : W \rightarrow X$ with adjoint $B^* : X \rightarrow W$,

$$(AB)^* = B^*A^*.$$

Definition 2.1.37. A linear operator A is *self-adjoint* if $A = A^*$.

Throughout this thesis, the operators that we consider are either matrices or integral operators, and so we discuss some relevant properties of these now.

Theorem 2.1.38. The adjoint of a real-valued matrix A is given by its matrix transpose, A^\top .

Corollary 2.1.39. A symmetric matrix is self-adjoint.

The integral operators that we use throughout this thesis take the form of a specific integral equation, known as a Fredholm integral.

Definition 2.1.40. A Fredholm operator, A , takes the form

$$[Af](t) = \int_a^b K(t, s)f(s)ds, \quad (2.1.3)$$

where $K(t, s)$ is continuous and is referred to as the kernel.

Theorem 2.1.41. The adjoint of the Fredholm operator, A , defined in Equation 2.1.3 is given by

$$[A^*f](t) = \int_a^b \overline{K(s, t)}f(s)ds,$$

where $\overline{K(s, t)}$ is the complex conjugate.

Corollary 2.1.42. A Fredholm operator with a Hermitian kernel, that is $\overline{K(s, t)} = K(t, s)$, is a self-adjoint operator.

We present two more properties about the operator norm of adjoint operators, which we will use multiple times throughout Chapter 3, Chapter 4.

Theorem 2.1.43. *If $A : V \rightarrow V$ is a linear operator and $A^* : V \rightarrow V$ is its adjoint, then:*

- $\|A\|_{op} = \|A^*\|_{op}$,
- $\|AA^*\|_{op} = \|A\|_{op}\|A^*\|_{op}$.

Finally, we use these properties to prove a theorem that we will use in this thesis.

Theorem 2.1.44. *For some self-adjoint linear operator $A : V \rightarrow V$, the operator composed with itself 2^n times is also self-adjoint where n is a positive integer. That is,*

$$A^{2^n} = (A^{2^n})^*.$$

Proof. We prove this by induction. For $n = 1$, we get

$$A^2 = AA = A^*A^* = (A^*)^2.$$

We assume that $A^{2^n} = (A^{2^n})^*$ for some $n > 1$ and now consider $A^{2^{n+1}}$.

$$\begin{aligned} A^{2^{n+1}} &= A^{2^n} A^{2^n} \\ &= (A^{2^n})^* (A^{2^n})^* \\ &= (A^{2^{n+1}})^*, \end{aligned}$$

where the second line follows from the induction hypothesis. □

The following corollary follows from Theorem 2.1.44, Theorem 2.1.43.

Corollary 2.1.45. *For some self-adjoint linear operator $A : V \rightarrow V$, $\|A^{2^n}\|_{op} = \|A\|_{op}^{2^n}$, where n is a positive integer.*

■ 2.1.5 Phase-type distributions

When we investigate composite strategies that use a giving-up time in Section 5.1.2, the distribution of the giving-up time depends on how we construct our transition matrix. Once the forager has given up, no further transitions occur, and so the transition matrix has one absorbing state with the rest being transient. This motivates the use of discrete phase-type distributions, which we now introduce. The following definition comes from Latouche and Ramaswami [46].

Definition 2.1.46. Let Z be a discrete-time Markov chain on the states $\{0, \dots, n\}$ with initial probability vector $(\tau_0, \boldsymbol{\tau})$ and transition matrix

$$P = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{t} & T \end{bmatrix},$$

where $\boldsymbol{\tau}$ is a row vector of size n , T is an $n \times n$ matrix, \mathbf{t} is a column vector of size n , and $\mathbf{0}$ is a row vector of zeros of length n . We have $T_{ij} \geq 0$ and $t_i \geq 0$ for $1 \leq i, j \leq n$ with $\mathbf{t} + T\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is column vector of length n filled with 1s.

The distribution of the time X until absorption into the absorbing state 0 is called the discrete phase-type distribution with representation $(\boldsymbol{\tau}, T)$, and we denote the distribution $PH(\boldsymbol{\tau}, T)$.

Some examples of the discrete phase-type distribution are geometric, mixture of geometric, and negative-binomial distributions, with exponential, hyperexponential, and Erlang distributions, respectively, as continuous analogues.

Theorem 2.1.47. *The phase-type distribution is dense in the field of all positive-valued distributions.*

Due to Theorem 2.1.47, we can represent any positive-valued distribution using phase-type distributions. However, the phase-type distribution is light-tailed, and so representing a heavy-tailed distribution is only an approximation, although we can make this approximation as precise as we like, at the expense of large matrices.

■ 2.1.6 Technical terms in foraging

Numerous terms and conventions have formed throughout the history of animal foraging theory. In this section we outline some of the main terms, as well as discuss alternatives that are used. The definitions below are the terms that we use in this thesis, in order to maintain consistency between the existing literature in Section 2.2 and our results.

Definition 2.1.48. A *food target*, or simply *target*, is some item — usually food — that a forager is searching for. All food targets are treated as equal and indistinguishable.

A food target is also sometimes referred to as a *food patch*, although food patch may also be used to refer to an area with multiple targets in close proximity to each other.

Definition 2.1.49. The forager's *radius of vision*, often denoted r_v , is the distance within which a forager can detect food in any direction. It does not necessarily have to be based on the forager's vision — any form of detection is still referred to as the radius of vision.

The radius of vision is often referred to by various similar names such as *vision radius* or *perception range*.

Definition 2.1.50. *Destructive foraging* is foraging under which food targets cannot be revisited more than once.

Destructive foraging represents a food patch being destroyed after it has been eaten. For example, foraging for an animal carcass may be considered as destructive foraging.

Definition 2.1.51. *Non-destructive foraging* is foraging under which food patches can be revisited multiple times.

This represents a food patch that is not destroyed after being eaten. For example, foraging for fruit trees may represent non-destructive foraging.

Definition 2.1.52. A *step* or *relocation step* refers to a continuous movement from one point to another by a forager, as opposed to a single physical step. During a step, a forager cannot decide to change its length and/or direction, and must wait until the end of a step.

A step is often also referred to as a *flight*.

Definition 2.1.53. The length of a step is called the *step-length*.

Definition 2.1.54. The probability distribution of the step-lengths is called the *step-length distribution*.

Definition 2.1.55. A *Lévy flight* is a random walk in which the step-length distribution is heavy-tailed and the time between steps is either instantaneous or at equally spaced time intervals, and thus not dependent on the length of a step. To be a true Lévy flight, the direction of steps must be uniformly distributed in all possible directions.

Definition 2.1.56. A *Lévy walk* is a random walk in which the step-length distribution is heavy-tailed and the time between the steps is related to the length

of the step. To be a true Lévy walk, the direction of steps must be uniformly distributed in all possible directions.

The terms Lévy walk and Lévy flight are often (incorrectly) used interchangeably throughout the animal foraging literature. One reason for this confusion developing is that many papers do not consider the time between the steps, but rather only the length, and so there is effectively no difference between the two terms.

Usually, in order to obtain a Lévy flight a power-law distribution with $1 < \mu \leq 3$ is used as the step-length distribution, although any heavy-tailed distribution can be used. The advantage of using a power-law distribution when investigating Lévy flights is that depending on the parameter μ , a power-law distribution may be either heavy-tailed ($1 < \mu < 3$) or light-tailed ($\mu \geq 3$), and hence will either be a Lévy flight or a Brownian motion. We generate power-law random walks with various parameters in Figure 2.1.1 to demonstrate this.

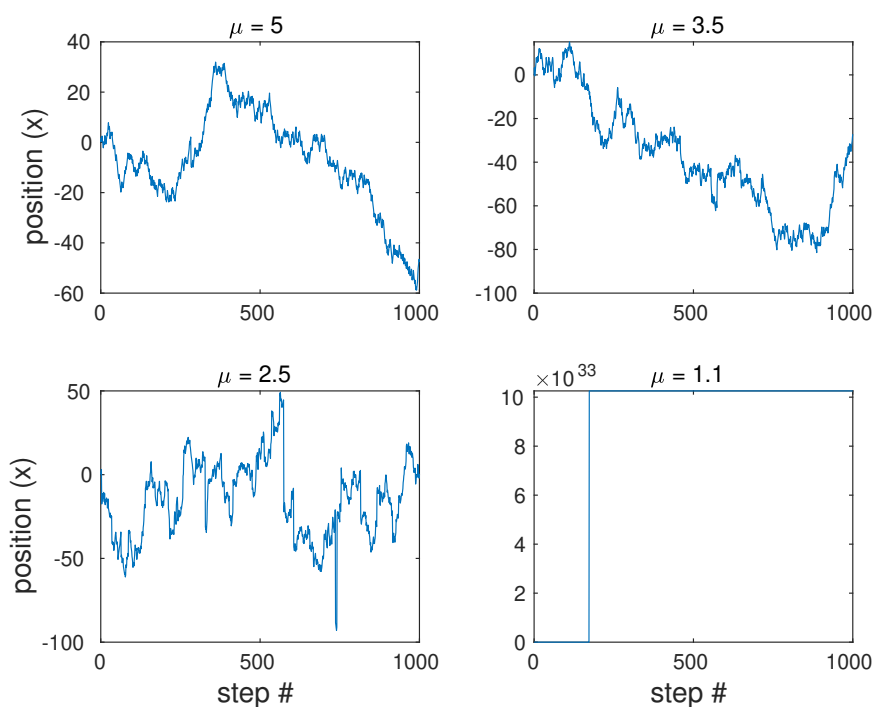


Figure 2.1.1: A realisation of a random walk with power-law step-length distribution, simulated over 1000 time steps for $x_{\min} = 1$ and four different choices of parameter μ . The top-left ($\mu = 5$) and top-right ($\mu = 3.5$) plots correspond to Brownian motion in the scaling-limit, and the bottom-left ($\mu = 2.5$) and bottom-right ($\mu = 1.1$) are Lévy flights.

■ 2.2 Literature Review

James *et al.* [42] gave a thorough overview of stochastic optimal foraging theory in 2011, and discussed both the latest results as well as the history of the field. Our literature review follows a similar structure to their paper, by first examining in detail a few of the early papers in the field, before discussing various extensions that have been made to the basic model. Since this thesis does not investigate any actual movement data, but rather focusses on the theoretical side of the problem, we emphasise the findings and methods of papers that focus on theoretical models. Furthermore, we place the heaviest emphasis on papers that consider various regime-switching strategies or where model parameters such as the radius of vision may vary.

■ 2.2.1 Early evidence for Lévy flight foraging strategies

Empirical evidence in favour of Lévy flights

The earliest evidence of Lévy flights in foraging data came in the form of scale-invariance in the activity periods of *Drosophila*, a genus of fruitfly. Cole [24] examined the activity periods of said *Drosophila*, and observed that they appeared to be undergoing activity constantly. Upon magnification, this constant activity turned out to be multiple bouts of activity, with short bouts of inactivity between. Further magnification revealed that these smaller bouts of activity were once again made up of even smaller episodes of activity with inactivity between these, suggesting a self-similar structure. By plotting on a log-log plot the total number of intervals of inactivity as a function of the length of time required to consider a period inactive, a slope of -1.37 ± 0.12 was found, and thus it was concluded that the temporal distribution of the activity of fruitflies has a fractal structure. The slope being less than -1 implied that the larger the threshold of time that is used to discern between activity and inactivity is, the total amount of time spent inactive increases. Finally, Cole concluded that if we assumed the flies moved at a constant rate during active phases, then the fractal switching of movement activity would translate into fractal space use, producing a Lévy flight.

Viswanathan *et al.* [85] investigated the presence of long-range correlations in the movement data of wandering albatrosses by considering a random walk, $Y(t)$, to represent the net displacement of an albatross. The *mean square displacement* of the albatross is given by

$$F(t) = \sqrt{\mathbb{E}[\Delta Y(t)^2] - \mathbb{E}[\Delta Y(t)]^2},$$

where $\Delta Y(t) := Y(t_0 + t) - Y(t_0)$.

An uncorrelated random walk, or a Markov process for a sufficiently large t , would result in

$$F(t) \sim t^\alpha,$$

with $\alpha = 1/2$, due to the Central Limit Theorem [85]. The mean square displacement, when plotted on a log-log plot against time, demonstrated a power law relationship with $\alpha = 0.84 \pm 0.2 \neq 1/2$, implying that long-range correlations do, in fact, exist in the wandering albatross data. By plotting the sum of the power spectra against the time period used on a linear plot, and against the frequency on a log-log plot, they were able to further confirm the existence of long-range correlations as well as scale invariance. To determine the origin of the scale invariance, the authors plotted the distribution of flight-time intervals on a log-log plot, and found that the data approximately fitted a power-law distribution, with a slope of $\mu \approx 2$.

The authors also used numerical simulations to determine the mean square displacement fluctuations that would be observed under a Lévy flight model in which the probability density for a flight time, t , is given by

$$p(t) \sim (t + 1)^{-\mu},$$

where the '+1' is introduced to account for the bird spending exactly one unit of time each time it lands. Fixing $\mu = 2$ in this model, based on the observed slope from the flight-time data of wandering albatrosses, the mean square displacement was found via numerical simulations. The simulation of this model yielded long-range correlations with $\alpha = 0.8 \pm 0.05$, which was consistent with the long-range correlation found from the data.

The reasons for the landing points of wandering albatrosses being spatially scale-invariant was speculated on by Viswanathan *et al.* [85], with one possible reason being that the distribution of food is also scale invariant. Another possible explanation is related to the lifetime distribution of the thermal columns that birds use to produce lift. A potential advantage may also be found by considering foragers that operate as a flock, rather than an individual. An individual Lévy walker visits t new sites in the time that an individual Brownian walker visits $t/\ln t$ sites, whereas a flock or swarm comprised of N Lévy walkers finds Nt sites, while a swarm of Brownian walkers visit $t \ln(N/\ln t)$, making the improved efficiency of the Lévy flight pattern considerable when dealing with large flocks. Since evidence of Lévy flights was found for albatrosses, the authors suggest further work should be done examining other animals for Lévy flight patterns.

Using the same methods as previous papers [24, 85], Viswanathan *et al.* [86] investigated the foraging behaviour of bees, wandering albatrosses, both fenced and unfenced deer, and amoeba. Using log-log plots and line of best fit, the flight-length distribution was found to fit a power-law distribution with $\mu \approx 2$, which matched the theoretical optimal strategy that they found. The amoeba data was found to fit a power-law with $2 \leq \mu \leq 2.5$, which also corresponds to a Lévy walk.

Recent reanalysis of the empirical evidence has contradicted some of these conclusion. In particular, Edwards *et al.* [30] devised a maximum likelihood estimation (MLE) method to determine more rigorously whether movement data fits a particular distribution. Using this, they revisited many of the original animals that were investigated using higher resolution datasets and found that most did not undergo Lévy flights. Towards the end of the literature review we discuss the more recent statistical methods, the reanalyses of the empirical results, and some further problems with data analysis in the field of animal foraging, in further detail.

Basic theoretical models for animal foraging

Following the empirical evidence of Lévy flights, Viswanathan *et al.* [86] investigated what search strategy provides a theoretical optimal for a simplified model. Their model was two-dimensional and they looked at both *destructive* and *non-destructive foraging*. The *search efficiency* was defined to be

$$\eta = \frac{1}{\mathbb{E}[\ell] N},$$

where $\mathbb{E}[\ell]$ is the mean flight distance, and N is the mean number of steps taken to find food. The step lengths taken by foragers were drawn from a power-law distribution. By assuming the *mean free path* between successive targets, which we denote as λ , are all equal, and denoting the forager's *radius of vision* as r_v , the mean flight distance was expressed as

$$\mathbb{E}[\ell] = \frac{\int_{r_v}^{\lambda} \ell^{1-\mu} d\ell + \lambda \int_{\lambda}^{\infty} \ell^{-\mu} d\ell}{\int_{r_v}^{\infty} \ell^{-\mu} d\ell},$$

which can be solved exactly. The mean number of steps taken to find a target is denoted either N_d or N_n , for destructive and non-destructive foraging, respectively.

To solve for N_d , the authors defined $\mu - 1$ as the fractal dimension of the sites, and then used an existing result [35] to determine that the mean number of steps until a target is found to scale as $N_d \approx (\lambda/r_v)^{\mu-1}$, for $1 \leq \mu \leq 3$, which is valid

for the destructive case. If $\mu \geq 3$, which corresponds to Brownian motion, then the mean steps scale as $N_d \approx (\lambda/r_v)^2$. To solve for non-destructive foraging, they showed that $N_n \approx N_d^{1/2}$ using a one-dimensional Brownian walker. More specifically, a Brownian walker at the middle of an interval has $N_d = \lambda^2/(2D)$, where D is the diffusion constant, and a Brownian walker at some small distance, r_0 , from the edge of an interval has $N_n = (\lambda - r_0)r_0/(2D)$. This implies that $N_n \approx N_d^{1/2}$, and the authors assumed that this relation remains true even for $\mu < 3$, and thus arrived at $N_n \approx (\lambda/r_v)^{(\mu-1)/2}$ for $1 \leq \mu \leq 3$.

Then, with their approximate expressions for the efficiency, Viswanathan *et al.* [86] determined the most efficient strategy for a range of different scenarios. When target sites are plentiful, $\lambda \leq r_v$ and so $N_d \approx N_n \approx 1$, meaning the efficiency is independent of search strategy. For sparse targets, the efficiency for destructive foraging was found to asymptotically increase as $\mu \rightarrow 1$, corresponding to ballistic motion. For non-destructive foraging when targets are sparse, an optimum was shown to exist at $\mu = 2 - \delta$, where $\delta \approx [\ln(\lambda/r_v)]^{-2}$, which corresponds to a Lévy flight.

Viswanathan *et al.* [86] verified the validity of their conclusions by performing numerical simulations, which were found to match the analytic approximations they had derived. Further, the fact that the optimal strategy for non-destructive foraging matched previous empirical results [24, 85] strengthened the validity of the model. Thus, their model seemed to be simple enough to obtain some analytic results, while still being complex enough to consider various different scenarios and maintain realism. This paper became one of the foundations of stochastic animal foraging theory, with many papers building upon this model.

Bartumeus *et al.* [6] considered a one-dimensional foraging model, which they showed can be reduced to a single interval with absorbing boundaries at either end representing food targets. They considered a forager doing a random walk with an arbitrary step-length distribution, and found an exact expression for the expected total length until food is found, and hence the foraging efficiency. Some of the results of this paper already existed [20, 21], although not in the context of animal foraging. They confirmed the optimal efficiency of a power-law random walk occurs at $\mu = 2$, representing a Lévy walk.

Further empirical investigation and methodology

The evidence of Lévy flights in animal foraging data [24, 85, 86] sparked an interest in investigating the step-length distribution of many different species of animals, with many papers following these first three. Most species investigated

were found to have some evidence of Lévy flight movement paths (e.g. [51, 62, 72, 73, 80]), although this was not the case for every species (e.g. [2]).

Much of the investigation of animal movement data used the same methods as the original Lévy flight papers of Cole [24], and Viswanathan *et al.* [85]. The most commonly used technique was plotting the distribution of step-lengths on a log-log plot to determine the parameter of the power-law distribution. Mårell *et al.* [51] investigated semidomesticated female reindeer, for which the line of best fit on a log-log plot of the step-length distribution had a slope of $\mu = 1.8$ or $\mu = 2$ depending on the time-scale. Ayala-Orozco *et al.* [62] studied the movement patterns of free-ranging spider monkeys and found that the fitted power-law distribution has parameter $\mu = 2.18$. Austin *et al.* [2] found that only 15.3% of grey seals had movement data that fit a negative power law distribution based upon regression lines on a log-log plot, and suggested that the lack of evidence for Lévy flight may be indicative of non-randomly distributed food items. Sims *et al.* [80] used a log-log plot to determine the power-law parameters for penguins ($\mu = 1.7$, $r^2 = 0.91$), sharks ($\mu = 2.4$, $r^2 = 0.90$), as well as sea turtles and various bony fish, which had parameters between that of penguins and sharks. Reynolds *et al.* [70] investigated *Drosophila* again, and found that a power-law distribution with a parameter of $\mu \approx 2$ provided the best fit. Reynolds *et al.* [72] found the frequency distribution of the appetitive *Agrotis segetum* moths to fit a power-law distribution with $\mu = 1.5$. This value of μ is consistent with the optimal parameter for a biased Lévy flight, which has $1 < \mu \leq 2$. Reynolds *et al.* [72] reasoned that a biased strategy occurred due to the way the experiment was conducted, with the detection of pheromones depending on whether or not the target was upwind or downwind.

Some other methods were also used in investigating animal movement data. Reynolds *et al.* [73] considered the mean square displacement of the flight lengths of honey bees, $F(t)$, finding $F(t) \propto t^\alpha$, where $\alpha = 0.85 \neq 0.5$, which is indicative of long-term power-law correlation. This scale-free behaviour is consistent with a Lévy walk with $\mu \approx 2$, and the scaling exponent found can be reproduced with a truncated Lévy flight with $\mu = 2$. Sims *et al.* [80] also plotted the mean square displacement of the displacement of various sea creatures, and found values of α between 0.80 and 1.24, which implies a long-term correlation and indicates scale-invariance in the movement pattern. Ayala-Orozco *et al.* [62] also considered an exponential distribution ($r^2 = 0.79$) for the step-length distribution, although the power-law ($r^2 = 0.89$) was shown to provide a better fit. Mårell *et al.* [51] found the fractal dimension in reindeer movement lengths to be between 1.0 and 1.5, with the time-scale having no effect. The fractal dimension of Brownian

motion is 2, and thus they concluded that Brownian motion was not a good fit for reindeer movement, which implies that a Lévy flight may be a better fit.

Evidence was also found for heavy-tailed distributions in various other aspects of animal movement, beyond just the distribution of step-lengths. Ayala-Orozco *et al.* [62] found evidence of a heavy-tailed distribution of waiting times in the movement of spider monkeys, with a power-law distribution ($\mu = 1.7$, $r^2 = 0.86$) providing the best fit. Sims *et al.* [80] examined horizontal krill density in the current of a water column passing an echosounder. The frequency distribution of changes in krill density was plotted on a log-log plot, and found to fit a power-law distribution with $\mu = 1.7$. Since krill is consumed by some of the predators from the study, this finding suggested that the existence of Lévy flight behaviour in predator movement may be in response to prey distribution.

Sims *et al.* [80] suggested two possible explanations for the existence of Lévy flight behaviour in animal foraging movement: animal search patterns are adapted to stochastically optimise their search for food and are essentially ‘blind’ at the spatio-temporal scale that they search at, or search patterns that are apparently optimal simply arise as a function of the distribution of the prey. The authors provided further support for the latter explanation, by showing that Lévy walk search strategies have advantages in landscapes with fractal prey distributions compared to other prey distributions. Through numerical simulations of vertical diving movement, the encounter rate of both Lévy walks and random searches in both Lévy and fractal prey fields are tested, finding that Lévy walks in Lévy environments had encounter rates 14% higher than in random environments.

■ 2.2.2 Extensions to the basic foraging model

Foraging for moving targets

Bartumeus *et al.* [3] considered the case of foraging for moving targets, which can be thought of as a *predator* searching for moving *prey*. The authors performed a simulation study of a single forager and a single moving target in one-dimension on an interval of length L , with periodic boundary conditions, this domain is equivalent to a circle with perimeter L . Varying the size of the system, L , effectively varies the density of food targets. Both the searcher and the target move with a constant velocity, and each step j that they take is in a random direction and of length ℓ_j , where ℓ_j is chosen from a power-law distribution. Efficiency was defined to be the encounter rate per unit distance travelled by the searcher. The ratio of *searcher velocity* to *target velocity* was denoted v , and the ratio of size — usually representing either physical size or detection radius

— between the searcher and the target was r . Both the target and the searcher adopt either Brownian ($\mu = 3$) or Lévy ($\mu = 2$) searches. The simulation was repeated, and after an encounter the target is destroyed and a new one created in a random location. As the system size gets larger, the relative efficiency of a Lévy walk search compared to a Brownian search increases. When the targets move according to a Brownian motion, a Lévy walk search is the most efficient, except for the extreme case of very large (approximately $r \geq 10$) and fast moving targets (approximately $v \geq 10$). However, when the food targets follow a Lévy walk, a Brownian search often outperforms a Lévy search; for example, when target density is high ($L = 25$), Brownian search outperforms Lévy search when $r > 1$ and $v > 1$. The authors found that with Brownian targets, size ratio and velocity ratio are both equally important, but with Lévy targets, velocity ratio is more important than size ratio.

James *et al.* [41] continued investigating this same model of moving searcher and target, and confirmed the results that a Lévy search outperforms a Brownian search for Brownian targets, with its relative advantage greater when v is smaller. However, the authors did not restrict themselves to just $\mu = 3$ and $\mu = 2$ for the searcher as Bartumeus *et al.* [3] did, but instead investigated an entire spectrum of values from $\mu = 1$ to $\mu = 3$, with the food target following either a ballistic path ($\mu \rightarrow 1$), Lévy walk ($\mu = 2$), or Brownian motion ($\mu = 3$). They found that a ballistic search actually outperforms both a Lévy search and a Brownian search in all but one case — the unrealistic case when the searcher and target have identical speeds ($v = 1$) and travel in the same direction, never encountering one another. The authors also derived an analytic expression for the mean efficiency of a ballistic search, η , based on a geometric approach

$$\eta = \begin{cases} 2(1 - v^2)/L & \text{for } v \leq 1, \\ [2(v^2 - 1)]/(Lv) & \text{for } v \geq 1. \end{cases}$$

James *et al.* [41] also extended the model by considering many moving targets, each moving independently of each other, which they argued is a more accurate approximation of a higher-dimensional setting. They assumed the searcher begins at 0, and considered the case of evenly-spaced targets at $(2n - 1)L/2$, $n \in \mathbb{Z}$, as well as the case with uniformly distributed targets, both cases having the same target density. A predator using a ballistic search was once again shown to be more efficient than either a Brownian or Lévy search.

Faustino *et al.* [33] considered the same model as James *et al.* [41], although only the case of $v = 1$, and rather than using a pure power-law distribution, they instead used a truncated power-law distribution. The optimal search strategy

was found to be when $\mu = 1$, which does not break down in this case as it did in James *et al.* [41], since it is not actually ballistic motion. James *et al.* [41] declared that the efficiency of this truncated power-law search with $\mu = 1$ is the same as the ballistic search when $v \geq 1$, and is outperformed by the ballistic search when $v < 1$. James *et al.* [41] concluded that the superior efficiency of the ballistic search against all others, and of the Lévy search over the Brownian search, is due to the lack of backtracking allowing more ground to be covered by the searcher. James *et al.* [41] also pointed out that the work of Bartumeus *et al.* [3] is somewhat flawed: it considered only the size ratio r_t/r_p and not the absolute values, although the sum of their sizes, $r_t + r_p$, relative to the system size, L , is actually what determines the target density. Keeping $r_t + r_p$ fixed and varying the ratio r_t/r_p , Faustino *et al.* [33] showed that the efficiency does not change.

Bartumeus *et al.* [4] performed another simulation study, this time considering a range of μ between 1 and 3, as well as varying target density, target size, target cruising velocity, and both destructive and non-destructive searches, on 1D, 2D, and 3D search spaces. Their results confirmed the results of James *et al.* [41] for the destructive case, and found that $\mu \approx 2$ is most efficient for the non-destructive case, with the exact value depending on the system's parameters. They noted that although increasing the dimension reduces the relative efficiency of one strategy over another, it does not change what the optimal strategy is. When targets travel slower than the searcher, the encounter rates are affected only by the search strategy of the searcher, and when searchers travel slower, the encounter rates are affected only by the movement strategy of the target. That is, the encounter rate depended only on the movement strategy of the faster moving organism. When both searcher and target move at similar speeds, the most efficient search strategy for the searcher was to use a Lévy search with parameter as far away from the parameter of the targets Lévy movement as possible (e.g. $\mu = 3$ if the target is moving according to $\mu \rightarrow 1$, and $\mu \rightarrow 1$ if the target is moving according to $\mu = 3$).

Food targets with a revisitability delay

Raposo *et al.* [63] studied the role of a *delay time* — or *revisitability delay* — which was denoted τ . After an animal finds a food target, the same target cannot be revisited until τ time has elapsed since originally visiting it. A revisitability delay essentially provides a continuum between non-destructive ($\tau = 0$) and destructive foraging ($\tau = \infty$). Raposo *et al.* began with the model outlined by Viswanathan *et al.* [86] and extended it by adding a delay time to the targets. To determine

the optimal strategy, the authors used a result from the one-dimensional case found by Buldyrev *et al.* [20], that if a forager begins at a point $x = z\lambda$ in the interval $[0, \lambda]$ with $0 < z < 1$ and $\lambda/\ell_{\min} \rightarrow \infty$, the optimal parameter for a power-law strategy is

$$\mu_{\text{opt}} = 2 + \frac{2}{\log z} + o\left(\frac{1}{\log z}\right).$$

Raposo *et al.* [63] related the distance $z\lambda$ to the forager's velocity and the delay time by

$$z = \frac{v\tau}{\lambda}.$$

This is based on the assumption that the forager moves at a constant velocity v away from the target during the delay time, and then begins its search again once the delay time expires, which will be at point $z\lambda$. Using this, the authors showed that the case of $\tau \rightarrow 0$ implies that $z \rightarrow 0$ and $\log z \rightarrow -\infty$ and hence $\mu_{\text{opt}} = 2$. Similarly, when $z > 1/e^2$ — which occurs when $\tau > \lambda/(ve^2)$ — the optimal strategy is $\mu < 1$, corresponding to straight line motion with constant velocity. Although the optimal efficiency is not known exactly for higher dimensions, Raposo *et al.* [63] argued that the results of the one-dimensional case have validity since in d dimensions the forager moves through a one-dimensional corridor of cross section proportional to r_v^{d-1} , and they confirmed this for the two-dimensional case using simulations. The authors concluded that the most efficient strategy is strongly dependent on the delay time of the targets as long as $\ell_{\min} \geq r_v$, where the efficiency is defined as the number of targets found per distance travelled, with $1 < \mu_{\text{opt}}(\tau) \leq 2$, based on the two limiting cases of destructive and non-destructive foraging.

Raposo *et al.* [63] also considered a more general cost function, $f(\mathbb{E}[L])$, which is a function of the expected distance to find a single target. This definition of efficiency can take into account various phenomena that make the model more realistic, such as energy gain due to calories and nutrients in the targets. The optimal values of μ for a Lévy walk do not change when using this more general notion of efficiency, although depending on the cost function the values that μ can take may be limited.

Santos *et al.* [76] further investigated a delay time, and extended the results of Raposo *et al.* [63] by effectively interpolating between the two known results, which are the limiting cases of destructive ($\tau \rightarrow \infty$) and non-destructive foraging ($\tau = 0$). As found originally by Viswanathan *et al.* [86], the expected number of steps to find a food target scales as $N_s \sim \lambda^{(\mu-1)/2}$ for a search beginning at $x_0 = r_v$ (non-destructive), and $N_s \sim \lambda^{\mu-1}$ for $x_0 = \lambda/2$ (destructive). Based on

this, Santos *et al.* [76] determined the more general form for the mean number of flights,

$$N_s \sim \lambda^{(\mu-1)/\Gamma},$$

where $\Gamma = 2$ and $\Gamma = 1$ are the two limiting cases for non-destructive and destructive foraging, respectively. The authors then considered $\Gamma(\tau)$, as a monotonically decreasing function with $\Gamma \rightarrow 2$ when $\tau \rightarrow 0$ and $\Gamma \rightarrow 1$ when $\tau \rightarrow \infty$. Using this function, Santos *et al.* [76] arrived at an expression for the efficiency depending only on the value of Γ as well as the ratio λ/r_v . The authors then plotted the efficiency of a Lévy search for a range of different values of τ , showing the crossover between the two different limiting cases. Santos *et al.* [76] also considered the effect of a more general cost function which takes into account the forager's energy, and concluded that optimal parameters will not be affected by this different notion of efficiency, matching the conclusions of Raposo *et al.* [63].

Two-dimensional and three-dimensional foraging

Santos *et al.* [78] considered a two-dimensional search on a lattice, with two different types of lattice topology: a *square lattice*, with 4 possible directions from each node, and a *triangular lattice*, with 6 possible directions from each node. Targets were assumed to be distributed homogeneously among the nodes of the lattice, and both the non-destructive and destructive cases were considered. A forager would choose a uniformly random direction and a step-length from a power-law distribution. If a target is within a distance r_v of a forager, the forager may move directly to it if there is a straight-line path towards it, meaning zigzag paths were restricted. Using this simplified model, the authors were able to obtain some analytic results. A Lévy walk with $\mu \approx 2$ was only optimal for low target densities, with a ballistic path being the optimal strategy for higher target densities.

Santos *et al.* [77] *et al.* built upon the previous two-dimensional lattice model [78], by introducing *defects* that were randomly scattered throughout the lattice. Under their model, foragers would travel in a straight line until they came into contact with a defect, after which they would choose a new direction and distance. One way of thinking about defects is as food patches that are found to be empty upon arriving at them. The optimal strategy was found to heavily depend on the density of targets and defects, as well as the boundary conditions employed. For $\mu > 2$, the average step length is small and so truncation does not happen very often; consequently, the density of food patches and defects matter somewhat less, and the boundary conditions have almost no effect on the optimal strategy. Raposo *et al.* [64] also considered a very similar model to Santos *et al.*

[77], involving a lattice with defects. Using both analytic approximations and simulations they came to similar conclusions as Santos *et al.* [77], finding that as the presence of defects increases, the relative advantage of one strategy over any other decreases.

■ 2.2.3 Composite search strategies for improved efficiency

Foraging strategies with a giving-up time

Benhamou [8] considered *composite Brownian walk* strategies as alternatives to Lévy walk strategies when searching in a *patchy environment*. A composite Brownian walk is a mixture of two random walks, both with exponentially-distributed step sizes. A patchy environment was here defined to have uniformly distributed patches of food, and within these patches the density of food is much higher than the density of food outside of the patches. The forager considered by Benhamou began with an extensive search, in which large steps were drawn from an exponential distribution, and switched to the intensive search once a food patch was reached, with step-lengths also drawn from an exponential distribution, with different parameters resulting in a smaller mean. After some time, referred to as the giving-up time, had elapsed, the forager would switch back into the extensive movement mode. In a patchy food environment with destructive targets, the forager begins its next search in the vicinity of a target since multiple targets are clumped together, and so this scenario is equivalent to the non-destructive scenario of Viswanathan *et al.* [86]. Through simulations, Benhamou [8] showed that a composite Brownian walk appeared to be more efficient than a Lévy walk for finding food in a patchy environment.

Plank and James [61] derived some analytic results for a similar model in one dimension, in which a forager follows a Brownian motion for its intensive search, before giving-up and undergoing ballistic motion for its extensive search. They reasoned that, due to the Central Limit Theorem, the discrete model of Benhamou [8] will converge to their model after a large number of steps, as long as the variance of the step-length distribution is finite. In deriving an expression for the density of the total distance travelled to find food, they used an existing result for the density of Brownian motion with an absorbing barrier at $x = 0$, from Grimmett and Stirzaker [36]. This also means that the results are only valid for strategies which have Brownian motion as the intensive search strategy, but not a Lévy walk, for example. With their model, Plank and James [61] derived an analytic expression for the optimal giving-up time, as well as showed that the composite Brownian walk strategy outperforms a Lévy walk strategy, in

agreement with the results of Benhamou [8].

Adaptive Lévy walks as a more general giving-up time strategy

Reynolds [67] explained that the composite Brownian walk search of Benhamou [8] can actually be viewed as a special case of an *adaptive Lévy walk search*. An adaptive Lévy walk involves switching between a Lévy walk with $\mu \rightarrow 1$ for the extensive search, and a Lévy walk with $\mu = 3$ for the intensive search. Then, Reynolds argued, that when considering the CBW as an adaptive Lévy walk the results of Benhamou [8] follow directly from the Lévy walk search theory of Viswanathan *et al.* [86], since a Lévy walk with $\mu \rightarrow 1$ was optimal for sparse targets, and $\mu = 3$ for densely distributed targets. Benhamou [9] replied to Reynolds [67], pointing out that the composite Brownian walk could not be thought of as an adaptive Lévy flight for two reasons: the inverse power-law has infinite variance when $1 < \mu < 3$ and so $\mu = 3$ does not correspond to a Lévy flight, and when $\mu \rightarrow 1$ a Lévy flight tends to a straight line of infinite length and so must be truncated, effectively making the variance finite. Benhamou's criticism is mostly valid, since to be Brownian motion (in the scaling limit), a strategy must have a step-length distribution with finite variance, whereas a Lévy walk uses a power-law distribution with infinite variance, by definition. However, this disagreement between Reynolds and Benhamou is ultimately over little more than semantics, since if although a power-law with $\mu = 3$ is technically not a Lévy walk, there is no mathematical reason as to why the step-length distribution cannot have $\mu = 3$. For example, if the adaptive Lévy walk search model was instead referred to as an adaptive power-law search model, since the power-law can produce both a Lévy walk and a Brownian motion, then the conclusions of Reynolds [67] would still be true.

Reynolds [66] then considered the adaptive Lévy walk model but allowed for any value of μ during the extensive phase, which reduces to the CBW model of Benhamou [8] when $\mu \rightarrow 1$. This time it was used to investigate non-destructive foraging ($x_0 = r_v$), which as previously shown [67] is equivalent to foraging in a patchy environment. It was determined that the optimal extensive parameter was between $\mu = 1$ and $\mu = 2$, with a higher food target density moving the optimal closer to $\mu = 2$. James *et al.* [42] noted that Reynolds [66] only considered a fixed giving-up time, whereas Plank and James [61] optimised over all giving-up times.

Nolting [56] investigated composite strategies such as giving-up time strategies, as well as a new type of strategy that a forager may use, which they called the *optimal zone forager*. As with the giving-up time strategy, the optimal zone

forager switches between an intensive and extensive search, although this time depending on its current location. Thus, the forager must have *a priori* knowledge of where the food targets are located, although it cannot use this knowledge to move directly to a resource, but rather it can only change modes. For a strategy that uses a Brownian motion and ballistic motion for its intensive and extensive search, Nolting reasoned that the optimal zone forager is ideal.

Searching for difficult-to-detect targets

A one-dimensional composite model was also considered by Bénichou *et al.* [14], in which a forager switches between an intensive phase, also known as the *searching phase*, and *extensive phase*, also known as the *relocation phase*. In this model, the forager was incapable of detecting targets during the relocation phase, and so the model was referred to as searching for hidden targets. The time spent in each phase was exponentially distributed, and the forager could switch back-and-forth between phases an unlimited number of times. As with many of the other composite strategies (e.g. [8, 61]), the intensive phase consists of Brownian motion, and the extensive phase is ballistic motion which is always in the same direction. This persistence in the direction of the extensive search was referred to as *orientational memory*, and was also addressed by later papers. A uniform density of patches is assumed, and so the problem is reduced to finding patches in a single interval, as done in other papers (e.g. [6]). The authors considered a forager that begins at the middle of an interval, which corresponds to destructive foraging. Using an analytic argument, they found that the optimal search strategy under their model occurs when the average duration of the searching phase, denoted f_1 , is proportional to the average duration in the relocation phase, denoted f_2 , by either $f_1 \propto f_2^{3/5}$ or $f_1 \propto f_2^{2/3}$ depending on the frequency of the switching.

Bénichou *et al.* [13] considered the effect of how difficult a target is to detect, and whether this changed the optimal search strategy. Two types of targets were considered, those that are easy to detect, and those that are difficult to detect, referred to as *hidden*. A difficult to detect target requires a forager to pass through the area multiple times before it is finally located, whereas an easily detectable target emits a strong enough stimulus to allow a forager walking past to detect it immediately, as long as it is moving slowly enough. Composite strategies were again considered, with exponentially distributed time between intensive Brownian and extensive ballistic search modes. For the case of easily detectable targets, it was shown that staying in the active search phase is optimal, as opposed to any composite strategy. For hidden targets, the optimal strategy was shown to be a composite strategy where the time spent in each phase is

as short as possible, which matches the findings of Bénichou *et al.* [14]. After removing the forager's orientational memory, the optimal search strategy was found to differ. With orientational memory, the efficiency scales asymptotically with the mean time spent in each state, whereas without orientational memory the optimal efficiency occurs at finite values for the mean times in each state, which depends on the target density. The efficiency of searching for hidden targets with orientational memory was found to always be greater than that of searching without orientational memory.

Lomholt *et al.* [49] generalised the model of Bénichou [13] by allowing the time spent in the extensive phase to be drawn from either an exponential or Lévy distribution. It was found that strategies with Lévy distributed time spent in the extensive phase outperform the previously investigated strategies with exponentially distributed time in each phase. Switching between a Brownian motion in the intensive phase and a Lévy walk in the extensive phase was shown to offer the most efficiency, as well as also being less sensitive to changes in the target density when compared to other strategies. Of course, these conclusions relied on the forager not being able to detect food during its extensive phase, as discussed by Bénichou *et al.* [13].

Bénichou *et al.* [15] extended the hidden targets destructive foraging model [14] to two dimensions. They again considered composite strategies, with exponentially distributed time between diffusive searching and ballistic motion. Taking an approximate analytic approach, they determined that composite two-state strategies optimise the search time required to find a target, as opposed to one-state strategies such as Lévy flights.

Reynolds [69] considered the original model of Viswanathan *et al.* [86], but allowed food to be detected only during a step of length $\ell < \ell_0$, where ℓ_0 is some fixed threshold. The optimal parameter for a power-law search was found to be $\mu \approx 2$ for both non-destructive and destructive targets. If ℓ_0 is the threshold between a step being considered either intensive or extensive, then Reynolds showed that there is a power-law distributed amount of time spent in each phase.

■ 2.2.4 Statistical methods for analysing animal movement data

As discussed above, many early papers [24, 51, 62, 70, 72, 73, 80, 85, 86] found empirical evidence of Lévy flights in animal movement data. Most of these [24, 62, 71, 72, 80, 85, 86] analysed the movement data by plotting the distribution of step-lengths on a log-log plot, and some [51, 62, 73, 80] also considered the mean square displacement of the flight lengths. However, more recent reanalysis has

put these methods into question.

The effect of subsampled data on the apparent step-length distribution

Benhamou [8] questioned the validity of using subsampled data to analyse animal movement. Most animal movement data sets represent a subsample of the animal's movement, since the time between recordings of successive locations is usually significantly longer than time-scale at which an animal moves. Reynolds [67] addressed this issue by simulating true Lévy walk searches with $\mu = 2$ and plotting the distribution of subsampled movement lengths, which was still found to fit a log-log plot, though now with $\mu = 1.6$. Reynolds suggested that subsampling would result in an effective value of μ less than the true value.

However, Plank and Codling [60] found that the sampling rate at which movement data is collected can affect the apparent properties of the movement path, and can even cause misidentification. Using simulations, they show that it is possible to misidentify a composite correlated random walk which switches between two different phases, as a Lévy walk, even when using appropriate statistical procedures such as maximum likelihood methods that were previously outlined [31, 32, 61]. Plank and Codling [60] also showed that when a Lévy walk is subsampled, it can be misidentified as an exponential distribution, with a larger μ making an apparent exponential more likely. The point of greatest uncertainty between an exponential and power-law distribution occurs at $\mu \approx 2$. This conclusion contradicts the claim of Reynolds [67], who had not compared the goodness-of-fit with that of another distribution.

To avoid the issues with using data that is a subsample of an animal's true movement path, some papers have instead defined turning points (or reorientation points) which correspond to locations at which a forager changes direction by a significant amount. Reynolds [73] used both *local* and *non-local* determination of a turning point, and found that the statistical properties of the movement distributions do not change significantly between the two. When determining the turning points *locally*, a new point was said to have occurred when three successive recorded positions of the forager involve a change in direction greater than some critical angle, which in this case was 90° . *Non-local* determination instead used the cumulative change in angle since the last turning point, with a new point occurring once this was greater than the critical angle. Since the bee movement distributions found with each of the two methods were not significantly different to each other, this indicated that most direction changes are abrupt, and the authors found that the direction of flight paths are distributed uniformly.

Reynolds [68, 72] also used local turning point determination when investigating the movement of moths and bees again, and found that the choice of critical angle had no significant effect on the results.

Codling and Plank [22], however, found that whether or not a movement path is identified as having a power-law distribution depends on both the sampling rate and the designation of turning angles. They simulate correlated composite random walks and Lévy walks and find that there is no standard method of turning-point identification that is robust for all cases, with non Lévy walks often displaying Lévy-like characteristics.

Improved identification using MLE methods

Viswanthan *et al.* [87] proposed a criterion that is necessary, but not sufficient, to establish true superdiffusive behaviour in animal movement data. If τ is the estimated correlation time of a general correlated random walk, then the data must display superdiffusive behaviour on scales larger than τ for there to be true superdiffusive behaviour.

Edwards *et al.* [32] revisited the movement pattern of albatrosses, now with a higher resolution dataset than the original study [85]. They discussed some of the inadequacies of the usual data analysis methods used throughout the animal foraging literature (e.g [24, 62, 85, 86]), and outlined a MLE method to determine whether movement data fits a power-law distribution. They analytically solved the MLEs for both a power-law distribution and an exponential, and used these to determine the Akaike's Information Criterion (AIC) and the relative likelihood of each model. Using their MLE method and the new albatross dataset, they determined that there was no evidence of Lévy flight behaviour in the movement of albatrosses, but rather they followed a gamma distribution. After correcting the original albatross data [85], they also determined that the original conclusions of Lévy flights being present in the data were spurious. Furthermore, they applied their method to the original deer and bumblebee datasets [86], and determined that none exhibited evidence of Lévy flights.

White *et al.* [89] compared various methods of fitting data to a distribution, including logarithmic binning and MLE methods. Power-law datasets were generated using Monte Carlo methods, and then using each method the distribution and parameters were determined, before finally comparing the performance of the methods by considering their bias and variance. Logarithmic binning was found to give an incorrect exponent, and adjustments were necessary to obtain the correct slope. Further, binning methods in general performed poorly, and

White *et al.* [89] recommended avoiding these whenever possible. MLE methods were shown to be the best approach, and produced valid confidence intervals for the estimated exponents.

Further issues with identifying Lévy walks in foraging data

Plank and James [61] simulated a Lévy walk and a composite random walk and investigated fitting the movement data to step-length distributions. They determined the best-fit parameters and the log-likelihood for each model, and in both cases found an exponential distribution provided a better fit than the power law distribution. They conclude that both a Lévy and a non-Lévy process can produce a non-Lévy pattern, which complements the results of Benhamou [8], who showed that both a Lévy and non-Lévy process can produce a Lévy pattern.

Following some of the inadequacies found with the early evidence for Lévy flights in foraging data [32], Edwards [30] readdressed many of the previous datasets to determine if Lévy flights really are prevalent in ecology. He used the modern MLE and AIC approach [32] to compare Lévy flights to other simple models, and examined 17 different datasets from previous studies [1, 2, 5, 16, 18, 50, 51]. The possible distributions were: unbounded power-law (Lévy flight), bounded power-law (truncated Lévy flight), unbounded exponential, and bounded exponential. Almost all of the original estimates of the parameters were found to lie outside 95% confidence intervals for the newly calculated parameters. Further, the evidence of Lévy flights was overturned for all 17 datasets. The gray seal dataset was found to have come from a bounded power-law distribution, or a truncated Lévy flight, and the possibility of the hunter-gatherer dataset coming from a bounded power-law distribution also could not be ruled out. Three of the 17 datasets were found to fit an unbounded exponential distribution, and the remaining 12 datasets did not match any of the four tested distributions. Thus, Edwards concluded that Lévy flight movement patterns are not as common a phenomena in ecology as once thought.

Empirical evidence for composite strategies

There has been empirical evidence of animals using composite (or intermittent) strategies, including honeybees [84], fish [39], birds [55], turtles [84], weasels [37], slime moulds [47], beetles [34], and others [10, 43, 45, 53, 66].

Further, many animals have been found to switch between an intensive and extensive search based on sensory cues from their environment [7, 25, 27, 29, 38, 48, 52, 54, 59, 81, 82]. For example, some parasites were shown to use chemical

cues to decide when to use an intensive search [7, 81, 82], and in a similar way Procellariiform seabirds were shown to use dimethyl sulphide to determine when an intensive search should be used [54].

Some of the evidence of composite strategies is based on the qualitative behaviour of animals, while there are also papers that examine the step-length distributions of movement data using various quantitative methods, such as state space models (SSMs), [43], hidden Markov models (HMMs) [44, 58], Bayesian inference [57] and deep-learning models [19].

Chapter 3

One-dimensional foraging model

■ 3.1 General assumptions

We consider a forager in one dimension by first defining the search space as the set of all real numbers, \mathbb{R} . That is, the forager can search anywhere along a single dimension. Targets are equispaced along this search space, a distance of λ apart. Figure 3.1.1 shows what this scenario may look like.

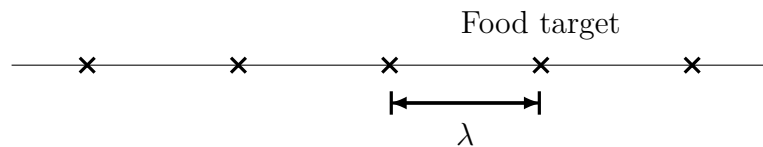


Figure 3.1.1: The search space for the one-dimensional foraging model. Food patches (crosses) are deterministically spaced along the search space, a distance of λ apart.

In reality, food patches will not usually be distributed deterministically and so more realistic one-dimensional models would allow food patches to be distributed along \mathbb{R} according to some probability distribution. Our assumption of deterministically spaced patches makes solving for the search efficiency easier, and we investigate the effects of making this assumption in Section 3.4.1.

Two types of food patch behaviours are considered: *destructive foraging*, in which the target is destroyed after the forager reaches it, and *non-destructive foraging*, in which the target is not destroyed and can be revisited an unlimited number of times.

Our model assumes that the forager has a radius of vision, r_v , with which it can sense any targets that are within r_v of its location. After a target enters the

forager’s radius of vision, the forager moves directly to the target. Initially, we consider a model where the forager has a constant radius of vision, but eventually extend our results in Section 3.3 to allow the radius of vision to switch based on the state of an underlying Markov chain.

The forager’s movement along the x -axis is governed by a random process. In Section 3.2, we consider a strategy in which the forager uses a random walk, with a general step-length distribution. This is extended to a Markov-modulated random walk strategy in Section 3.3. We are interested only in discrete-time random processes, in which the forager takes discrete “steps”. Our notion of a step corresponds to a “reorientation” step, in which the animal decides which direction and for what distance it should travel for its next movement, rather than steps representing physical steps made by the forager. We also disallow the forager from jumping over or skipping targets, and any step that would cause this to occur is truncated. In one dimension, this assumption is perfectly valid, as there is no way that an animal can pass a target without it at one point being at the target itself.

Our primary aim is to determine the most efficient foraging strategy, where the efficiency is given by

$$\eta = \frac{1}{\mathbb{E}[L]},$$

where L is the total distance travelled to find the first food patch. This notion of efficiency is common throughout the literature (e.g [6, 86]) and only considers the distance required to find a single patch. However, in the case of destructive foraging the food patches are becoming more sparse over time, and so we would expect the efficiency to be decreasing over time. Because of this, we investigate this notion of efficiency and compare it with others in Section 3.4.2.

Since we are only considering the distance travelled to find a single patch, we can reduce the search space to $[0, \lambda]$, where the two ends of this interval correspond to the targets. Figure 3.1.2 shows what this new model looks like. In the reduced model, the assumption of destructive targets corresponds to a search that starts at $\lambda/2$, and that of non-destructive targets corresponds to a search that starts at r_v . This choice of starting locations has been used throughout the literature (e.g. [6]), although we put this on more rigorous grounding in Section 3.4.2. For now, we provide a brief intuitive explanation of why this is the case. For a non-destructive forager, after reaching a target and beginning a new search, the target it just found is still available and so its search is beginning right next to a target, at r_v . For a destructive forager, without loss of generality assume that it finds the target at 0. Then, for its next search the two nearest patches are at $-\lambda$

and λ , and its location is halfway between them at 0. After scaling this becomes equivalent to starting at $\lambda/2$ on the interval $[0, \lambda]$.



Figure 3.1.2: The search space for a single search, which without loss of generality is the interval $[0, \lambda]$. The two food patches are at $x = 0$ and $x = \lambda$.

To determine the efficiency of a search, we shall first derive an expression for $\mathbb{E}[L]$. Although the overall aim is to determine an expression for the efficiency, we are also interested in determining other quantities for a search, such as the total number of steps taken. The derivation of many of these quantities is similar in structure, so rather than showing each separately, we first derive the total “cost” of a strategy, based on some general cost function for each step. Using this result, we can determine the efficiency and the total steps taken, among other things, with relative ease.

The derivation of the efficiency of the random walk strategy comes mostly from Bartumeus *et al.* [6], although some of the results were first found outside of the context of animal foraging [20, 21]. The reason for including the derivation in this thesis is twofold.

First, the derivation for the Markov-modulated random walk is fairly similar in structure to that of the random walk strategy. Some of the techniques that are used require conditions that, while not obvious in the Markov-modulated random walk case, are quite obviously true in the case of the random walk. Because of this, the paper by Bartumeus *et al.* [6] does not give these details for the random walk strategy. In particular, applying the Dirac delta function (Theorem 3.2.5) and showing the fact that the operator norm is less than unity (Lemma 3.2.11) are not treated by Bartumeus *et al.* [6]. We include these details that were missed as they help make the derivation for the Markov-modulated case clearer. Our notation is also reasonably different from Bartumeus *et al.* [6] in order to maintain consistency between the unmodulated and modulated strategies.

Second, the random walk strategy is simple enough to allow us a good insight into why many of the model assumptions are needed, and how relaxing any of these may be difficult. The notation becomes more complex when Markov-modulation is considered, and much of the intuition is lost.

■ 3.2 Random walk strategies

The first type of strategy we investigate is the random walk. Recall from Definition 2.1.14, that a random walk is a path of successive steps on some mathematical space. For the one-dimensional model in question, the mathematical space is the search interval $[0, \lambda]$, although since r_v is constant we may just consider the interval $[r_v, \lambda - r_v]$ as the entire search space. The forager moves according to a random walk, beginning at some known location $x_0 = a$, for $a \in [r_v, \lambda - r_v]$.

We denote the location of the forager after taking i steps as X_i , which is a random variable, or as x_i for a single realisation, for $i \geq 0$. The length of the i th step taken, $\ell_i = X_i - X_{i-1}$, is distributed according to some probability distribution, $\ell_i \sim p(\ell)$, which we call the *step-length distribution*. According to this definition, the length of a step is negative for steps to the left, and positive for steps to the right. We require $p(\ell)$ to be symmetric about zero, which gives the forager an equal chance of moving in either direction. The step length is drawn from a distribution with values from $(-\infty, \infty)$, but since the search space is finite, steps moving beyond the boundary will be truncated. Since a search ends once the forager has left the search interval and hence located food, we are not concerned with where the forager would have ended up after the final step, and instead are concerned with the cost attributed to the final step. Thus, we avoid having to reconstruct $p(\ell)$ to account for the truncation by instead adjusting the cost function (see Equation 3.2.2). We also require that the step-length distribution has a minimum step size, and optionally may have a maximum step size. That is, we consider step-length distributions that have $|\ell| \in [\ell_{\min}, \ell_{\max}]$, with $\ell_{\min} > 0$ since a forager must move every step, and $\ell_{\max} \geq \ell_{\min}$. For unbounded step-length distributions we have $\ell_{\max} \rightarrow \infty$. Once a food patch has been located, the forager stops moving and hence all future positions are the same, given by $X_\tau = X_{\tau+1} = \dots$, where τ is a stopping time corresponding to the time at which the first target is found, and so may take any positive integer value.

■ 3.2.1 Total cost of a random walk strategy

The total cost accumulated to reach a food patch is denoted $Q(x_0)$, where x_0 is the starting location of the random walk. The total cost is, of course, the sum of the cost of every step,

$$Q(x_0) = \sum_{n=0}^{\tau-1} q(X_n, X_{n+1}), \quad (3.2.1)$$

where $q(X_n, X_{n+1})$ is non-negative for all n . The cost of a single step, $q(X_n, X_{n+1})$, is a function of the start and end locations of a step, although not of the step

number itself. Recall that we must truncate the cost function for steps that travel beyond the boundaries, and so our cost function becomes

$$q(X_n, X_{n+1}) = \begin{cases} q^*(X_n, r_v) & \text{if } X_{n+1} < r_v, \\ q^*(X_n, X_{n+1}) & \text{if } r_v \leq X_{n+1} \leq \lambda - r_v, \\ q^*(X_n, \lambda - r_v) & \text{if } X_{n+1} > \lambda - r_v, \end{cases} \quad (3.2.2)$$

where $q^*(X_n, X_{n+1})$ is the function representing the *untruncated* cost of a step from X_n to X_{n+1} .

As an example, the untruncated cost function we use when considering the distance travelled on the n th step is

$$q^*(X_n, X_{n+1}) = |X_{n+1} - X_n|,$$

and hence the cost function will be

$$q(X_n, X_{n+1}) = \begin{cases} |r_v - X_n| & \text{if } X_{n+1} < r_v, \\ |X_{n+1} - X_n| & \text{if } r_v \leq X_{n+1} \leq \lambda - r_v, \\ |\lambda - r_v - X_n| & \text{if } X_{n+1} > \lambda - r_v. \end{cases}$$

We can rewrite Equation 3.2.1 to obtain an expression for the total cost where the summation limits do not depend on the random variable τ :

$$\begin{aligned} Q(x_0) &= \sum_{n=0}^{\tau-1} q(X_n, X_{n+1}) \\ &= \sum_{n=0}^{\infty} \mathbb{1}_{(\tau \geq n+1)} q(X_n, X_{n+1}). \end{aligned} \quad (3.2.3)$$

Lemma 3.2.1. *The expected value of the total cost is given by*

$$\mathbb{E}[Q(x_0)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda - r_v} \rho_n(x_n) \mathbb{E}[q(X_n, X_{n+1}) \mid X_n = x_n] dx_n, \quad (3.2.4)$$

where $\rho_n(x_n)$ is the probability density function for the forager's location after taking n steps.

Here, and throughout, we use an explicit notation for the dummy variable of integration, such as dx_n in order to assist with the interpretation of each expression.

Proof. Taking the expectation of (Equation 3.2.3) gives

$$\mathbb{E}[Q(x_0)] = \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{(\tau \geq n+1)} q(X_n, X_{n+1})]$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \rho_n(x_n) \mathbb{E} \left[\mathbb{1}_{(\tau \geq n+1)} q(X_n, X_{n+1}) \mid X_n = x_n \right] dx_n \\
&= \sum_{n=0}^{\infty} \int_{r_v}^{\lambda - r_v} \rho_n(x_n) \mathbb{E} [q(X_n, X_{n+1}) \mid X_n = x_n] dx_n,
\end{aligned}$$

where the second line follows from the law of total expectation, and the third line is due to the fact that $\mathbb{1}_{(\tau \geq n+1)}$ will be zero for all x_n outside of $[r_v, \lambda - r_v]$ and will be one for all x_n within this interval. \square

The expression for the average cost of a step, $\mathbb{E} [q(X_n, X_{n+1}) \mid X_n = x_n]$, will of course depend on our definition of cost. For all of the definitions that we are interested in, we can find this expectation explicitly (see Section 3.2.2). Thus, we now have only the term $\rho_n(x_n)$ to deal with.

The probability that a forager is at location x_n after taking n steps, is equivalent to the probability that the forager is at x_{n-1} on the previous step, before taking a step of length $x_n - x_{n-1}$, for any starting point x_{n-1} . That is,

$$\rho_n(x_n) = \int_{r_v}^{\lambda - r_v} \rho_{n-1}(x_{n-1}) p(x_n - x_{n-1}) dx_{n-1}. \quad (3.2.5)$$

If we apply this recursively, we obtain n integrals for $\rho_n(x_n)$. Since we have an infinite sum over n , this is going to be difficult to deal with. To make it easier, we define an integral operator.

Definition 3.2.2. We define the operator \mathcal{L} , which acts on any real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}^+$, as

$$[\mathcal{L}f](x_n) = \int_{r_v}^{\lambda - r_v} p(x_n - x_{n-1}) f(x_{n-1}) dx_{n-1},$$

where $p(x_n - x_{n-1})$ is the probability density of the step length.

In some circumstances we wish to treat the situation in which we have the initial condition $X_0 = a$, almost surely. Let ρ_a be the measure defining this initial condition and we explicitly define

$$[\mathcal{L}\rho_a](x_n) = \int_{r_v}^{\lambda - r_v} p(x_n - x_{n-1}) \delta(x_{n-1} - a) dx_{n-1} = p(x_n - a).$$

Then, \mathcal{L}^n represents the recursive application of the operator n times, that is, $[\mathcal{L}^k f](x) = [\mathcal{L}[\mathcal{L}^{k-1} f]](x)$, etc.

Theorem 3.2.3. *The operator \mathcal{L} is self-adjoint.*

Proof. Our operator \mathcal{L} is a Fredholm operator, from Equation 2.1.3, with kernel $p(x - x')$. Since the step-length distribution is symmetric, we get that $p(x - x') = p(x' - x)$ and hence the kernel is symmetric. Therefore, by Corollary 2.1.42, \mathcal{L} is self-adjoint. \square

By definition of the operator \mathcal{L} and Equation 3.2.5, we get the recursive relation

$$\rho_n(x_n) = [\mathcal{L}\rho_{n-1}](x_n) \quad \text{for } n \geq 0, x_n \in \mathbb{R}. \quad (3.2.6)$$

Lemma 3.2.4. *For any $n \geq 0$,*

$$\rho_n(x) = [\mathcal{L}^n \rho_0](x) \quad \text{for } x \in \mathbb{R}.$$

Proof. By induction. Using Equation 3.2.6 we know that the base case, $n = 1$, is true

$$[\mathcal{L}\rho_0](x) = \rho_1(x).$$

Then, for any $k > 1$ we assume

$$[\mathcal{L}^k \rho_0](x) = \rho_k(x),$$

and for $n = k + 1$ we get

$$[\mathcal{L}^{k+1} \rho_0](x) = [\mathcal{L}[\mathcal{L}^k \rho_0]](x) = [\mathcal{L}\rho_k](x) = \rho_{k+1}(x),$$

by Equation 3.2.5. \square

We can substitute this result into Equation 3.2.4 to get

$$\mathbb{E}[Q(x_0)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda-r_v} [\mathcal{L}^n \rho_0](x_n) \mathbb{E}[q(X_n, X_{n+1}) \mid X_n = x_n] dx_n.$$

Further, since the function $q(X_n, X_{n+1})$ does not depend on n itself but only on two physical points in space we get

$$\mathbb{E}[Q(x_0)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda-r_v} [\mathcal{L}^n \rho_0](x_n) \mathbb{E}[q(X_0, X_1) \mid X_0 = x_n] dx_n. \quad (3.2.7)$$

Since the forager's initial position is known to be $x_0 = a$, we can express the probability density function for the forager's initial position as

$$\rho_0(x_0) = \delta(x_0 - a),$$

where δ is the Dirac delta function, for which

$$\int_{\mathcal{A}} f(x)\delta(x)dx = \begin{cases} 0 & \text{if } 0 \notin \mathcal{A}, \\ f(0) & \text{if } 0 \in \mathcal{A} \setminus \partial\mathcal{A}, \\ f(0)/2 & \text{if } 0 \in \partial\mathcal{A}, \end{cases}$$

and constrained to satisfy the identity

$$\int_{-\infty}^{\infty} \delta(x)dx = 1.$$

Theorem 3.2.5. *The expected total cost of a random walk strategy starting at $x_0 = a$ is*

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} [\mathcal{L}^n h](a),$$

where $h(x) = \mathbb{E}[q(X_0, X_1) \mid X_0 = x]$.

We note that the proof of Theorem 3.2.5 follows fairly simply from the fact that \mathcal{L} is self-adjoint. However, the proof below is presented instead since it more closely matches the proof used for the Markov-modulated case.

Proof. We begin with Equation 3.2.7, and make the substitution $h(x) = \mathbb{E}[q(X_0, X_1) \mid X_0 = x]$, to obtain

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda-r_v} [\mathcal{L}^n \rho_0](x_n)h(x_n)dx_n.$$

First, we separate the $n = 0$ term from the rest of the summation. For the remaining terms in the summation, we separate out the final operator that will be applied,

$$\mathbb{E}[Q(a)] = \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0)h(x_0)dx_0 + \sum_{n=1}^{\infty} \int_{x_n=r_v}^{\lambda-r_v} [\mathcal{L}[\mathcal{L}^{n-1}\rho_0]](x_n)h(x_n)dx_n,$$

and then use the definition of \mathcal{L} to get

$$\begin{aligned} \mathbb{E}[Q(a)] &= \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0)h(x_0)dx_0 \\ &\quad + \sum_{n=1}^{\infty} \int_{x_n=r_v}^{\lambda-r_v} \int_{x_{n-1}=r_v}^{\lambda-r_v} p(x_n - x_{n-1}) [\mathcal{L}^{n-1}\rho_0](x_{n-1})dx_{n-1}h(x_n)dx_n. \end{aligned}$$

Continuing in this manner,

$$\begin{aligned} \mathbb{E}[Q(a)] &= \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0)h(x_0)dx_0 \\ &\quad + \sum_{n=1}^{\infty} \int_{x_n=r_v}^{\lambda-r_v} \cdots \int_{x_0=r_v}^{\lambda-r_v} \prod_{i=0}^{n-1} p(x_{i+1} - x_i)\rho_0(x_0)h(x_n)dx_i dx_n. \end{aligned}$$

We combine the two terms together again, and reverse the order of integrations,

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) \int_{x_1=r_v}^{\lambda-r_v} \cdots \int_{x_n=r_v}^{\lambda-r_v} \prod_{i=0}^{n-1} p(x_{i+1} - x_i) h(x_n) dx_n dx_i.$$

which we are allowed to do according to Theorem 2.1.10 and the extension we outlined in Corollary 2.1.11. We use the fact that both $p(x)$ and $\rho_0(x)$ are PDFs and therefore are non-negative and measurable, and we know that $h(x)$ is a measurable function since the conditional expectation of a random variable is also a random variable, and is non-negative since $q(X_i, X_{i+1})$ is defined to be non-negative.

Next, we replace $p(x_{i+1} - x_i)$ with $p(x_i - x_{i+1})$ for all $i = 0, \dots, n-1$ using the symmetry of the step-length distribution, to arrive at

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) \int_{x_1=r_v}^{\lambda-r_v} \cdots \int_{x_n=r_v}^{\lambda-r_v} \prod_{i=0}^{n-1} p(x_i - x_{i+1}) h(x_n) dx_n dx_i.$$

The inner integral over x_n may be rewritten using the definition of \mathcal{L} , giving

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) \int_{x_1=r_v}^{\lambda-r_v} \cdots \int_{x_{n-1}=r_v}^{\lambda-r_v} \prod_{i=0}^{n-2} p(x_i - x_{i+1}) [\mathcal{L}h](x_{n-1}) dx_{n-1} dx_i.$$

The next innermost integral, over x_{n-1} , can be rewritten to get

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) \int_{x_1=r_v}^{\lambda-r_v} \cdots \int_{x_{n-2}=r_v}^{\lambda-r_v} \prod_{i=0}^{n-3} p(x_i - x_{i+1}) [\mathcal{L}^2h](x_{n-2}) dx_{n-2} dx_i.$$

Continuing in this manner results in

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) [\mathcal{L}^n h](x_0) dx_0.$$

Now, we substitute in the Dirac delta function for the forager's initial location,

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \int_{x_0=r_v}^{\lambda-r_v} [\mathcal{L}^n h](x_0) \delta(x_0 - a) dx_0,$$

and evaluate this integral to get

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} [\mathcal{L}^n h](a), \tag{3.2.8}$$

which completes the proof. □

To simplify this expression further, we must first prove some results about the operator norm of \mathcal{L} . In particular, for Equation 3.2.8 to converge, we must show that \mathcal{L} has an operator norm less than unity, which we will build up to over the next few lemmas.

Lemma 3.2.6. *For step-length distributions with $\ell_{\max} > \lambda/2 - r_v$, $\|\mathcal{L}\|_{op} < 1$, where $\|\cdot\|_{op}$ is the operator norm.*

Proof. Recall from Definition 2.1.30 the definition of the operator norm is

$$\|\mathcal{L}\|_{op} = \sup \left\{ \frac{\|\mathcal{L}f\|}{\|f\|} : \|f\| \neq 0 \right\},$$

where, in this case, we choose $\|\cdot\|$ to be the uniform norm, given by

$$\|f\|_{\infty} = \sup\{|f(x)| : x \in (r_v, \lambda - r_v)\}.$$

Then,

$$\begin{aligned} [\mathcal{L}f](x) &= \int_{r_v}^{\lambda - r_v} p(x' - x)f(x')dx' \\ &\leq C \int_{r_v}^{\lambda - r_v} p(x' - x)dx', \end{aligned}$$

where $C := \|f\|_{\infty}$.

We can rewrite this integral as two integrals, representing steps to either the left or the right, giving us

$$[\mathcal{L}f](x) \leq C \left(\int_{\ell_{\min}}^{x - r_v} p(\ell)d\ell + \int_{\ell_{\min}}^{\lambda - r_v - x} p(\ell)d\ell \right).$$

If $\ell_{\max} \leq x - r_v$, then every possible step to the left must land within $[r_v, \lambda - r_v]$, and so the first integral will evaluate to 0.5. If $\ell_{\max} > x - r_v$ then there is a nonzero probability that a step to the left will land beyond the boundary and so the first integral will evaluate to something less than 0.5. Similarly, if $\ell_{\max} \leq \lambda - r_v - x$ the right integral will evaluate to 0.5, and if $\ell_{\max} > \lambda - r_v - x$ the right integral will evaluate to something less than 0.5. Thus, we get that

$$[\mathcal{L}f](x) \leq \begin{cases} C & \text{if } \ell_{\max} \leq x - r_v \text{ and } \ell_{\max} \leq \lambda - r_v - x, \\ C^* < C & \text{otherwise.} \end{cases}$$

Rearranging the conditions on ℓ_{\max} , we get

$$[\mathcal{L}f](x) \leq \begin{cases} C & \text{if } \ell_{\max} + r_v \leq x \leq \lambda - r_v - \ell_{\max}, \\ C^* < C & \text{otherwise.} \end{cases}$$

Note that if $\ell_{\max} > \lambda/2 - r_v$, then there are no values of x that satisfy $\ell_{\max} + r_v \leq x \leq \lambda - r_v - \ell_{\max}$. So, for $\ell_{\max} > \lambda/2 - r_v$,

$$[\mathcal{L}f](x) \leq C^* \text{ for all } x \in [r_v, \lambda - r_v],$$

which implies that

$$\|\mathcal{L}f\|_{\infty} \leq C^* < C.$$

Therefore, for $\ell_{\max} > \lambda/2 - r_v$,

$$\frac{\|\mathcal{L}f\|_{\infty}}{\|f\|_{\infty}} \leq \frac{C^*}{C} < 1,$$

and so by the definition of the operator norm we get $\|\mathcal{L}\|_{op} < 1$. \square

Using Lemma 3.2.6, we know that the operator norm is less than 1 for some step-length distributions with a large enough ℓ_{\max} value. Note that the condition on ℓ_{\max} in Lemma 3.2.6 is a sufficient condition and not a necessary condition. There may exist some distributions with $\ell_{\max} \leq \lambda/2 - r_v$ for which $\|\mathcal{L}\|_{op} < 1$. In fact, as the next few lemmas show, there definitely is. We now prove some results for the operator under more relaxed conditions of ℓ_{\max} .

Lemma 3.2.7. *For step-length distributions with $\ell_{\max} > \frac{\lambda - 2r_v}{2^k}$, for $k \geq 1$, $\|\mathcal{L}^k\|_{op} < 1$, where $\|\cdot\|_{op}$ is the operator norm.*

Proof. In the same manner as the proof of Lemma 3.2.6, we could write the integrals corresponding to $[\mathcal{L}^k f](x)$ and rearrange these. However, we can note that over k steps, the maximum possible distance that a forager can travel from its starting point in either direction is $k\ell_{\max}$. Thus, a forager beginning at some position x that is within $k\ell_{\max}$ of either boundary has a non-zero probability of reaching a food target. Then, as with the the proof of Lemma 3.2.6 we can write

$$[\mathcal{L}^k f](x) \leq \begin{cases} C & \text{if } k\ell_{\max} \leq x - r_v \text{ and } k\ell_{\max} \leq \lambda - r_v - x, \\ C^* < C & \text{otherwise,} \end{cases}$$

where, once again, $C := \|f\|_{\infty}$. Rearranging the conditions, we get

$$[\mathcal{L}^k f](x) \leq \begin{cases} C & \text{if } k\ell_{\max} + r_v \leq x \leq \lambda - r_v - k\ell_{\max}, \\ C^* < C & \text{otherwise.} \end{cases}$$

Then, if $\ell_{\max} > \frac{\lambda - 2r_v}{2^k}$, for some $k \geq 1$, then there are no values of x which satisfy $k\ell_{\max} + r_v \leq x \leq \lambda - r_v - k\ell_{\max}$. So if $\ell_{\max} > \frac{\lambda - 2r_v}{2^k}$, for some $k \geq 1$,

$$[\mathcal{L}^k f](x) \leq C^* \text{ for all } x \in [r_v, \lambda - r_v]$$

which implies that

$$\|\mathcal{L}f\|_\infty \leq C^* < C.$$

Therefore, for $\ell_{\max} > \frac{\lambda-2r_v}{2^k}$, for some $k \geq 1$,

$$\frac{\|\mathcal{L}^k f\|_\infty}{\|f\|_\infty} \leq \frac{C^*}{C} < 1,$$

and so by definition of the operator norm we get $\|\mathcal{L}^k\|_{op} < 1$. \square

As with Lemma 3.2.6, the condition on ℓ_{\max} in Lemma 3.2.7 is also a sufficient condition but not a necessary condition.

Remark 3.2.8. Lemma 3.2.6, Lemma 3.2.7 are also true if we instead use the operator norm induced by

$$\|f\|_1 = \int_{r_v}^{\lambda-r_v} |f(x)| dx.$$

Consider

$$\begin{aligned} \|\mathcal{L}f\|_1 &= \int_{r_v}^{\lambda-r_v} |[\mathcal{L}f](x)| dx \\ &= \int_{r_v}^{\lambda-r_v} \left| \int_{r_v}^{\lambda-r_v} p(x-x') f(x') dx' \right| dx \\ &= \int_{r_v}^{\lambda-r_v} f(x') \int_{r_v}^{\lambda-r_v} p(x-x') dx dx', \end{aligned} \quad (3.2.9)$$

where the order of integration is exchanged by the exact same justification as in the proof of Theorem 3.2.5. Then, $\|\mathcal{L}f\|_1$ will be strictly less than $\|f\|_1$ under the same conditions as Lemma 3.2.6, and for the same reasons. Lemma 3.2.7 follows similarly.

Lemma 3.2.9. *If $\|\mathcal{L}^k\|_{op} < 1$, for $k = 2^n$ where n is some positive integer, then $\|\mathcal{L}\|_{op} < 1$.*

Proof. We know that \mathcal{L} is self-adjoint by Theorem 3.2.3. Then using Corollary 2.1.45 we know that $\|\mathcal{L}^k\|_{op} = \|\mathcal{L}\|_{op}^k$. Then,

$$\|\mathcal{L}^k\|_{op} < 1 \iff \|\mathcal{L}\|_{op}^k < 1 \iff \|\mathcal{L}\|_{op} < 1,$$

which completes the proof. \square

Remark 3.2.10. Not only is Lemma 3.2.9 true for all values of ℓ_{\max} , but \mathcal{L} can actually be any self-adjoint linear operator, not necessarily the linear operator \mathcal{L} that we have defined. This is important because we shall also use Lemma 3.2.9 in Chapter 4.

Now, we can present a lemma that is a stronger version of Lemma 3.2.6, since it no longer has any requirements on ℓ_{\max} .

Lemma 3.2.11. *For the operator \mathcal{L} defined in Definition 3.2.2, $\|\mathcal{L}\|_{op} < 1$.*

Proof. First note that for any value of ℓ_{\max} , we can choose a value of k such that $\|\mathcal{L}^k\|_{op} < 1$ using Lemma 3.2.7, since $\frac{\lambda - 2r_v}{2^k}$ is decreasing to zero and $\ell_{\max} > 0$.

Then, we know that $\|\mathcal{L}^k\|_{op} < 1$ for every possible value of ℓ_{\max} and hence every possible step-length distribution. However, to use Lemma 3.2.9 we require a k^* of the form 2^n for some positive integer n . Thus, we let $k^* = \min\{2^n : 2^n \geq k\}$, for which we also have $\|\mathcal{L}^{k^*}\|_{op} < 1$ for every step-length distribution. Then, we can use Lemma 3.2.9 to conclude that $\|\mathcal{L}\|_{op} < 1$. \square

Theorem 3.2.12. *Equation 3.2.8 is a convergent Neumann series and can be rewritten as*

$$\mathbb{E}[Q(a)] = [(\text{Id} - \mathcal{L})^{-1}h](a), \quad (3.2.10)$$

where $h(x) = \mathbb{E}_{X_1|X_0}[q(X_0, X_1) \mid X_0 = x]$.

Proof. From Lemma 3.2.11, we know that the operator norm of \mathcal{L} is strictly less than unity. Therefore, by Theorem 2.1.33,

$$[(\text{Id} - \mathcal{L})^{-1}f] = \sum_{n=0}^{\infty} [\mathcal{L}^n f], \quad (3.2.11)$$

for any function f , where Id is the identity operator. Now, recall Equation 3.2.8 is

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} [\mathcal{L}^n h](a),$$

and so Equation 3.2.11 implies that

$$\mathbb{E}[Q(a)] = [(\text{Id} - \mathcal{L})^{-1}h](a),$$

which completes the proof. \square

It is worth noting that the same Neumann series may be applied to Section 3.2.1 to give

$$\mathbb{E}[Q(x_0)] = \int_{r_v}^{\lambda - r_v} [(\text{Id} - \mathcal{L})^{-1}\rho_0](x_n)h(x_n)dx_n. \quad (3.2.12)$$

After discretisation of the search space, both this expression and Equation 3.2.10 are sufficient to numerically determine the total cost. Although Equation 3.2.12 is not as simple analytically, there is no significant difference in computation time (see Chapter 4) when compared to Equation 3.2.10.

■ 3.2.2 Average total distance travelled

We now use the results of the previous section to determine the average total distance travelled by a forager undergoing a random walk strategy. Since we are concerned only with distance travelled, and not direction, we can take absolute values of the step size, giving us the untruncated cost function

$$q^*(X_i, X_{i+1}) := |X_{i+1} - X_i|,$$

where X_i is the position of the forager after taking i steps. Then, the cost function becomes

$$q(X_i, X_{i+1}) = \begin{cases} |r_v - X_i| & \text{if } X_{i+1} < r_v, \\ |X_{i+1} - X_i| & \text{if } r_v \leq X_{i+1} \leq \lambda - r_v, \\ |\lambda - r_v - X_i| & \text{if } X_{i+1} > \lambda - r_v. \end{cases} \quad (3.2.13)$$

Using this definition of cost, the total cost, $Q(x_0)$, represents the total distance travelled to detect a food patch. However, we may also want to consider the distance needed to actually reach the food after detecting it. Thus, the total distance travelled will be

$$L(x_0) = Q(x_0) + r_v,$$

since the forager detects the food a distance r_v away, and moves directly towards it. When taking the expected value, the constant term remains as a constant term

$$\mathbb{E}[L(x_0)] = \mathbb{E}[Q(x_0)] + r_v.$$

The inclusion of the constant term, r_v , is of little significance when determining the optimal efficiency, but we include it for completeness.

Recalling that the forager begins at $x_0 = a$, and using Equation 3.2.10, we get

$$\mathbb{E}[L(x_0)] = [(\text{Id} - \mathcal{L})^{-1}h](a) + r_v,$$

where $h(x) = \mathbb{E}_{X_1|X_0}[q(X_0, X_1) | X_0 = x]$.

Note, that this result is the same as the result from Bartumeus *et al.* [6], with some differences in notation. They have denoted the expected distance travelled on the i th step, given the step began at x_{i-1} , as $\langle |\ell_i(x_{i-1})| \rangle$, as opposed to our expression $h(x_{i-1})$. As mentioned earlier, there is no dependence on i and so this becomes $\langle |\ell| \rangle$. They have also not included the r_v term, and instead consider the distance travelled to *detect* a food patch, rather than to *reach* a food patch.

Making these notational changes results in

$$\langle L \rangle(a) = [(\text{Id} - \mathcal{L})^{-1} \langle |\ell| \rangle](a),$$

which matches Equation (24) found by Bartumeus *et al.* [6].

The expectation term, h , to which the operator is applied can be solved similarly to a normal expectation, but we must take into consideration the possibility of truncation. We consider distributions that have a parameter representing the minimum step-size, ℓ_{\min} , such as the power-law distribution. The existence of an $\ell_{\min} > 0$ was a crucial part of Lemma 3.2.11 and hence required to show that the Neumann series would converge. However, we can still consider step-length distributions with $\ell_{\min} = 0$ as long as $\ell_{\max} > \ell_{\min}$, where the strict inequality is now needed so that $\ell_{\max} \neq 0$ to ensure convergence.

When evaluating h , we must separate the integral into separate cases depending on whether or not truncation occurs, according to Equation 3.2.13. Thus, we get

$$\begin{aligned} h(a) &= \mathbb{E}[q(X_0, X_1) \mid X_0 = a] \\ &= \int_{-\infty}^{\infty} q(a, x)p(x - a)dx \\ &= \int_{-\infty}^{r_v} |r_v - a|p(x - a)dx + \int_{r_v}^{\lambda - r_v} |x - a|p(x - a)dx + \int_{\lambda - r_v}^{\infty} |\lambda - r_v - a|p(x - a)dx. \end{aligned}$$

The first integral represents steps to the left that are truncated at the boundary, the second integral represents steps that do not reach a boundary and so are not truncated, and the third integral represents steps to the right that are truncated. We now break the middle integral into two integrals representing steps that are not truncated, and move to either the left or to the right, and also introduce a minimum step-size $\ell_{\min} \geq 0$, so

$$\begin{aligned} h(a) &= \int_{-\infty}^{r_v} |r_v - a|p(x - a)dx + \int_{r_v}^{a - \ell_{\min}} |x - a|p(x - a)dx \\ &\quad + \int_{a + \ell_{\min}}^{\lambda - r_v} |x - a|p(x - a)dx + \int_{\lambda - r_v}^{\infty} |\lambda - r_v - a|p(x - a)dx. \end{aligned}$$

Note that this is only valid for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$, since otherwise the other limits of integration will also need to be adjusted. These cases are considered in Section A.1, as well as consideration of a maximum step-size, $\ell_{\max} \geq \ell_{\min}$, and obtaining simplified expressions for h for a range of different step-length distributions. Finally, we rearrange the integrals to give

$$\begin{aligned} h(a) &= \int_{r_v}^{a - \ell_{\min}} (a - x)p(x - a)dx + \int_{a + \ell_{\min}}^{\lambda - r_v} (x - a)p(x - a)dx \\ &\quad + (a - r_v) \int_{-\infty}^{r_v} p(x - a)dx + (\lambda - r_v - a) \int_{\lambda - r_v}^{\infty} p(x - a)dx, \quad (3.2.14) \end{aligned}$$

which is valid for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$.

By substituting $\ell = x - a$, we can write Equation 3.2.14 as

$$h(a) = \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} \ell p(\ell) d\ell \\ + (a - r_v) \int_{-\infty}^{-(a-r_v)} p(\ell) d\ell + (\lambda - r_v - a) \int_{\lambda-r_v-a}^{\infty} p(\ell) d\ell. \quad (3.2.15)$$

The key difference between these two representations is that the ranges of integration in Equation 3.2.14 represent the locations where the forager steps to, whereas in Equation 3.2.15 they represent the distance that the forager travels.

If we denote the cumulative distribution function (CDF) of the step-length as $F_\ell(x)$, then using the symmetry of our step-length distribution we can rewrite Equation 3.2.15 as

$$h(a) = \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} \ell p(\ell) d\ell \\ + (a - r_v) F_\ell(-(a - r_v)) + (-\lambda - r_v - a) F_\ell(\lambda - r_v - a).$$

Our final result for the average total step length, Section 3.2.2, cannot be solved analytically for most choices of step-length distribution. Therefore, we discuss how it can be solved numerically in Chapter 4.

■ 3.2.3 Average number of steps

Previously we let τ be the time until a target was first found. Now we want to find $\mathbb{E}[\tau]$, and we note that in this context τ is usually denoted N . We can now use Theorem 3.2.12, which gives the expression for the expected total cost, to determine $\mathbb{E}[\tau]$. In this case, each step regardless of start or end location has a cost of 1. We also do not need to consider the constant term, r_v , as we did in determining the distance, since it is assumed that after detecting the food patch the animal moves directly to it without reorientation, and therefore incurs no further cost.

Once again, using Equation 3.2.10 from Section 3.2.1 we obtain

$$\mathbb{E}[\tau(a)] = [(\text{Id} - \mathcal{L})^{-1} \mathbf{1}],$$

where the ‘1’ is included just to make it clear that the operator is applied to a constant 1. As with the average total distance travelled, this expression cannot be solved analytically so we discuss solving this numerically in Chapter 4.

■ 3.3 Markov-modulated random walk strategies

As discussed in Chapter 1, by considering a Markov-modulated random walk strategy, we can generalise many of the other search strategies that have been investigated, as well as consider other new strategies.

We begin by considering a Markov-modulated random walk strategy, which is similar to the random walk strategy from Section 3.2, but now the step-length distribution of the random walk depends on the state of a discrete-time Markov chain, Z , with a state space that we denote as \mathcal{S} . We assume that the change of state of Z occurs after a step has already been finished, but before the next step begins. We say J is the number of states in \mathcal{S} , that is $|\mathcal{S}| = J$, where J is finite, and let Z_n be the state of Z after taking n steps (equivalently at the beginning of the $n + 1$ th step). Thus, for each state $j = 1, \dots, J$, we have a corresponding probability distribution $p_j(\ell)$ from which the length of a step is determined, rather than a single distribution as in the random walk strategy. We also allow for the forager's radius of vision to change depending on the state of Z , which we denote as r_j while Z is state j . We denote the transition matrix of Z as $P(x)$, where x is the current location of the forager. From Lemma 3.3.11 onwards, to proceed any further we are forced to make the assumption that the transition matrix does not depend on the current location, that is, $P(x) = P(y)$ for all $x, y \in [0, \lambda]$, in which case we can denote the transition matrix as P .

The more general case, with the transition matrix depending on x , would allow us to model location-aware foragers, such as done by Nolting [56]. Our first expression for the expected total cost, Theorem 3.3.8, allows for P to be a function of x .

As before, we are considering an overall search space $[0, \lambda]$. We maintain the meaning of X_i , x_i , and ℓ_i from Section 3.2. We assume that the forager begins at a known location $x_0 = a$, for some $a \in (0, \lambda)$. However, we do not make any assumption about what state the underlying Markov chain, Z , begins in. We are not able to reduce the search space to $[r_v, \lambda - r_v]$ since the radius of vision is not a constant as in the unmodulated scenario.

Once again, we begin by deriving an expression for the expectation of the total cost, and then show how this can be used to determine the average distance travelled and the average number of steps taken. Ideally, we would like an expression for the expectation of the total cost incurred in any given state, which, by summation over each state, will also give us an expression for the expectation of the total cost overall. We achieve this in the following section, although the

expression for the total cost incurred in a single state (Theorem 3.3.8) does not incorporate the Dirac delta for the starting location, and the final expression (Equation 3.3.13) which incorporates the Dirac delta can only be used to find the total cost incurred in all states, and not a single state individually.

■ 3.3.1 Total cost of a Markov-modulated strategy

We define $Q_j(x_0)$ to be the total cost that is accumulated in state j to detect a food target. This is the sum of the costs of all steps taken while in state j .

We define $q_j^*(x_n, x_{n+1})$ to be the *untruncated* cost function for a step at the *beginning* of which Z is in state j and the forager begins at location x_n , and ends at location x_{n+1} . Then, as in the unmodulated case, we can define the true cost function, for each $j \in \mathcal{S}$, as

$$q_j(X_n, X_{n+1}) = \begin{cases} q_j^*(X_n, r_j) & \text{if } X_{n+1} < r_j, \\ q_j^*(X_n, X_{n+1}) & \text{if } r_j \leq X_{n+1} \leq \lambda - r_j, \\ q_j^*(X_n, \lambda - r_j) & \text{if } X_{n+1} > \lambda - r_j. \end{cases}$$

The total cost accumulated in state j is then given by

$$Q_j(x_0) = \sum_{n=0}^{\tau-1} \mathbb{1}_{(Z_n=j)} q_j(X_n, X_{n+1}). \quad (3.3.1)$$

Due to the indicator function, a cost is attributed to a state only if the Markov chain begins in that state. Once again, τ is the stopping time representing the time until the forager first detects a target.

Then, the total cost accumulated regardless of state is

$$Q(x_0) = \sum_{j=1}^J Q_j(x_0).$$

We can rewrite Equation 3.3.1 to remove the random variable τ from the summation limits, by introducing another indicator function,

$$\begin{aligned} Q_j(x_0) &= \sum_{n=0}^{\tau-1} \mathbb{1}_{(Z_n=j)} q_j(X_n, X_{n+1}) \\ &= \sum_{n=0}^{\infty} \mathbb{1}_{(Z_n=j)} q_j(X_n, X_{n+1}) \mathbb{1}_{(\tau \geq n+1)}. \end{aligned} \quad (3.3.2)$$

Recall from the random walk strategy in Section 3.2, the probability density of the forager's location after taking n steps was denoted $\rho_n(x_n)$. For the Markov-modulated random walk, we must also take into account the current state of the

Markov chain Z , hence we define $\rho_{n,j}(x_n)$ as the probability density of the forager being at location x_n and Z being in state j after taking n steps ($Z_n = j$). Then we can define the vector

$$\boldsymbol{\rho}_n(x_n) = (\rho_{n,1}(x_n), \dots, \rho_{n,J}(x_n)). \quad (3.3.3)$$

Now, we present a lemma which may be considered the Markov-modulated analogue of Lemma 3.2.1.

Lemma 3.3.1. *The expected value of the total cost accumulated in the j th state is given by*

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{r_j}^{\lambda-r_j} \rho_{n,j}(x_n) \mathbb{E}[q_j(X_n, X_{n+1}) \mid X_n = x_n, Z_n = j] dx_n.$$

where $\rho_{n,j}(x_n)$ is the probability density function for the forager's location after taking n steps with $Z_n = j$.

Proof. Taking the expectation of Equation 3.3.2 gives

$$\begin{aligned} \mathbb{E}[Q_j(x_0)] &= \sum_{n=0}^{\infty} \mathbb{E} \left[\mathbb{1}_{(\tau \geq n+1)} \mathbb{1}_{(Z_n=j)} q_j(X_n, X_{n+1}) \right] \\ &= \sum_{n=0}^{\infty} \sum_{k \in \mathcal{S}} \int_{-\infty}^{\infty} \rho_{n,k}(x_n) \mathbb{E} \left[\mathbb{1}_{(\tau \geq n+1)} \mathbb{1}_{(Z_n=j)} \right. \\ &\quad \left. \times q_j(X_n, X_{n+1}) \mid X_n = x_n, Z_n = k \right] dx_n, \end{aligned}$$

where we have conditioned on both the forager's location after n steps, x_n , as well as the state of Z after n steps, Z_n . Now, realising that the second indicator function implies that the expression is zero unless the forager is in state j , and the first indicator function implies that the expression is zero unless the forager is within $(r_j, \lambda - r_j)$, we get

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{r_j}^{\lambda-r_j} \rho_{n,j}(x_n) \mathbb{E}[q_j(X_n, X_{n+1}) \mid X_n = x_n, Z_n = j] dx_n,$$

which completes the proof. \square

We use a new function $h_j(x)$ to denote this expectation term,

$$h_j(x) = \mathbb{E}[q_j(X_n, X_{n+1}) \mid X_n = x, Z_n = j],$$

and hence

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{r_j}^{\lambda-r_j} \rho_{n,j}(x_n) h_j(x_n) dx_n. \quad (3.3.4)$$

As in the original random walk strategy, we must now deal with the density function $\rho_{n,j}(x_n)$, which we do using operators.

Definition 3.3.2. We define the operator $\mathcal{L}_{i,j}$, for $i, j \in \mathcal{S}$, which acts on any real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}^+$, as

$$[\mathcal{L}_{i,j}f](x_n) = \int_{r_i}^{\lambda-r_i} p_i(x_n - x_{n-1})f(x_{n-1})P_{i,j}(x_n)dx_{n-1},$$

recalling that r_i is the forager's radius of vision while Z is in state i and P is the transition matrix of Z , and hence $P_{i,j}$ is the (i, j) th element of the transition matrix.

As with the unmodulated case, we may sometimes wish to treat the situation in which we have the initial condition $X_0 = a$, almost surely. Let ρ_a be the measure defining this initial condition and we explicitly define

$$[\mathcal{L}_{i,j}\rho_a](x_n) = \int_{r_v}^{\lambda-r_v} p_i(x_n - x_{n-1})\delta(x_{n-1} - a)P_{i,j}(x_n)dx_{n-1} = P_{i,j}(x_n)p_i(x_n - a).$$

This operator that we have defined, $\mathcal{L}_{i,j}$, is very similar in structure to the integral operator in Section 3.2.1 (Definition 3.2.2). The only differences are the different limits of integration and the inclusion of the $P_{i,j}(x_n)$ term, which accounts for the probability of Z switching states.

Definition 3.3.3. We define the operator \mathcal{L} , which acts on any positive, real-valued function $f : \mathbb{R} \rightarrow (\mathbb{R}^+)^{m \times J}$ with m being any positive integer, as

$$[\mathcal{L}f](x_n) = \left[\begin{array}{c} \left(\begin{array}{cccc} f_{1,1} & f_{1,2} & \cdots & f_{1,J} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,J} \end{array} \right) \left(\begin{array}{cccc} \mathcal{L}_{1,1} & \mathcal{L}_{1,2} & \cdots & \mathcal{L}_{1,J} \\ \mathcal{L}_{2,1} & \mathcal{L}_{2,2} & \cdots & \mathcal{L}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{J,1} & \mathcal{L}_{J,2} & \cdots & \mathcal{L}_{J,J} \end{array} \right) \\ \end{array} \right] (x_n),$$

where \mathcal{L} has dimension $J \times J$ and is applied to f in the same order that matrix multiplication is done. That is, the (i, j) th element of $[\mathcal{L}f](x_n)$ is given by $\sum_{k=1}^J [\mathcal{L}_{k,j}f_{i,k}](x_n)$.

When raised to a power, the innermost operator acts first, and continues in order from the inside to the outside, that is, $[\mathcal{L}^n f](x) = [\mathcal{L}[\mathcal{L}[\dots[\mathcal{L}f]\dots]](x)$.

The subscript on the $\mathcal{L}_{i,j}$ operators represent Z switching between state i to state j . The operator \mathcal{L} , being made up entirely of these $\mathcal{L}_{i,j}$ operators, essentially handles the transition matrix for Z , as well as the forager's location.

Lemma 3.3.4. *The operator \mathcal{L} , as defined in Definition 3.3.3, gives the recursive relation*

$$\rho_n(x_n) = [\mathcal{L}\rho_{n-1}](x_n) \text{ for } x_n \in \mathbb{R}.$$

Proof. First, we note that the density of the forager's location on the n th step, at the end of which Z has transitioned to state j can be expressed by the recursive relation

$$\rho_{n,j}(x_n) = \sum_{i=1}^J \int_{r_i}^{\lambda-r_i} p_i(x_n - x_{n-1}) \rho_{n-1,i}(x_{n-1}) P_{i,j}(x_n) dx_{n-1}, \quad (3.3.5)$$

which has the same justification as it does in the case of the random walk (see Equation 3.2.5), now with the extra term $P_{i,j}(x_n)$ representing any transition of Z that may end up in state j , which also gives us the sum over i .

The careful reader may observe that if $r_i < r_j$, then there is a possibility that the forager steps to a position $r_i < x_n < r_j$ (or $\lambda - r_j < x_n < \lambda - r_i$). This implies that $\rho_{n,j}(x_n)$ is non-zero, even though the food patch is within the forager's new radius of vision and hence has already been located, which seems to be an error. However, the integration at time $n + 1$, according to Equation 3.3.5, must be over $(r_j, \lambda - r_j)$, which excludes this value of x_n , and hence the forager still locates its food at the end of the n th step and terminates its search.

Equation 3.3.5 can be written with our operator as

$$\rho_{n,j}(x_n) = \sum_{i=1}^J [\mathcal{L}_{i,j} \rho_{n-1,i}](x_n), \quad (3.3.6)$$

from Definition 3.3.2. Recall from Equation 3.3.3 that

$$\boldsymbol{\rho}_{n-1}(x_n) = (\rho_{n-1,1}(x_n), \dots, \rho_{n-1,J}(x_n)).$$

Applying the operator \mathcal{L} to this vector,

$$[\mathcal{L} \boldsymbol{\rho}_{n-1}](x_n) = \begin{bmatrix} \left(\rho_{n-1,1} & \rho_{n-1,2} & \cdots & \rho_{n-1,J} \right) & \begin{pmatrix} \mathcal{L}_{1,1} & \mathcal{L}_{1,2} & \cdots & \mathcal{L}_{1,J} \\ \mathcal{L}_{2,1} & \mathcal{L}_{2,2} & \cdots & \mathcal{L}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{J,1} & \mathcal{L}_{J,2} & \cdots & \mathcal{L}_{J,J} \end{pmatrix} \end{bmatrix} (x_n),$$

and multiplying out gives

$$[\mathcal{L} \boldsymbol{\rho}_{n-1}](x_n) = \left(\sum_{k=1}^J [\mathcal{L}_{k,1} \rho_{n-1,k}](x_n), \dots, \sum_{k=1}^J [\mathcal{L}_{k,J} \rho_{n-1,k}](x_n) \right).$$

Finally, applying Equation 3.3.6 results in

$$[\mathcal{L} \boldsymbol{\rho}_{n-1}](x_n) = (\rho_{n,1}(x_n), \dots, \rho_{n,J}(x_n)) = \boldsymbol{\rho}_n(x_n),$$

which completes the proof. \square

Lemma 3.3.5. *The operator \mathcal{L} gives the expression*

$$\rho_n(x) = [\mathcal{L}^n \rho_0](x),$$

for any $n \geq 0$, and $x \in \mathbb{R}$.

Proof. This is by induction, following analogous arguments to the proof of Lemma 3.2.4. \square

We are now able to use these results to simplify Equation 3.3.4 to

$$\mathbb{E} [Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{r_j}^{\lambda - r_j} [\mathcal{L}^n \rho_0]_j(x_n) h_j(x_n) dx_n. \quad (3.3.7)$$

Just as for the random walk strategy, we would like to rewrite the infinite sum of \mathcal{L}^n as $(\text{Id} - \mathcal{L})^{-1}$ using a Neumann series. Thus, we now prove some properties of the operator \mathcal{L} . In particular, we show that the operator norm of \mathcal{L} is less than 1. Recall that for the unmodulated random walk we proved that $\|\mathcal{L}\|_{op} < 1$ with Lemma 3.2.6, Lemma 3.2.7, Lemma 3.2.9, Lemma 3.2.11. Rather than prove the Markov-modulated case across multiple theorems again, we can instead rely on some of the results of the unmodulated case to simplify the proof.

Definition 3.3.6. We define for a function $f : \mathbb{R} \rightarrow (\mathbb{R}^+)^{m \times J}$, a norm $\|\cdot\|$ given by

$$\|f\| = \max_{i=1, \dots, m} \sum_{j=1}^J \int_0^\lambda |f_{i,j}(x)| dx.$$

This is essentially the maximum column sum of the matrix, where we are also taking the L^1 -norm for each element, since f is a matrix-valued function. The L^1 -norm for functions was discussed in Remark 3.2.8 following Lemma 3.2.6, so we will be able to rely on some previous results.

Lemma 3.3.7. $\|\mathcal{L}\|_{op} < 1$, where $\|\cdot\|_{op}$ is the operator norm given by

$$\|\mathcal{L}\|_{op} = \sup \left\{ \frac{\|\mathcal{L}f\|}{\|f\|} : \|f\| \neq 0 \right\},$$

with $\|\cdot\|$ taken from Definition 3.3.6.

Proof. For some matrix function f , we have

$$\|f\| = \max_{i=1, \dots, m} \sum_{k=1}^J \int_0^\lambda |f_{i,k}(x)| dx =: C.$$

Consider

$$[\mathcal{L}f](x) = \begin{pmatrix} \sum_{k=1}^J [\mathcal{L}_{k,1}f_{1,k}](x) & \sum_{k=1}^J [\mathcal{L}_{k,2}f_{1,k}](x) & \cdots & \sum_{k=1}^J [\mathcal{L}_{k,J}f_{1,k}](x) \\ \sum_{k=1}^J [\mathcal{L}_{k,1}f_{2,k}](x) & \sum_{k=1}^J [\mathcal{L}_{k,2}f_{2,k}](x) & \cdots & \sum_{k=1}^J [\mathcal{L}_{k,J}f_{2,k}](x) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^J [\mathcal{L}_{k,1}f_{m,k}](x) & \sum_{k=1}^J [\mathcal{L}_{k,2}f_{m,k}](x) & \cdots & \sum_{k=1}^J [\mathcal{L}_{k,J}f_{m,k}](x) \end{pmatrix},$$

which we now take the norm of to give

$$\begin{aligned} \|\mathcal{L}f\| &= \max_{i=1,\dots,m} \sum_{j=1}^J \int_0^\lambda |[\mathcal{L}f](x)_{i,j}| dx \\ &= \max_{i=1,\dots,m} \sum_{j=1}^J \int_0^\lambda \left| \sum_{k=1}^J [\mathcal{L}_{k,j}f_{i,k}](x) \right| dx \\ &= \max_{i=1,\dots,m} \sum_{j=1}^J \int_0^\lambda \left| \sum_{k=1}^J \int_{r_k}^{\lambda-r_k} p_k(x-x') f_{i,k}(x') P_{k,j}(x) dx' \right| dx \\ &= \max_{i=1,\dots,m} \sum_{j=1}^J \sum_{k=1}^J \int_0^\lambda \int_{r_k}^{\lambda-r_k} p_k(x-x') f_{i,k}(x') P_{k,j}(x) dx' dx \\ &\leq \max_{i=1,\dots,m} \sum_{j=1}^J \sum_{k=1}^J \int_0^\lambda \int_0^\lambda p_k(x-x') f_{i,k}(x') P_{k,j}(x) dx' dx, \end{aligned}$$

where the final line follows from the fact that everything inside the integrals is positive. Rearranging this further gives

$$\begin{aligned} \|\mathcal{L}f\| &\leq \max_{i=1,\dots,m} \sum_{j=1}^J \sum_{k=1}^J \int_0^\lambda P_{k,j}(x) \int_0^\lambda p_k(x-x') f_{i,k}(x') dx' dx \\ &= \max_{i=1,\dots,m} \sum_{k=1}^J \int_0^\lambda \int_0^\lambda p_k(x-x') f_{i,k}(x') dx' dx =: C^*. \end{aligned} \quad (3.3.8)$$

Recall the unmodulated operator \mathcal{L} from Definition 3.2.2, which applied to a function $f_{i,k}$ gives

$$[\mathcal{L}f_{i,k}](x) = \int_{r_v}^{\lambda-r_v} p(x-x') f_{i,k}(x') dx',$$

where $p(x)$ is some step-length distribution, and r_v is any non-negative constant with $r_v \leq \lambda$. From Lemma 3.2.11 and Remark 3.2.8 we know that $\|\mathcal{L}f_{i,k}\|_1 < \|f_{i,k}\|_1$ for any $i = 1, \dots, m$ and $k = 1, \dots, J$, which means

$$\int_{r_v}^{\lambda-r_v} \int_{r_v}^{\lambda-r_v} p(x-x') f_{i,k}(x') dx' dx < \int_{r_v}^{\lambda-r_v} f_{i,k}(x) dx.$$

This inequality is still true for $r_v = 0$ and for any choice of step-length distribution, so we can substitute this into Equation 3.3.8 to get

$$\max_{i=1,\dots,m} \sum_{k=1}^J \int_0^\lambda \int_0^\lambda p_k(x-x') f_{i,k}(x') dx' dx < \max_{i=1,\dots,m} \sum_{k=1}^J \int_{r_v}^{\lambda-r_v} f_{i,k}(x) dx,$$

which implies $C^* < C$. Thus, we get that

$$\frac{\|\mathcal{L}f\|_1}{\|f\|_1} \leq \frac{C^*}{C} < 1 \implies \|\mathcal{L}\|_{op} < 1,$$

which completes the proof. □

Theorem 3.3.8. *Equation 3.3.7 is a convergent Neumann series, and so the expectation of the total cost accumulated in state j can be expressed as*

$$\mathbb{E}[Q_j(x_0)] = \int_{r_j}^{\lambda-r_j} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) h_j(x_n) dx_n.$$

Proof. From Lemma 3.3.7, we know that the operator norm of \mathcal{L} is strictly less than unity. Therefore, by Theorem 2.1.33,

$$[(\text{Id} - \mathcal{L})^{-1} f] = \sum_{n=0}^{\infty} [\mathcal{L}^n f], \quad (3.3.9)$$

for any function f , where Id is the identity operator. Recall Equation 3.3.7 was given by

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{r_j}^{\lambda-r_j} [\mathcal{L}^n \boldsymbol{\rho}_0]_j(x_n) h_j(x_n) dx_n.$$

The order of the integral and summation may be changed according to Corollary 2.1.13 since both $[\mathcal{L}^n \boldsymbol{\rho}_0]_j(x_n)$ and $h_j(x_n)$ are non-negative and measurable. Therefore, Equation 3.3.9 implies that

$$\mathbb{E}[Q_j(x_0)] = \int_{r_j}^{\lambda-r_j} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) h_j(x_n) dx_n. \quad \square$$

Just as we did for Equation 3.2.12, we note that Equation 3.3.7 may be used to numerically solve for $\mathbb{E}[Q_j(x_0)]$ after the discretisation of the search space. This expression allows us to solve for any initial location of the forager, $\boldsymbol{\rho}_0(x_0)$, which is demonstrated in Chapter 4. However, recall from the analysis of the unmodulated strategy that we were able to obtain a much simpler analytic expression involving only \mathcal{L} and h , using properties of the Dirac delta function, among other things. We proceed in a similar manner, in order to explain why an expression as simple as this does not quite work for the cost across a single state, rather works only for the cost over all states.

Recall the assumption that the initial location of a forager is known to be $x_0 = a$, but we do not assume anything about the initial state of Z . That is,

$$\boldsymbol{\rho}_0(x_0) = \delta(x_0 - a) \mathbf{z}_0,$$

where $\mathbf{z}_0 = (\Pr(Z_0 = 1), \dots, \Pr(Z_0 = J))$. This also means that

$$\rho_{0,k}(x_0) = \delta(x_0 - a) \Pr(Z_0 = k).$$

In the unmodulated case, the Dirac delta was applied in the proof of Theorem 3.2.5. The essential part of the proof was applying the operator \mathcal{L} to h in reverse. That is, the outermost application of \mathcal{L} to the density ρ became the innermost application to the function h , and vice versa. This reversal is somewhat obfuscated in the unmodulated case since all \mathcal{L} are equivalent. However, in the Markov-modulated strategy, each operator $\mathcal{L}_{i,j}$ corresponds to having a radius of vision r_i , and hence limits of integration, that depends on i . Thus, the order of application of these operators does matter. To deal with this issue, we define a new operator, which is a slightly modified version of the operator $\mathcal{L}_{i,j}$ from Definition 3.3.2.

Definition 3.3.9. We define the operator $\mathcal{L}_{i,j}^*$, for $i, j \in \mathcal{S}$, which acts on any positive, real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}^+$, as

$$[\mathcal{L}_{i,j}^* f](x_n) = \int_{r_j}^{\lambda - r_j} p_i(x_n - x_{n-1}) f(x_{n-1}) P_{i,j}(x_n) dx_{n-1},$$

recalling that r_j is the forager's radius of vision while Z is in state j and P is the transition matrix of Z , and hence $P_{i,j}$ is the (i, j) th element of the transition matrix.

The key difference between $\mathcal{L}_{i,j}$ from Definition 3.3.2 and the modified operator in Definition 3.3.9 is that the integration is now performed over the interval associated with j instead of i . Clearly, if the radius of vision is the same in both states, $r_i = r_j$, we get that $\mathcal{L}_{i,j}$ is equivalent to $\mathcal{L}_{i,j}^*$.

We then also define a new operator, \mathcal{L}^* , which is simply the matrix operator \mathcal{L} , but with all entries $\mathcal{L}_{i,j}$ replaced with $\mathcal{L}_{i,j}^*$.

Definition 3.3.10. We define the operator \mathcal{L}^* , which acts on any positive, real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}_+^{m \times J}$ with m being any positive integer, as

$$[\mathcal{L}^* f](x_n) = \left[\begin{array}{cccc} f_{1,1} & f_{1,2} & \cdots & f_{1,J} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,J} \end{array} \right] \left[\begin{array}{cccc} \mathcal{L}_{1,1}^* & \mathcal{L}_{1,2}^* & \cdots & \mathcal{L}_{1,J}^* \\ \mathcal{L}_{2,1}^* & \mathcal{L}_{2,2}^* & \cdots & \mathcal{L}_{2,J}^* \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{J,1}^* & \mathcal{L}_{J,2}^* & \cdots & \mathcal{L}_{J,J}^* \end{array} \right] (x_n),$$

where \mathcal{L}^* has dimension $J \times J$ and is applied to f in the same order that matrix multiplication is done.

In the special case of the radius of vision being constant across all states, we get that \mathcal{L} is equivalent to \mathcal{L}^* . Using these new definitions, we now present a lemma, which follows roughly the same procedure as the first part of the proof for Theorem 3.2.5.

Lemma 3.3.11. *If the transition matrix P is independent of the forager's location, then the expected total cost accumulated in state j of a Markov-modulated random walk starting at $x_0 = a$ can also be given by*

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J z_{0,k_0} \left[\mathcal{L}_{k_0,k_1}^* \cdots \left[\mathcal{L}_{k_{n-1},j}^* h_j \right] \cdots \right] (a),$$

where k_i represents the state of Z after taking i steps.

Proof. We begin with

$$\mathbb{E}[Q_j(x_0)] = \sum_{n=0}^{\infty} \int_{x_n=r_j}^{\lambda-r_j} [\mathcal{L}^n \rho_0]_j(x_n) h_j(x_n) dx_n. \quad (3.3.10)$$

We may use the definition of our operator \mathcal{L} to rewrite $[\mathcal{L}^n \rho_0]_j(x_n)$ as

$$\sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \left[\mathcal{L}_{k_{n-1},j} \left[\mathcal{L}_{k_{n-2},k_{n-1}} \cdots \left[\mathcal{L}_{k_0,k_1} \rho_{0,k_0} \right] \cdots \right] \right] (a),$$

and thus Equation 3.3.10 becomes

$$\begin{aligned} \mathbb{E}[Q_j(x_0)] &= \sum_{n=0}^{\infty} \int_{x_n=r_j}^{\lambda-r_j} h_j(x_n) \\ &\times \left[\sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \left[\mathcal{L}_{k_{n-1},j} \left[\mathcal{L}_{k_{n-2},k_{n-1}} \cdots \left[\mathcal{L}_{k_0,k_1} \rho_{0,k_0} \right] \cdots \right] \right] (x_n) \right] dx_n. \end{aligned}$$

We can expand out the integral using the definition of $\mathcal{L}_{i,j}$,

$$\begin{aligned} \mathbb{E}[Q_j(x_0)] &= \sum_{n=0}^{\infty} \int_{x_n=r_j}^{\lambda-r_j} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \int_{x_{n-1}=r_{k_{n-1}}}^{\lambda-r_{k_{n-1}}} \cdots \int_{x_0=r_{k_0}}^{\lambda-r_{k_0}} \prod_{i=0}^{n-1} [p_{k_i}(x_{i+1} - x_i) \\ &\times \rho_{0,k_0}(x_0) P_{k_i,k_{i+1}}(x_{i+1}) dx_i] h_j(x_n) dx_n. \end{aligned}$$

We rearrange the order of the sums and integrals to get,

$$\begin{aligned} \mathbb{E}[Q_j(x_0)] &= \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \int_{x_0=r_{k_0}}^{\lambda-r_{k_0}} \rho_{0,k_0}(x_0) \int_{x_1=r_{k_1}}^{\lambda-r_{k_1}} \cdots \int_{x_n=r_j}^{\lambda-r_j} \prod_{i=0}^{n-1} [p_{k_i}(x_i - x_{i+1}) \\ &\times P_{k_i,k_{i+1}}(x_{i+1})] h_j(x_n) dx_n dx_{n-1} \cdots dx_1 dx_0 \end{aligned}$$

where we have switched $p_{k_i}(x_{i+1} - x_i)$ for $p_{k_i}(x_i - x_{i+1})$, due to symmetry. It is at this point we require $P_{k_i,k_{i+1}}(x_i) = P_{k_i,k_{i+1}}(x_{i+1})$ for all $x_i, x_{i+1} \in [0, \lambda]$ if we

wish to simplify our expression any further. Thus, we use the assumption that the transition matrix P is independent of the forager's location.

Using the operator from Definition 3.3.9, we may rewrite the innermost integral of our expression, to get

$$\mathbb{E} [Q_j(x_0)] = \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \int_{x_0=r_{k_0}}^{\lambda-r_{k_0}} \rho_{0,k_0}(x_0) \int_{x_1=r_{k_1}}^{\lambda-r_{k_1}} \cdots \int_{x_{n-1}=r_{k_{n-1}}}^{\lambda-r_{k_{n-1}}} \prod_{i=0}^{n-2} p_{k_i}(x_i - x_{i+1}) P_{k_i, k_{i+1}} [\mathcal{L}_{k_{n-1}, j}^* h_j](x_{n-1}) dx_i.$$

In this same fashion, we may rewrite the innermost $n - 2$ integrals, giving

$$\mathbb{E} [Q_j(x_0)] = \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \int_{x_0=r_{k_0}}^{\lambda-r_{k_0}} \rho_{0,k_0}(x_0) \left[\mathcal{L}_{k_0, k_1}^* \cdots \left[\mathcal{L}_{k_{n-1}, j}^* h_j \right] \cdots \right] (x_0) dx_0.$$

Recall that the initial condition for $\boldsymbol{\rho}_0$ implies that $\rho_{0,k_0}(x_0) = \delta(x_0 - a) z_{0,k_0}$, and so we can integrate over the Dirac delta function, to get

$$\mathbb{E} [Q_j(x_0)] = \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J z_{0,k_0} \left[\mathcal{L}_{k_0, k_1}^* \cdots \left[\mathcal{L}_{k_{n-1}, j}^* h_j \right] \cdots \right] (a). \quad (3.3.11)$$

□

In analogue to the non-Markov-modulated strategy, we would like to write our expression only in terms of \mathcal{L}^* , \mathbf{h} , and \mathbf{z}_0 , evaluated at the initial location a , where $\mathbf{h}(x)$ is the row vector with i th element $h_i(x)$. Consider $\mathbf{h}^\top(x) \mathbf{z}_0$, which is a $J \times J$ matrix, with (i, j) th element $h_i(a) z_{0,j}$, and applying the operator \mathcal{L}^* to this matrix gives the matrix $[\mathcal{L}^*(\mathbf{h}^\top \mathbf{z}_0)](x)$, with (i, j) th element $\sum_{k=1}^J [\mathcal{L}_{k,j}^* h_i z_{0,k}](a)$. Repeatedly applying the operator \mathcal{L}^* gives,

$$\begin{aligned} & \sum_{n=0}^{\infty} \left[\mathcal{L}^{*n} (\mathbf{h}^\top \mathbf{z}_0) \right]_{jj} (a) \\ &= \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J \left[\mathcal{L}_{k_{n-1}, j}^* \left[\mathcal{L}_{k_{n-2}, k_{n-1}}^* \cdots \left[\mathcal{L}_{k_0, k_1}^* (z_{0,k_0} h_j) \right] \cdots \right] \right] (a), \end{aligned}$$

where the inner most operator is being applied to the function $z_{0,k_0} h_j(x)$. Since z_{0,k_0} is a constant it may be pulled out the front of all the operators due to their linearity,

$$\begin{aligned} & \sum_{n=0}^{\infty} \left[\mathcal{L}^{*n} (\mathbf{h}^\top \mathbf{z}_0) \right]_{jj} (a) \\ &= \sum_{n=0}^{\infty} \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J z_{0,k_0} \left[\mathcal{L}_{k_{n-1}, j}^* \left[\mathcal{L}_{k_{n-2}, k_{n-1}}^* \cdots \left[\mathcal{L}_{k_0, k_1}^* h_j \right] \cdots \right] \right] (a). \quad (3.3.12) \end{aligned}$$

We can see from Equation 3.3.12 that applying \mathcal{L}^{*n} to the matrix $\mathbf{h}^\top(x)\mathbf{z}_0$ produces the correct indexing for $z_{0,k_0}h_j$. However, the indexing for the $\mathcal{L}_{i,j}^*$ operators do not match that of Equation 3.3.11. Since all of the k_i terms are essentially dummy variables, the only index that matters in both Equation 3.3.11 and Equation 3.3.12 is the operator involving j . These are not in the same position, and so these expressions are not necessarily equivalent.

In fact, due to the way that the operators \mathcal{L} and \mathcal{L}^* are applied, there is no such expression involving \mathcal{L}^{*n} or \mathcal{L}^n and \mathbf{h} that will be equivalent to Equation 3.3.11.

If we instead consider the expectation of the total cost across all states, we no longer face this issue.

Theorem 3.3.12. *The expected total cost accumulated by a Markov-modulated random walk starting at $x_0 = a$, with transition matrix P independent to the forager's location is*

$$\mathbb{E}[Q(a)] = \text{Tr} \left(\sum_{n=0}^{\infty} [\mathcal{L}^{*n} (\mathbf{h}^\top \mathbf{z}_0)] (a) \right). \quad (3.3.13)$$

Proof. Beginning with the left hand side of Equation 3.3.13, we note that

$$\mathbb{E}[Q(a)] = \sum_{j=1}^J \mathbb{E}[Q_j(a)]$$

and so,

$$\mathbb{E}[Q(a)] = \sum_{n=0}^{\infty} \sum_{j=1}^J \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J z_{0,k_0} [\mathcal{L}_{k_0,k_1}^* \cdots [\mathcal{L}_{k_{n-1},j}^* h_j] \cdots] (a),$$

by Lemma 3.3.11. Now, considering the right hand side

$$\begin{aligned} \text{Tr} \left(\sum_{n=0}^{\infty} [\mathcal{L}^{*n} (\mathbf{h}^\top \mathbf{z}_0)] (a) \right) &= \sum_{j=1}^J \left[\sum_{n=0}^{\infty} [\mathcal{L}^{*n} (\mathbf{h}^\top \mathbf{z}_0)] (a) \right]_{jj} \\ &= \sum_{n=0}^{\infty} \sum_{j=1}^J \sum_{k_0=1}^J \cdots \sum_{k_{n-1}=1}^J z_{0,k_0} [\mathcal{L}_{k_{n-1},j}^* [\mathcal{L}_{k_{n-2},k_{n-1}}^* \cdots [\mathcal{L}_{k_0,k_1}^* h_j] \cdots]] (a), \end{aligned}$$

by Equation 3.3.12.

The expressions for the left-hand and right-hand side differ only by the indexing of the operators. Since the indices in both expressions are being summed over from 1 to J , they are all dummy variables, meaning the two expressions are equivalent. \square

Theorem 3.3.13. *Equation 3.3.13 is a convergent Neumann series, and so the expectation of the total cost can be expressed as*

$$\mathbb{E}[Q(a)] = \text{Tr} \left(\left[(\text{Id} - \mathcal{L}^*)^{-1} (\mathbf{h}^\top \mathbf{z}_0) \right] (a) \right).$$

Proof. This follows analogous arguments to those of the proof of Theorem 3.3.8. \square

Corollary 3.3.14. *The expectation of the total cost, for a Markov-modulated random walk with a constant radius of vision can be written as*

$$\mathbb{E}[Q(a)] = \text{Tr} \left(\left[(\text{Id} - \mathcal{L})^{-1} (\mathbf{h}^\top \mathbf{z}_0) \right] (a) \right).$$

Proof. When the radius of vision is constant across all states, $\mathcal{L}_{i,j}^*$ is equivalent to $\mathcal{L}_{i,j}$ for all $i, j \in \mathcal{S}$ and so \mathcal{L}^* is equivalent to \mathcal{L} . This is substituted directly into Equation 3.3.13. \square

In conclusion, for a Markov-modulated random walk, with the switching of both the step-length distribution and radius of vision, we have obtained an expression for the expectation of the total cost across a single state, Theorem 3.3.8. With P independent of the forager's location, we have also obtained an expression for the expectation of the total cost across all states, given a known starting location $x_0 = a$. Both of these expressions avoid the need for infinite summations through the use of convergent Neumann series.

■ 3.3.2 Average total distance travelled

As we did for the random walk strategy, we can now use our results for a generic cost function to determine the average total distance travelled for a Markov-modulated random walk strategy. We first look at determining the average total distance travelled over a single state.

Our untruncated cost function is defined as

$$q_j^*(X_n, X_{n+1}) := |X_{n+1} - X_n|$$

where X_n is the position of the forager after taking n steps. Then, the truncated cost is given by

$$q_j(X_n, X_{n+1}) = \begin{cases} |r_j - X_n| & \text{if } X_{n+1} < r_j, \\ |X_{n+1} - X_n| & \text{if } r_j \leq X_{n+1} \leq \lambda - r_j, \\ |\lambda - r_j - X_n| & \text{if } X_{n+1} > \lambda - r_j. \end{cases}$$

Recall, our function $h_j(x)$ was used to denote the expectation term,

$$h_j(x) = \mathbb{E}[q_j(X_0, X_1) \mid X_0 = x, Z_0 = j],$$

and hence, each element of $\mathbf{h}(x_0)$ can be solved by

$$\begin{aligned} h_j(x_0) &= \mathbb{E}[q_j(X_0, X_1) \mid X_0 = x_0, Z_0 = j] \\ &= \int_{r_j}^{a-\ell_{\min}} (a-x)p_j(x-a)dx + \int_{a+\ell_{\min}}^{\lambda-r_j} (x-a)p_j(x-a)dx \\ &\quad + (a-r_j) \int_{-\infty}^{r_j} p_j(x-a)dx + (\lambda-r_j-a) \int_{\lambda-r_j}^{\infty} p_j(x-a)dx, \end{aligned} \quad (3.3.14)$$

valid for $r_j + \ell_{\min} \leq a \leq \lambda - r_j - \ell_{\min}$. Equation 3.3.14 matches Equation 3.2.15 for the unmodulated case, with the difference being the radius of vision, r_j , and the step-length distribution now depending on j .

For any initial distribution, $\boldsymbol{\rho}_0$, using Theorem 3.3.8, the average total distance travelled in state j is

$$\mathbb{E}[L_j(x_0)] = \int_{r_j}^{\lambda-r_j} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) h_j(x_n) dx_n. \quad (3.3.15)$$

where h_j is determined by Equation 3.3.14.

When looking at the distance over a single state, we do not include the extra constant term r_v , since we do not attribute the final distance after detecting the food patch to any of the states.

If instead we are concerned with the total cost across all states, we can either sum Equation 3.3.15 across all states, or use Theorem 3.3.13, to give

$$\mathbb{E}[L(a)] = \text{Tr} \left([(\text{Id} - \mathcal{L}^*)^{-1} (\mathbf{h}^\top \mathbf{z}_0)] (a) \right),$$

where h_j is once again determined by Equation 3.3.14. However, we may want to include the distance that is travelled to the food patch after locating it. This will depend on what state the animal is in when it locates the food, since the radius of vision is not constant. Thus, we get

$$\mathbb{E}[L(x_0)] = \sum_{j=1}^J \left(\int_{r_j}^{\lambda-r_j} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) [h_j(x_n)] dx_n + P_j r_j \right),$$

where P_j is the probability that a target is detected during a step that has step-length distribution governed by state j . There is no obvious way to solve for this probability that is simple and does not require further use of the operator \mathcal{L} . Since the term makes such a small difference to the overall results, we do not solve P_j explicitly.

As with the random walk, the final result for the average total step length, either across a single state or across all states, cannot be solved analytically for any realistic choice of step-length distribution. Therefore, we discuss how it can be solved numerically in Chapter 4.

■ 3.3.3 Average number of steps

When considering the average number of steps taken by a Markov-modulated random walk, there is no possibility of truncation and so the cost function across a single state j is defined as

$$q_j(X_i, X_{i+1}) := \begin{cases} 1 & \text{if } Z_i = j \text{ and } X_i \neq X_{i+1} \\ 0 & \text{if } Z_i \neq j \text{ or } X_i = X_{i+1}, \end{cases}$$

since when $X_i = X_{i+1}$ a target has been found and no further cost is accumulated. Then, the j th element of the vector $\mathbf{h}(x)$ will be

$$h_j(x) = \mathbb{E}[q_j(X_0, X_1) \mid X_0 = x, Z_0 = j] = 1.$$

For the average number of steps in a state j , we get

$$\mathbb{E}[\tau_j(x_0)] = \int_{r_v}^{\lambda - r_v} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) dx_n.$$

For the average number of steps in total, we get

$$\mathbb{E}[\tau(a)] = \sum_{j=1}^J \mathbb{E}[N_j(a)],$$

or using Theorem 3.3.13,

$$\mathbb{E}[\tau(a)] = \text{Tr} \left([(\text{Id} - \mathcal{L}^*)^{-1} (\mathbf{1}^\top \mathbf{z}_0)](a) \right).$$

As with the average total distance travelled, this expression cannot be solved analytically so we discuss solving this numerically in Chapter 4.

■ 3.4 Analysis of the one-dimensional model assumptions

In constructing our one-dimensional foraging model, there were a number of assumptions that we had to make in order to make our model tractable. Many of the assumptions that we have made are prevalent throughout the animal foraging

literature. In particular, since our Markov-modulated foraging model is building upon the unmodulated foraging model of Bartumeus *et al.* [6], that is where many of our model's assumptions come from. Most of the assumptions that are made throughout foraging literature are made with some justification, although generally not very rigorously. In this section we investigate the effect of these assumptions on the results of the model, using numerical simulations.

We choose a random start location for the forager between $(0, \lambda)$, and simulate a one-dimensional random walk along the real line by sampling a random direction ($\theta = -1$ or $\theta = 1$) and a random step-length, using inverse-transform sampling. The restriction on jumping over targets, the truncation of steps, and the radius of vision are all treated the same as in the analytic model. Under destructive foraging conditions, after a target is found, it is removed from the set of possible targets. Under non-destructive foraging, a target is not removed from the set of possible targets, but rather, the animal continues its search a distance r_v away from the found target in either direction.

■ 3.4.1 Effects of the deterministically-spaced targets assumption

A realistic one-dimensional model should use targets that are randomly distributed along the real line, but in our analytic model we used deterministically-spaced targets. To test the effect of this assumption we compare the efficiency of searches for deterministic targets against searches for both uniformly and exponentially distributed targets.

Consider some sufficiently large integer, N (depending on the number of targets required to be found per search) to ensure that the animal cannot exit the search space. For the deterministic targets, we create a set of points $\{-N\lambda, -(N-1)\lambda, \dots, 0, \lambda, \dots, N\lambda\}$. For the uniformly distributed targets, we generate the same number of targets, $2N+1$, where the distance between successive targets is drawn from a uniform distribution on $[0, 2\lambda]$, which implies that the mean distance between targets will be λ , matching the mean in the case of the deterministic targets. Finally, we shift the position of all of the uniformly generated targets to recentre the middle target at $x = 0$. For the exponentially distributed targets, we do the same as we did for the uniformly distributed targets, although this time we generate the distances between targets using an exponential distribution with rate parameter $1/\lambda$, corresponding to a mean of λ .

We plot the mean efficiency and the 95% confidence intervals for a power-law

step-length strategy under destructive foraging, using 20 equispaced values across the range $1.1 \leq \mu \leq 3$ in Figure 3.4.1, each having different numbers of repetitions and targets found per search.

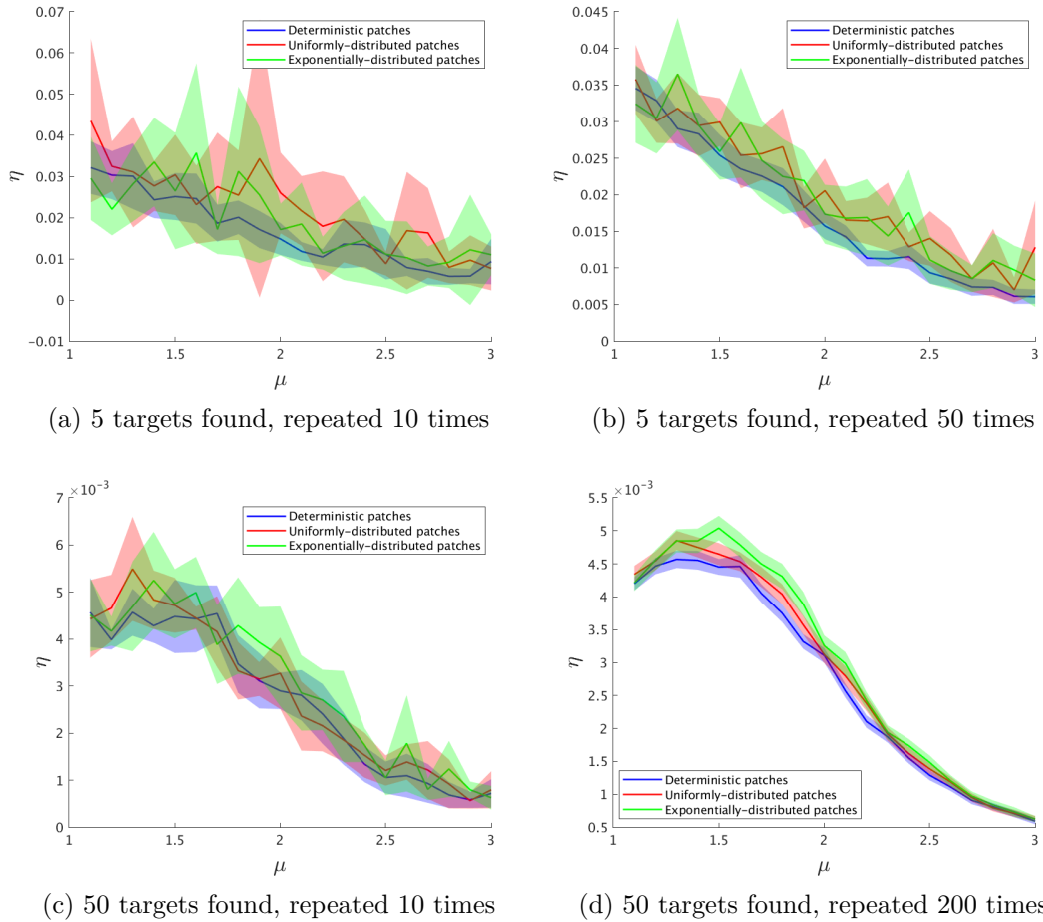


Figure 3.4.1: The mean efficiency of a random walk search strategy with a power-law step-length distribution, with parameter ranging from $1.1 \leq \mu \leq 3$ over 20 equispaced values, for destructive foraging ($x_0 = \lambda/2$) and with three different target distributions: deterministic, uniform, and exponential, each with an average distance between targets of $\lambda = 20$, and a radius of vision, $r_v = 1$. Each subplot ran the simulation for a different number of repetitions and had a different requirement on the number of targets to find.

Looking first at Figure 3.4.1a, there is clearly a large variance in the mean efficiency across values of μ for all target distributions. Comparing the distribution types across Figure 3.4.1a, Figure 3.4.1c, Figure 3.4.1b, the exponentially distributed targets seem to have the biggest variance in efficiency, with the uniform and deterministic target distributions seeming to have similar variances to each other.

The increase in the total targets found per search from Figure 3.4.1a to Figure 3.4.1c reduces the variances in our mean efficiency for all three target distribu-

tions, although there is still a large spike in the exponentially-distributed targets around $\mu \approx 1.6$. The increase from 5 targets found to 50 targets found also reduces the efficiency by a factor of 10. The efficiency seems to be very similar, if not the same, for the three different target distributions. Similarly, increasing the total number of times that the simulation is repeated from Figure 3.4.1a to Figure 3.4.1b smooths out the mean efficiency. The two key differences between repeating more searches versus running longer (in terms of targets) searches, is that when a new search begins the forager is reset to a random point within the first interval, and the targets are regenerated. Whereas after a target is found the animal resumes searching from where it found the target, and the targets become more scarce over time. Increasing the number of targets found seems to have a larger effect on reducing the variance than increasing the number of repetitions. Longer searches are also a more realistic model of the real world, since animals are generally not reset to a new location with a new set of targets.

Finally, in Figure 3.4.1d the number of repetitions is large enough that the efficiency curves are smooth and the confidence intervals are narrow. For $\mu \approx 1$, the efficiency is very similar between distributions, but as μ increases there is a clear difference in the efficiency between different target distributions. For values of $\mu \approx 3$, there is little if any difference in the efficiency between different target distributions. Therefore, the efficiency of a Lévy search depends on the distribution of targets, and for ballistic and Brownian searches there is not necessarily a dependence on the target distribution. However, the efficiency only changes slightly between different distributions, and the curves follow roughly the same shape, with the peak efficiency for each curve occurring close to each other. There are analytic models in the literature that allow for different target distributions (e.g. [6]), which we could potentially use to extend our model.

Although the deterministic target assumption will potentially affect the optimal parameters of any search strategy we investigate, the amount of difference between the target distributions seems small enough to justify making the simplification, especially given the relatively large variance in efficiency across multiple searches.

■ 3.4.2 Comparing the two different measures of efficiency

At the start of Chapter 3, we made the claim that optimising the efficiency of a search over the real line was equivalent to optimising the efficiency in finding the first target, given known starting conditions (either $x_0 = r_v$ or $x_0 = \lambda/2$). This claim is made throughout the animal foraging literature (e.g. [6]).

Intuitively, as more targets are found the search space becomes more sparse and

so the efficiency will decrease, which should mean that the efficiency over multiple targets is less than that of finding a single target. This can also be seen in Figure 3.4.1a, Figure 3.4.1c, Figure 3.4.1b, Figure 3.4.1d in the previous section, where the efficiency changed depending on how many targets were required to be found throughout each search. The two measures of efficiency are clearly not going to be equal to each other, although our claim is that optimising for either definition is equivalent, meaning that the maximum efficiency under both definitions occurs at the same location.

We define the efficiency of a search that finds N targets as

$$\eta_N = \frac{N}{L},$$

where L is the total distance travelled, and N is the number of targets found throughout the search, where the subscript N is used to make clear the dependence on the number of targets found.

Also recall the notion of efficiency used throughout Chapter 3,

$$\eta = \frac{1}{\mathbb{E}[L]}. \quad (3.4.1)$$

Note that under this definition of η , we are assuming a known starting point, as opposed to the η_N definition which assumes a random starting location. For the remainder of this section we include the dependence on the starting location explicitly. For destructive foraging,

$$\eta_d = \frac{1}{\mathbb{E}[L(\lambda/2)]},$$

where $L(x)$ represents the total distance to find a target for a search that begins at location x . For non-destructive foraging we have

$$\eta_n = \frac{1}{\mathbb{E}[L(r_v)]}.$$

The first case to consider is when the targets are non-destructive. Consider the efficiency across finding N targets, η_N , assuming that the forager starts at a random location between two targets, which may occur if we were to release an animal at random in some search space. Each time a target is reached, the animal begins its search a distance r_v away from the target in either direction. Since all intervals are statistically equivalent, and due to the symmetry of the step-length distribution, we can without loss of generality assume that the animal continues its search for the next target at location $x_0 = r_v$ in the interval $[0, \lambda]$. The search for the first target is unique in that it can begin at any point in the interval,

whereas the searches for each of the remaining $N - 1$ targets must begin at the same point, or a statistically equivalent point, since all searches will begin r_v from a target. If we define $L_i(r_v)$ as the distance travelled when finding the i th target, beginning from $x = r_v$, as , then we can say that $L_i(r_v)$ for $i \geq 2$ are independent and identically distributed. Using this notation, we may rewrite η_N as

$$\eta_N = \frac{N}{L_1(\lambda/2) + L_2(r_v) + \cdots + L_N(r_v)}.$$

When N is large, the random variable L_1 has a relatively small impact on the denominator, and so through the strong law of large numbers we know that

$$\lim_{N \rightarrow \infty} \eta_N = \frac{1}{\mathbb{E}[L_2(r_v)]} \text{ a.s.}$$

That is, for a non-destructive search, the long-term efficiency is the reciprocal of the expected distance travelled in a single search that begins at r_v . This is the same as our definition of η . We can consider η_N as the sample efficiency, which converges to η , the efficiency. In Figure 3.4.2 we plot both η and η_N for a range of N to further demonstrate this. The large N gets, the closer η_N gets to η .

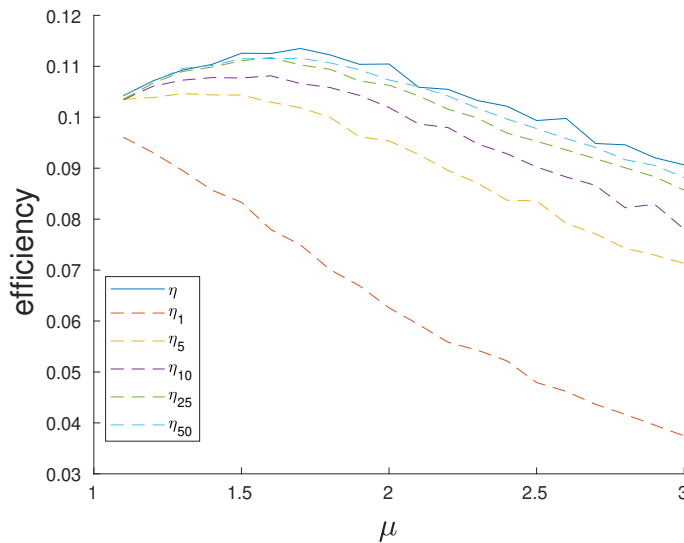


Figure 3.4.2: The efficiency, η , as defined in Equation 3.4.1 versus η_N plotted for a range of search lengths, $N = \{1, 5, 10, 25, 50\}$, for non-destructive foraging ($x_0 = r_v$ for η calculation) with $\lambda = 20$, $r_v = 1$, and $\ell_{\min} = 1$. Efficiency is averaged over 5000 simulations. The larger N is, the closer η_N is to η .

For the case of destructive foraging, an analytic argument is not as simple because as targets get more sparse, the starting location of the forager will depend on which targets have been destroyed. Instead, we simulate η_N and η for the destructive foraging case in Figure 3.4.3. Although η_N is clearly different to η , and there

does not seem to be convergence as N increases, we can still note that the peak efficiency occurs at the same location for both η and η_N .

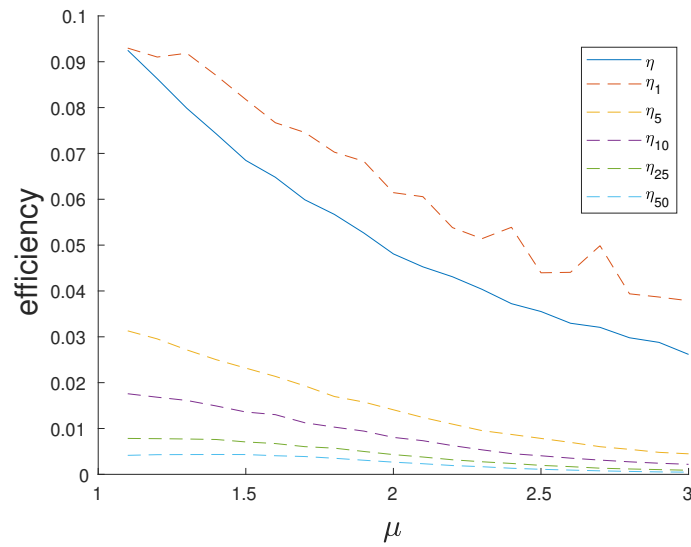


Figure 3.4.3: The efficiency, η , as defined in Equation 3.4.1 versus η_N plotted for a range of search lengths, $N = \{1, 5, 10, 25, 50\}$, for destructive foraging ($x_0 = \lambda/2$ for η calculation) with $\lambda = 20$, $r_v = 1$, and $\ell_{\min} = 1$. Efficiency is averaged over 5000 simulations. The only difference between η and η_1 is that the former begins at $x_0 = \lambda/2$ whereas the latter begins at a uniformly distributed location within $[0, \lambda]$.

Chapter 4

Discretisation of the one-dimensional search space

We stated at the end of Section 3.2.2, Section 3.2.3, Section 3.3.2, Section 3.3.3 that the exact analytic solutions we found for the expected total cost cannot feasibly be solved for any realistic choice of step-length distribution. In this section we outline how we may numerically solve these expressions via a spatial discretisation.

The discretisation of the unmodulated random walk in Section 4.2 once again can mostly be accredited to Bartumeus *et al.* [6]. However, we also derive a numerical approximation for Equation 3.2.12, an expression which avoids having to use properties of the Dirac delta function, and hence is valid for any distribution of starting locations. Bartumeus *et al.* [6] made the claim that the use of the Dirac delta was a crucial step in their derivation, but the existence of our expression — which has comparable computational efficiency — shows that this is not true.

The discretisation of the Markov-modulated random walk builds upon Section 4.2 in much the same way that Section 3.3 built upon Section 3.2. We are ultimately able to derive a simple expression for the cost of a Markov-modulated random walk, which amounts to solving a system of equations, which we solve using Matlab in Chapter 5. We also derive an alternative expression which avoids needing to know the initial location of a forager.

It is perhaps possible to argue that we could simply replace the functions and operators in the final analytic expressions with their discretised equivalents, resulting in the same numerical expressions found in this section. We opt, however, to demonstrate the derivation of the numerical expressions in full, to ensure correctness, as well as being able to shed some more light on some of the

issues found in the derivation of the analytic expressions, such as in deriving the cost for a single state of a Markov-modulated random walk.

■ 4.1 Notation and assumptions

The continuous search space $[0, \lambda]$ is replaced with a set of discrete positions at which the forager can be located. The positions are a distance Δx apart, and we denote the discretised point corresponding to $x = \lambda$ as M . Hence, the possible locations all take the form $j\Delta x$ with $j = 0, 1, \dots, M$. The discretisation length, Δx , is assumed to be much smaller than any of the original model parameters $(\ell_{\min}, \ell_{\max}, r_v, \lambda)$.

Then, discrete approximations of the model parameters are obtained with

$$\ell_{\min} = m_0\Delta x, \quad \ell_{\max} = m_m\Delta x, \quad r_v = m_r\Delta x, \quad m_0, m_m, m_r \in \mathbb{Z}.$$

The set of location variables $\{x_0, x_1, \dots, x_n\}$, which correspond to steps $\{0, 1, \dots, n-1, n\}$, correspond to steps ending at points $\{i_0, i_1, \dots, i_n\}$, where $x_m = i_m\Delta x$.

We use the notation $[\rho_n]_i$, $[h]_i$, and $\mathbb{E}[Q]_i$ to represent the functions $\rho_n(x)$ and $h(x)$, $\mathbb{E}[Q(x)]$ respectively, evaluated at the point $x = i\Delta x$. Then, $[\rho_n]$, $[h]$, and $\mathbb{E}[Q]$ are vectors used to denote the functions evaluated at every point in the discretised search space. Similarly, for the Markov-modulated case we use $[\rho_{n,j}]$, $[h_j]$, and $\mathbb{E}[Q_j]$ to denote the vector of discretised functions $\rho_{n,j}(x)$, $h_j(x)$, and $\mathbb{E}[Q_j(x)]$. We break with the convention of using boldface to denote vectors to be consistent with previous work [6], and since we use boldface to differentiate between the unmodulated and Markov-modulated versions of various quantities.

In discrete space, we represent the step-length distribution $p(\ell)d\ell$ using the matrix A . In continuous space we had $p(x_{m+1} - x_m)dx_m$, which corresponds to the probability of reaching location x_{m+1} in a step starting from x_m , on the $(m+1)$ th step. This is a jump of length $|x_{m+1} - x_m|$ which becomes $|i_{m+1} - i_m| \Delta x$ in the discretised space, which we denote as $[A]_{i_m, i_{m+1}}$. However, we must also take into account step-lengths that in continuous space are not integer multiples of Δx , which we do by integrating over all continuous values that will result in the same discrete value when discretised. Thus, we can find each element of A using

$$[A]_{j,k} = \int_{|j-k|\Delta x}^{(|j-k|+1)\Delta x} p(\ell)d\ell, \quad k \neq j. \quad (4.1.1)$$

Due to the symmetry of the jump probabilities, $[A]_{j,k} = [A]_{k,j}$ and hence the matrix A is symmetric. We require that every jump is greater than the minimum

jump size, and so $[A]_{j,k} = 0$ for $|j - k| < m_0$. For step-length distributions with a finite ℓ_{\max} , we also require $[A]_{j,k} = 0$, for $|j - k| \geq m_m$. Expressions for the matrix A based on Equation 4.1.1 are found for a variety of step-length distributions in Chapter A. Similarly, for the Markov-modulated case we define \mathbf{A} , which is made up of submatrices denoted A_i , each representing the step probabilities while Z is in state i . Thus, the elements of A_i are found in the same way as matrix A , with $p_i(\ell)$ and the corresponding values of m_0 and m_m .

In the discrete approximation, an integral over the continuous space becomes a sum over the discrete index. However, whether or not the end points are included in the summation is an important question which was not addressed in the continuous model, since the inclusion of endpoints make no difference when integrating. Our integrations over the continuous search space represented all of the possible points at which the forager was still searching for a target, and so the end points should not included, as these are where the forager would be able to detect the food. When the discrete radius of vision is m_r , the points from $m_r + 1$ to $(M - m_r - 1)$ are the locations where food hasn't yet been detected. Thus for an integral over $[r_v, \lambda - r_v]$, the equivalent summations are all from $m_r + 1$ to $M - m_r - 1$, rather than m_r to $M - m_r$. Thus, the effective search interval has a length of $(M - m_r - 1) - (m_r + 1) + 1 = M - 2m_r - 1$, and so the matrix A should have dimensions $(M - 2m_r - 1) \times (M - 2m_r - 1)$.

However, we instead define A to be an $(M - 1) \times (M - 1)$ matrix, by putting an outside layer of zeros — of width m_r — on each side. This way, the (i, j) th element of A will correspond to the probability of moving from the i th point in the search space to j th point. Similarly, the vector $[h]$ has a length of $M - 2m_r - 1$ for its non-zero elements, but we say $[h]$ has length $M - 1$, with m_r zeros on each end. Note that although we are only considering points within the search space, the forager may jump from within the search space to outside it, and thus the row sums of A will not necessarily equal 1. For the Markov-modulated case, the radius of vision may vary across states, so the dimensions of the non-zero elements of each A_i may vary across each state. Each matrix A_i will have the same dimensions as A , although the amount that the matrix needs to be padded with zeros will depend on the radius of vision in the corresponding state.

In deriving the analytic expressions for the expected total cost, we used the property of the Dirac delta function that

$$\int_{r_v}^{\lambda - r_v} \delta(x - a) dx = 1,$$

which in the discrete approximation becomes

$$\sum_{j=m_r+1}^{M-m_r-1} \delta_{j,i_a} = 1,$$

where δ_{j,i_a} is the Kronecker delta. The point i_a is the discretisation of the initial location $x_0 = a$, and so i_a is $a/\Delta x$ rounded to the nearest integer.

■ 4.2 Discretised expressions for a random walk strategy

Recall from Section 4.2 that we were able to write $\rho_n(x_n)$, the density function for the position of the animal at the end of the n th step, as an expression involving $\rho_{n-1}(x_{n-1})$,

$$\rho_n(x_n) = \int_{r_v}^{\lambda-r_v} \rho_{n-1}(x_{n-1}) p(x_n - x_{n-1}) dx_{n-1},$$

which applied recursively becomes

$$\rho_n(x_n) = \int_{x_{n-1}=r_v}^{\lambda-r_v} \cdots \int_{x_0=r_v}^{\lambda-r_v} \rho_0(x_0) \prod_{i=0}^{n-1} p(x_{i+1} - x_i) dx_i.$$

Replacing the integrals with their equivalent summations, we use this to find an expression for ρ_n in the discrete approximation, which is

$$[\rho_n]_{i_n} = \sum_{i_0=m_r+1}^{M-m_r-1} \cdots \sum_{i_{n-1}=m_r+1}^{M-m_r-1} [A]_{i_{n-1},i_n} [A]_{i_{n-2},i_{n-1}} \cdots [A]_{i_1,i_2} [A]_{i_0,i_1} [\rho_0]_{i_0},$$

where $[\rho_0]$ is the vector representing the function ρ_0 , as described in Section 4.1. Since the outer elements of the matrix A are 0, we can rewrite this as

$$[\rho_n]_{i_n} = \sum_{i_0=1}^{M-1} \cdots \sum_{i_{n-1}=1}^{M-1} [A]_{i_{n-1},i_n} [A]_{i_{n-2},i_{n-1}} \cdots [A]_{i_1,i_2} [A]_{i_0,i_1} [\rho_0]_{i_0}. \quad (4.2.1)$$

The sequence of products of elements of A in Equation 4.2.1 correspond to matrix products. Thus, we can rewrite it as

$$[\rho_n]_{i_n} = \sum_{i_0=1}^{M-1} [A^n]_{i_0,i_n} [\rho_0]_{i_0}. \quad (4.2.2)$$

Theorem 4.2.1. *The expectation of the total cost of a random walk in the discrete approximation is given by*

$$\mathbb{E}[Q] = \sum_{n=0}^{\infty} A^n [h], \quad (4.2.3)$$

where $\mathbb{E}[Q]$ is a vector of length $(M-1)$, with each element representing a different starting point for the forager.

Proof. Recall from Lemma 3.2.1, for the non-Markov-modulated random walk, the expectation of the total cost was given by

$$\mathbb{E}[Q(x_0)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda-r_v} \rho_n(x_n) h(x_n) dx_n,$$

where $h(x_n) = \mathbb{E}_{X_1|X_0}[q(X_1, X_0) | X_0 = x_n]$. In the discrete approximation, this becomes

$$\mathbb{E}[Q]_{i_0} = \sum_{n=0}^{\infty} \sum_{i_n=m_r+1}^{M-m_r-1} [\rho_n]_{i_n} [h]_{i_n} \Delta x,$$

where $\mathbb{E}[Q(x_0)]_{i_0}$ represents the expected total cost of a random walk beginning at point i_0 , and $[h]$ is the discretised vector of the function h , and thus is a vector with each element being given by the function h represented at a discrete point in the search space. Although there is no i_0 on the right-hand side of the equation, the dependence on i_0 , along with i_1, i_2, \dots is implicit in the vector $[\rho_n]$, as we shall next see.

Substituting in our expression for $[\rho_n]_{i_n}$ from Equation 4.2.2, as well as using the fact that the first m_r and last m_r elements of $[h]$ are 0, we get

$$\mathbb{E}[Q]_{i_0} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{i_0=1}^{M-1} [A^n]_{i_0, i_n} [\rho_0]_{i_0} [h]_{i_n} \Delta x. \quad (4.2.4)$$

Since the initial distribution is known to be $\rho_0(x) = \delta(x - a)$, and combining this with the Δx term, we get a Kronecker delta,

$$\mathbb{E}[Q]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{i_0=1}^{M-1} [A^n]_{i_0, i_n} \delta_{i_0, i_a} [h]_{i_n}.$$

Summing over the i_0 terms and using the corresponding property of the Kronecker delta we get

$$\mathbb{E}[Q]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} [A^n]_{i_a, i_n} [h]_{i_n},$$

which can be written as a matrix expression

$$\mathbb{E}[Q] = \sum_{n=0}^{\infty} A^n [h].$$

□

As mentioned above, the vector $[h]$ is found by evaluating $h(x)$ at every point in the discretised search space. For the average distance travelled we use Equation 3.2.15 or for the number of steps taken, we use $[h] = \mathbf{1}$.

Unfortunately, when evaluating Equation 4.2.3 we still have an infinite sum to deal with, and so evaluating this to a reasonable degree of accuracy may be slow since

we have to calculate each term in the summation until the result has converged sufficiently. As with our analytic expression, we can express this without the infinite summation by considering it as a Neumann series and using properties based on this. This time, instead of \mathcal{L} , we have matrix A as our operator. For the continuous case we split the proof over Lemma 3.2.6, Lemma 3.2.7, Lemma 3.2.9, Lemma 3.2.11, whereas we do this for the discrete approximation in a single combined lemma.

Lemma 4.2.2. $\|A\|_{op} < 1$, where $\|\cdot\|_{op}$ is the operator norm given by

$$\|A\|_{op} = \sup \left\{ \frac{\|Av\|}{\|v\|} : \|v\| \neq 0 \right\},$$

with $\|\cdot\|$ being the maximum column sum.

Proof. For some matrix v , with dimensions $(M-1) \times p$, where $p \in \mathbb{N}$, we use the norm given by

$$\|v\| = \max_{j=1,\dots,p} \sum_{i=1}^{M-1} |v_{i,j}| =: C.$$

Element (i, j) of the matrix Av is given by

$$[Av]_{i,j} = \sum_{k=1}^{M-1} A_{i,k} v_{k,j},$$

and so Av has norm

$$\begin{aligned} \|Av\| &= \max_{j=1,\dots,p} \sum_{i=1}^{M-1} \left| \sum_{k=1}^{M-1} |A_{i,k} v_{k,j}| \right| \\ &\leq \max_{j=1,\dots,p} \sum_{i=1}^{M-1} \sum_{k=1}^{M-1} |A_{i,k} v_{k,j}| \\ &= \max_{j=1,\dots,p} \sum_{k=1}^{M-1} |v_{k,j}| \sum_{i=1}^{M-1} |A_{i,k}|. \end{aligned}$$

where the second line has equality if and only if $v_{k,j} \geq 0$ for all k and j from 1 to $M-1$.

Now, consider the summation over $|A_{i,k}|$ individually. Recalling the definition of $A_{i,j}$ from Equation 4.1.1, and separating the summation into steps to either the left or the right, we can write

$$\begin{aligned} \sum_{i=1}^{M-1} |A_{i,k}| &= \sum_{i=1}^{k-1} A_{i,k} + \sum_{i=k+1}^{M-1} A_{i,k} \\ &= \sum_{i=1}^{k-1} \int_{|i-k|\Delta x}^{(|i-k|+1)\Delta x} p(\ell) d\ell + \sum_{i=k+1}^{M-1} \int_{|i-k|\Delta x}^{(|i-k|+1)\Delta x} p(\ell) d\ell, \end{aligned}$$

where the $i = k$ term in the summation is dropped since it is always 0. These integrals can be combined to give

$$\sum_{i=1}^{M-1} |A_{i,k}| = \int_{\Delta x}^{(k)\Delta x} |p(\ell)| d\ell + \int_{\Delta x}^{(M-k)\Delta x} |p(\ell)| d\ell.$$

Letting $x = k\Delta x$ and combining into a single integral, we get

$$\sum_{i=1}^{M-1} |A_{i,k}| \leq \int_0^\lambda p(x - x') dx' =: C^*,$$

with the inequality coming from the fact that there are some points between $(M-1)\Delta x$ and $M\Delta x$ that are being integrated over. The right hand side is the same as in the proof of Lemma 3.2.6, with $r_v = 0$. Following the same reasoning, the right hand side is less than or equal to 1 if $\ell_{\max} < x < \lambda - \ell_{\max}$ and strictly less than 1 otherwise. This tells us

$$\frac{\|Av\|}{\|v\|} \leq \frac{C^*}{C} < 1 \implies \|A\|_{op} < 1,$$

when $\ell_{\max} > \lambda/2$. Further, if we instead consider $A^n v$ for some $n \geq 1$, then

$$\begin{aligned} \|A^n v\| &= \max_{j=1, \dots, p} \sum_{i=1}^{M-1} \sum_{k_1=1}^{M-1} \sum_{k_2=1}^{M-1} \cdots \sum_{k_n=1}^{M-1} |A_{i,k_n} A_{k_n,k_{n-1}} \cdots A_{k_2,k_1} v_{k_1,j}| \\ &= \max_{j=1, \dots, p} \sum_{k_1=1}^{M-1} |v_{k_1,j}| \sum_{k_2=1}^{M-1} |A_{k_2,k_1}| \cdots \sum_{k_n=1}^{M-1} |A_{k_n,k_{n-1}}| \sum_{i=1}^{M-1} |A_{i,k_n}| \\ &\leq \max_{j=1, \dots, p} \sum_{k_1=1}^{M-1} |v_{k_1,j}| \sum_{k_2=1}^{M-1} |A_{k_2,k_1}| \cdots \sum_{k_n=1}^{M-1} |A_{k_n,k_{n-1}}| \int_0^\lambda p(x_i - x_{k_n}) dx_{k_n} \\ &\leq \sum_{k_1=1}^{M-1} \int_{r_v}^\lambda \cdots \int_0^\lambda \prod_{i=0}^{n-1} p(x_{i+1} - x_i) dx_i =: D \\ &\leq \max_{j=1, \dots, p} \sum_{k_1=1}^{M-1} |v_{k_1,j}| = \|v\|. \end{aligned}$$

Then, referring back to the $n = 1$ case above, as well as Lemma 3.2.7, we see that the only situation in which we may not have a strict inequality in the final line will be when $\ell_{\max} \geq \frac{\lambda}{2^k}$, and so we can conclude that

$$\frac{\|A^n v\|}{\|v\|} \leq \frac{D}{C} < 1 \implies \|A^n\|_{op} < 1$$

for $\ell_{\max} < \frac{\lambda}{2^k}$. We can choose an n such that $\|A^n\|_{op} < 1$ for every possible ℓ_{\max} . Then, we choose $n^* = 2^n$ and as we did in Lemma 3.2.9, we can show that $\|A^n\|_{op} < 1$ implies $\|A\|_{op} < 1$, where we have used the fact that A is a self-adjoint operator since it is symmetric. Thus, we conclude that $\|A\|_{op} < 1$ for every possible step-length distribution. \square

Theorem 4.2.3. *In the discrete approximation, the expected total cost of a random walk is given by*

$$\mathbb{E}[Q] = (\mathbb{I} - A)^{-1}[h],$$

where the i th element of the vector $\mathbb{E}[Q]$ represents the expected total cost of a walk that starts at point i .

Proof. Considering $\sum_{n=0}^{\infty} A^n$ as a Neumann series and using Lemma 4.2.2, we can write $\sum_{n=0}^{\infty} A^n$ as $(\text{Id} - A)^{-1}$, where Id is the identity operator, which in this case is the $(M - 1) \times (M - 1)$ identity matrix, \mathbb{I} . Substituting this into Equation 4.2.3, we get

$$\mathbb{E}[Q] = (\mathbb{I} - A)^{-1}[h].$$

□

Note that this expression is equivalent to solving the system $(\mathbb{I} - A)x = h$. We can solve this in Matlab using

$$Q = (\mathbb{I} - A) \setminus h;$$

Recall from Section 3.2.1, there was also an alternate expression for the expected total cost, Equation 3.2.12,

$$\mathbb{E}[Q(x_0)] = \int_{r_v}^{\lambda - r_v} [(\text{Id} - \mathcal{L})^{-1}\rho_0](x_n)h(x_n)dx_n.$$

We can derive a discrete space equivalent of this expression in a similar way.

Theorem 4.2.4. *In the discrete approximation, the expected total cost of a random walk is given by*

$$\mathbb{E}[Q] = [(\mathbb{I} - A)^{-1}[\rho_0]][h],$$

with $\mathbb{E}[Q]$ having dimensions $(M - 1) \times (M - 1)$ and $[\rho_0]$ and $[h]$ taken to be column and row vectors, respectively.

Proof. Beginning with Equation 4.2.4,

$$\mathbb{E}[Q] = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{i_0=1}^{M-1} [A^n]_{i_0, i_n} [\rho_0]_{i_0} [h]_{i_n} \Delta x,$$

and using the fact that the matrix A is symmetric, we get

$$\mathbb{E}[Q] = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} [A^n [\rho_0]]_{i_n} [h]_{i_n} \Delta x.$$

Using Lemma 4.2.2, we can rewrite the infinite summation $\sum_{n=0}^{\infty} A^n$ to get

$$\mathbb{E}[Q] = \sum_{i_n=1}^{M-1} [(\mathbb{I} - A)^{-1}[\rho_0]]_{i_n} [h]_{i_n} \Delta x.$$

Finally, writing this as a matrix multiplication, we get

$$\mathbb{E}[Q] = [(\mathbb{I} - A)^{-1}[\rho_0]] [h].$$

□

As before, we can solve this in Matlab using

$$Q = ((I-A) \setminus rho0) * h;$$

where Q is a $(M - 1) \times (M - 1)$ matrix since ρ_0 is a column vector and h is a row vector.

■ 4.3 Discretised expressions for a Markov-modulated random walk strategy

The discretisation of the Markov-modulated case follows similarly. Recalling Equation 3.3.5,

$$\rho_{n,j}(x_n) = \sum_{i=1}^J \int_{r_i}^{\lambda-r_i} p_i(x_n - x_{n-1}) \rho_{n-1,i}(x_{n-1}) P_{i,j} dx_{n-1},$$

and applying recursively, we get

$$\rho_{n,j_n}(x_n) = \sum_{j_{n-1}=1}^J \cdots \sum_{j_0=1}^J \int_{r_{j_{n-1}}}^{\lambda-r_{j_{n-1}}} \cdots \int_{r_{j_0}}^{\lambda-r_{j_0}} \rho_{0,j_0}(x_0) \left[\prod_{k=0}^{n-1} p_{j_k}(x_{k+1} - x_k) P_{j_k,j_{k+1}} dx_k \right]. \quad (4.3.1)$$

As discussed in the unmodulated case, integrals become summations in the discrete approximation. Therefore, Equation 4.3.1 becomes

$$[\rho_{n,j_n}]_{i_n} = \sum_{j_{n-1}=1}^J \cdots \sum_{j_0=1}^J \sum_{i_{n-1}=m_{j_{n-1}}+1}^{M-m_{j_{n-1}}-1} \cdots \sum_{i_0=m_{j_0}+1}^{M-m_{j_0}-1} [\rho_{0,j_0}]_{i_0} \left[\prod_{k=0}^{n-1} [A_{j_k}]_{i_k,i_{k+1}} [P]_{j_k,j_{k+1}} dx_k \right].$$

We can rewrite this as

$$[\rho_{n,j_n}]_{i_n} = \sum_{j_{n-1}=1}^J \cdots \sum_{j_0=1}^J \sum_{i_{n-1}=1}^{M-1} \cdots \sum_{i_0=1}^{M-1} [\rho_{0,j_0}]_{i_0} \left[\prod_{k=0}^{n-1} [A_{j_k}]_{i_k,i_{k+1}} [P]_{j_k,j_{k+1}} dx_k \right] \quad (4.3.2)$$

since the additional summation terms now included are zero anyway due to the structure of the A_j matrices. We can rewrite the $\rho_{0,j}$ vectors as a single vector by concatenating them to form the vector $[\rho_0]$. That is,

$$[\rho_0] = ([\rho_{0,1}], \dots, [\rho_{0,J}])^\top,$$

Then, the element $[\rho_{0,j}]_i$ is equivalent to $[\rho_0]_{(j-1)(M-1)+i}$, and similarly for ρ_1, ρ_2 , etc. Just as we did with the unmodulated case, we can express Equation 4.3.2 as a matrix product:

$$[\rho_{n,j_n}]_{i_n} = \sum_{j_0=1}^J \sum_{i_0=1}^{M-1} [\rho_{0,j_0}]_{i_0} [(\mathbf{A}^n)_{j_0,j_n}]_{i_0,i_n}, \quad (4.3.3)$$

where the matrix \mathbf{A} is given by

$$\mathbf{A} = \begin{pmatrix} [A_1]P_{1,1} & [A_1]P_{1,2} & \cdots & [A_1]P_{1,J} \\ [A_2]P_{2,1} & [A_2]P_{2,2} & \cdots & [A_2]P_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ [A_J]P_{J,1} & [A_J]P_{J,2} & \cdots & [A_J]P_{J,J} \end{pmatrix},$$

where each A_j matrix has size $(M-1) \times (M-1)$, so the matrix \mathbf{A} must have size $J(M-1) \times J(M-1)$. The expression $[(\mathbf{A}^n)_{j_0,j_n}]_{i_0,i_n}$ represents the (i_0, i_n) th element of the (j_0, j_n) th submatrix of \mathbf{A} .

Recall from Lemma 3.3.1, the cost incurred in a given state j is

$$\mathbb{E}[Q_j(a)] = \sum_{n=0}^{\infty} \int_{r_v}^{\lambda-r_v} \rho_{n,j}(x_n) h_j(x_n) dx_n,$$

which in the discretised search space becomes

$$\mathbb{E}[Q_j]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} [\rho_{n,j}]_{i_n} [h_j]_{i_n} \Delta x.$$

Substituting the expression for $[\rho_{0,j}]_{i_n}$ from Equation 4.3.3, we get

$$\mathbb{E}[Q_j]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J \sum_{i_0=1}^{M-1} [\rho_{0,j_0}]_{i_0} [(\mathbf{A}^n)_{j_0,j}]_{i_0,i_n} [h_j]_{i_n} \Delta x. \quad (4.3.4)$$

We combine the the initial distribution, $\boldsymbol{\rho}_0(x) = \delta(x-a)\mathbf{z}_0$ with the Δx term to obtain a Kronecker delta,

$$\mathbb{E}[Q_j]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J \sum_{i_0=1}^{M-1} \delta_{j_0,i_0,i_a} z_{0,j_0} [(\mathbf{A}^n)_{j_0,j}]_{i_0,i_n} [h_j]_{i_n} \Delta x,$$

where δ_{j_0,i_0,i_a} is Kronecker delta for (i_0, i_a) in the j_0 th subvector. Now, summing over i_0 and using the property of the Kronecker delta,

$$\mathbb{E}[Q_j]_{i_a} = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J \sum_{i_0=1}^{M-1} [(\mathbf{A}^n)_{j_0,j}]_{i_a,i_n} [h_j]_{i_n} \Delta x, \quad (4.3.5)$$

The \mathbf{A} term and the $[h]$ term do not constitute a full matrix multiplication, since we are only considering a single value of j . This relates back to the issue around Lemma 3.3.11 and our inability to find a simple expression for the expected cost over a single state. The probability of a forager's location on step n in some state j depends on the location on step $n - 1$ in *any* state.

To demonstrate this issue further, we split the \mathbf{A}^n term into submatrices, allowing us to write Equation 4.3.5 as a matrix multiplication. The matrix is split, by

$$(\mathbf{A}^n) = \begin{pmatrix} [(\mathbf{A}^n)_{1,1}] & [(\mathbf{A}^n)_{1,2}] & \cdots & [(\mathbf{A}^n)_{1,J}] \\ [(\mathbf{A}^n)_{2,1}] & [(\mathbf{A}^n)_{2,2}] & \cdots & [(\mathbf{A}^n)_{2,J}] \\ \vdots & \vdots & \ddots & \vdots \\ [(\mathbf{A}^n)_{J,1}] & [(\mathbf{A}^n)_{J,2}] & \cdots & [(\mathbf{A}^n)_{J,J}] \end{pmatrix},$$

where each $[(\mathbf{A}^n)_{i,j}]$ is an $(M - 1) \times (M - 1)$ submatrix, corresponding to the elements $((i - 1)(M - 1) + 1, (j - 1)(M - 1) + 1)$ up to $(i(M - 1), j(M - 1))$ of \mathbf{A}^n . That is, $[(\mathbf{A}^n)_{i,j}]$ is not the same as $[(\mathbf{A})_{i,j}]^n$, since for some $i, j \in \mathcal{S}$, $[(\mathbf{A}^2)_{i,j}]$ will actually depend on $[(\mathbf{A})_{i,j}]$ for every $i, j \in \mathcal{S}$. Thus, Equation 4.3.5 becomes

$$\mathbb{E}[Q_j]_{i_a} = \sum_{n=0}^{\infty} \sum_{j_0=1}^J z_{0,j_0} [(\mathbf{A}^n)_{j_0,j}] [h_j]. \quad (4.3.6)$$

This expression, although looking simpler than Equation 4.3.5, highlights an important problem. To actually determine the values of $[(\mathbf{A}^n)_{i,j}]$, for some n , we require knowing the values of the entire (\mathbf{A}^{n-1}) matrix. This also means we must know the values of $[(\mathbf{A}^{n-1})_{i,j}]$ for every i and j . Thus, we still have to evaluate (\mathbf{A}^n) in its entirety for every single term in the summation over n . This makes sense, since a walk that finishes in state j could have previously been in any other state and so all possible states must be taken into account. A consequence of this is that we cannot consider the infinite sum over the submatrix $(\mathbf{A})_{j_0,j}^n$, but rather have to consider the full matrix \mathbf{A}^n , extracting the corresponding submatrix at the end.

We can still, however, remove the infinite summation using a Neumann series but we must keep the entire matrix \mathbf{A} in our expression.

Lemma 4.3.1. *The matrix \mathbf{A} has an operator norm less than unity.*

Proof. We first show that \mathbf{A}^\top has an operator norm less than unity. Any matrix v , with dimensions $J(M - 1) \times p$, where $p \in \mathbb{N}$, has a norm given by

$$\|v\| = \max_{j=1, \dots, p} \sum_{i=1}^{J(M-1)} |v_{i,j}|.$$

If we consider the matrix v as J individual matrices concatenated vertically, each with dimensions $(M - 1) \times p$, denoted by

$$v = ([v_1], [v_2], \dots, [v_J])^\top,$$

we can write this norm as

$$\|v\| = \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{i=1}^{M-1} |[v_m]_{i,j}| =: C.$$

Element (i, j) of the matrix $\mathbf{A}v$ is given by

$$[\mathbf{A}^\top v]_{i,j} = \sum_{k=1}^{J(M-1)} \mathbf{A}_{i,k}^\top v_{k,j},$$

or alternatively

$$[(\mathbf{A}^\top v)_m]_{i,j} = \sum_{n=1}^J \sum_{k=1}^{M-1} [(\mathbf{A}_{m,n}^\top)]_{i,k} [v_n]_{k,j},$$

where $(\mathbf{A}^\top v)_m$ is the m th concatenated matrix making up $\mathbf{A}^\top v$. Then, $\mathbf{A}^\top v$ has norm

$$\begin{aligned} \|\mathbf{A}^\top v\| &= \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{i=1}^{M-1} \left| \sum_{n=1}^J \sum_{k=1}^{M-1} [(\mathbf{A}_{m,n}^\top)]_{i,k} [v_n]_{k,j} \right| \\ &\leq \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{i=1}^{M-1} \sum_{n=1}^J \sum_{k=1}^{M-1} [(\mathbf{A}_{m,n}^\top)]_{i,k} |[v_n]_{k,j}| \\ &= \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{i=1}^{M-1} \sum_{n=1}^J \sum_{k=1}^{M-1} P_{n,m} [A_n]_{i,k} |[v_n]_{k,j}| \\ &= \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{n=1}^J P_{n,m} \sum_{i=1}^{M-1} \sum_{k=1}^{M-1} [A_n]_{i,k} |[v_n]_{k,j}| =: C^*. \end{aligned}$$

Now, using the results from the unmodulated case, we know that for any n ,

$$\sum_{i=1}^{M-1} \sum_{k=1}^{M-1} [A_n]_{i,k} [v_n]_{k,j} < \sum_{k=1}^J |[v_n]_{k,j}|.$$

Substituting this in,

$$\begin{aligned} C^* &< \max_{j=1, \dots, p} \sum_{m=1}^J \sum_{n=1}^J P_{n,m} \sum_{k=1}^{M-1} |[v_n]_{k,j}| \\ &= \max_{j=1, \dots, p} \sum_{n=1}^J \sum_{k=1}^{M-1} |[v_n]_{k,j}| = \|v\| = C, \end{aligned}$$

since the row sum of the transition matrix P must be 1 for every n . Therefore,

$$\frac{\|\mathbf{A}^\top v\|}{\|v\|} \leq \frac{C^*}{C} < 1 \implies \|\mathbf{A}^\top\|_{op} < 1.$$

Since the adjoint of a matrix is the transpose, and using Theorem 2.1.43, we get $\|\mathbf{A}\|_{op} = \|\mathbf{A}^\top\|_{op} < 1$. \square

Theorem 4.3.2. *In the discrete approximation, the expected total cost across state j of a Markov-modulated random walk is given by*

$$\mathbb{E}[Q_j]_{i_a} = \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J [(\mathbb{I} - \mathbf{A})_{j_0,j}^{-1}]_{i_a,i_n} z_{0,j_0} [h_j]_{i_n}, \quad (4.3.7)$$

where the i_a th element of $\mathbb{E}[Q_j]$ represents the cost incurred in state j of a search that begins at discrete point i_a .

Proof. Considering $\sum_{n=0}^{\infty} \mathbf{A}^n$ as a Neumann series and using Lemma 4.3.1, we can write $\sum_{n=0}^{\infty} \mathbf{A}^n$ as $(\text{Id} - \mathbf{A})^{-1}$, where Id is the identity operator, which in this case is the $J(M-1) \times J(M-1)$ identity matrix, \mathbb{I} . Substituting this into Equation 4.3.5, we get

$$\mathbb{E}[Q_j]_{i_a} = \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J [(\mathbb{I} - \mathbf{A})_{j_0,j}^{-1}]_{i_a,i_n} z_{0,j_0} [h_j]_{i_n}.$$

□

Although we are able to express the expected total cost incurred in a single state without an infinite summation, this expression isn't as neat as our matrix multiplication in Equation 4.3.6.

Recall we were able to find a simple expression for the expectation of the total cost for *all* states of a Markov-modulated random walk. This same reasoning is valid in the discrete approximation.

Theorem 4.3.3. *In the discrete approximation, the expected total cost of a Markov-modulated random walk is given by*

$$\mathbb{E}[Q]_{i_a} = \sum_{j_0=1}^J [(\mathbb{I} - \mathbf{A})^{-1} z_{0,j_0} h]_{(j_0-1)(M-1)+i_a},$$

where the i th element of $\mathbb{E}[Q]$ represents the expected total cost of a walk that starts at point i , and the vector $[(\mathbb{I} - \mathbf{A})^{-1} h]$ has length $J(M-1)$.

Proof. Summing over Equation 4.3.7, we get

$$\begin{aligned} \mathbb{E}[Q]_{i_a} &= \sum_{j=1}^J \mathbb{E}[Q_j]_{i_a} \\ &= \sum_{j=1}^J \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J [(\mathbb{I} - \mathbf{A})_{j_0,j}^{-1}]_{i_a,i_n} z_{0,j_0} [h_j]_{i_n}, \end{aligned}$$

which is equivalent to

$$\mathbb{E}[Q]_{i_a} = \sum_{j_0=1}^J [(\mathbb{I} - \mathbf{A})^{-1} z_{0,j_0} h]_{(j_0-1)(M-1)+i_a}.$$

□

Recall, we had another analytic expression for the expected cost incurred in a single state,

$$\mathbb{E}[Q_j(x_0)] = \int_{x_n=r_v}^{\lambda-r_v} [(\text{Id} - \mathcal{L})^{-1} \boldsymbol{\rho}_0]_j(x_n) h_j(x_n) dx_n,$$

which doesn't require the starting location of the search to be known exactly. Following a similar procedure to that of above, we can find a discretised version of this expression.

Theorem 4.3.4. *In the discrete approximation, the expected total cost incurred in state j of a Markov-modulated random walk can also be expressed as*

$$\mathbb{E}[Q_j] = \sum_{i_n=1}^{M-1} \left[(\boldsymbol{\rho}_0 (\mathbb{I} - \mathbf{A})^{-1})_j \right]_{i_n} [h_j]_{i_n} \Delta x. \quad (4.3.8)$$

where the i_a th element of $\mathbb{E}[Q_j]$ represents the cost incurred in state j of a search that begins at discrete point i_a .

Proof. Beginning with Equation 4.3.4,

$$\mathbb{E}[Q_j] = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} \sum_{j_0=1}^J \sum_{i_0=1}^{M-1} [\boldsymbol{\rho}_0]_{j_0} [(\mathbf{A}^n)_{j_0,j}]_{i_0,i_n} [h_j]_{i_n} \Delta x,$$

we can rewrite this as a matrix multiplication, to get

$$\mathbb{E}[Q_j] = \sum_{n=0}^{\infty} \sum_{i_n=1}^{M-1} [\boldsymbol{\rho}_0 (\mathbf{A}^n)_j]_{i_n} [h_j]_{i_n} \Delta x,$$

where $[\boldsymbol{\rho}_0 (\mathbf{A}^n)_j]$ is a vector of length $J(M-1)$. The order of the two summations can be changed, and we can once again use the fact that A has an operator norm to rearrange this. We rewrite $\sum_{n=0}^{\infty} (\mathbf{A}^n)$ as $((\text{Id} - \mathbf{A})^{-1})$, and our expression becomes

$$\mathbb{E}[Q_j] = \sum_{i_n=1}^{M-1} \left[(\boldsymbol{\rho}_0 (\mathbb{I} - \mathbf{A})^{-1})_j \right]_{i_n} [h_j]_{i_n} \Delta x.$$

□

Theorem 4.3.5. *In the discrete approximation, the expected total cost of a Markov-modulated random walk is also given by*

$$\mathbb{E}[Q] = [\rho_0(\mathbb{I} - \mathbf{A})^{-1}h]\Delta x.$$

Proof. Beginning with Equation 4.3.8, from Theorem 4.3.4, we sum over all possible states

$$\mathbb{E}[Q] = \sum_{j=1}^J \sum_{i_n=1}^{M-1} \left[\rho_0 \left((\mathbb{I} - \mathbf{A})^{-1} \right)_{j,i_n} \right] [h_j]_{i_n} \Delta x.$$

We can further simplify this expression by writing it as a matrix product, resulting in

$$\mathbb{E}[Q] = [\rho_0(\mathbb{I} - \mathbf{A})^{-1}h]\Delta x.$$

□

Results for the one-dimensional model

In this chapter, we implement the discrete expressions for the expected total length of our one-dimensional model, as derived in Chapter 4. First, we discuss how our Markov-modulated model can be used to recover various other models used throughout the literature through careful selection of the model parameters. Where relevant, we also extend these models, or consider strategies that were not originally considered. Finally, we investigate the most efficient search strategy according to our model, considering two-state and three-state Markov chains.

■ 5.1 Special cases of the Markov-modulated random walk strategy

Our Markov-modulated random walk model can be seen as a more general version of many of the specific models that have been discussed in the literature. To demonstrate this, we now demonstrate how careful selection of parameters in our model allows us to recover some of the more important models throughout the literature, or at the very least, high quality approximations to these models.

■ 5.1.1 Unmodulated random walk

The random walk strategy with no Markov-modulation unsurprisingly exists as a special case of the Markov-modulated walk strategy, in which Z has only a single state. Consider the final expression found for the average total cost,

Theorem 3.3.13,

$$\mathbb{E}[Q(a)] = \text{Tr} \left(\left[(\text{Id} - \mathcal{L})^{-1} (\mathbf{h}^\top \mathbf{z}_0) \right] (a) \right).$$

For the unmodulated case, we can consider an underlying Markov chain Z , which only has a single state ($J = 1$), and hence $\mathbf{z}_0 = [1]$, r_v is constant, and the vector $\mathbf{h}(x)$ is the scalar function,

$$h(x) = \mathbb{E}_{X_1|X_0} [q(X_0, X_1) \mid X_0 = x].$$

The operator \mathcal{L} now only consists of the element $\mathcal{L}_{1,1}$, which is

$$[\mathcal{L}_{1,1}f](x_n) = \int_{r_v}^{\lambda - r_v} p_1(x_n - x_{n-1}) f(x_{n-1}) P_{1,1}(x_n) dx_{n-1},$$

which, after substituting in $P_{1,1} = 1$, is equivalent to our operator \mathcal{L} from the unmodulated section. Substituting everything into Theorem 3.3.13, and leaving off the trace since this is a scalar, we arrive at

$$\mathbb{E}[Q(a)] = \left[(\text{Id} - \mathcal{L})^{-1} h \right] (a),$$

which matches with the expression found by Bartumeus *et al.* [6], and discussed in Section 3.2.

Using our Markov-modulated model, we set $P = [1]$, and $\mathbf{z}_0 = [1]$ and investigate the efficiency over four different distributions: power-law, bounded power-law, exponential, and bounded exponential. We choose model parameters that match Bartumeus *et al.* [6] to make comparisons easier. These are $\lambda = 1000$, $r_v = 1$, and $\ell_{\min} = 1$. We also consider the effect of discretisation size on these results, plotting the efficiency for decreasing values of Δx , beginning with $\Delta x = 1$ and ending with $\Delta x = 0.2$, which was used by Bartumeus *et al.* [6].

Figure 5.1.1a, Figure 5.1.1b both match the results of Bartumeus *et al.* [6] exactly. The optimal parameter, μ , for destructive foraging with a power-law distribution is $\mu \rightarrow 1$, and for non-destructive foraging is $\mu \approx 2$, which matches the conclusions of the literature (e.g. [6, 86]).

With the bounded power-law distribution, Figure 5.1.2a shows that the efficiency appears to be very similar to the unbounded power-law, with the biggest difference coming as $\mu \rightarrow 1$. An explanation for this is that the closer μ to 1, the more likely that there are steps that are large enough to be truncated. When μ is larger, especially $\mu \geq 3$, the distribution has a finite variance and the chances of taking a step larger than $\ell_{\max} = 100$ is very low. The non-destructive case is

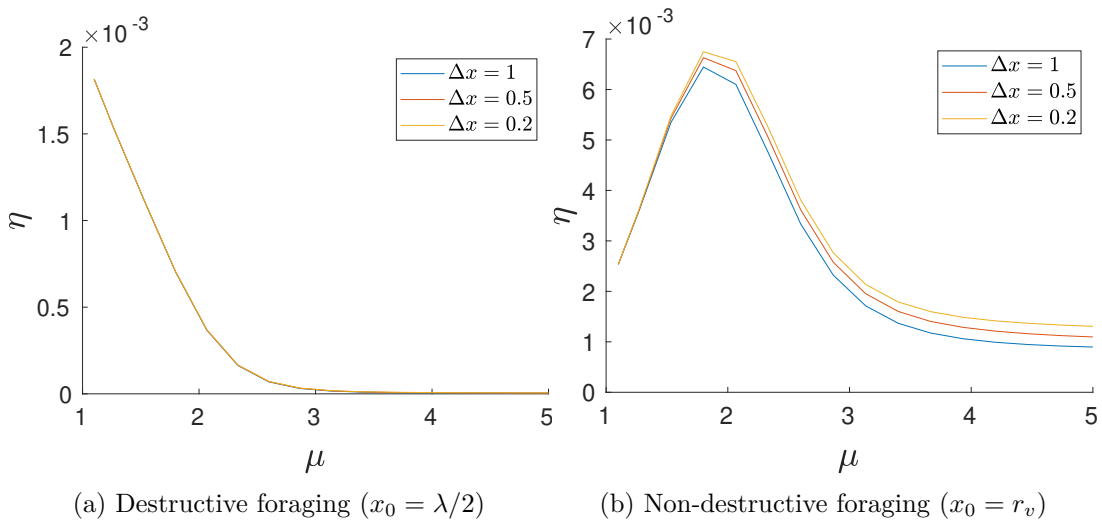


Figure 5.1.1: Search efficiency, η , versus distribution parameter μ , of a power-law step-length distribution strategy with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, and simulations run with three different discretisation sizes, $\Delta x = \{1, 0.5, 0.2\}$.

shown in Figure 5.1.2b, although the efficiency curve is approximately the same shape, there the peak efficiency is shifted slightly towards a larger μ . The peak efficiency ($\eta \approx 3 \times 10^{-3}$) is also much lower than that of the unbounded power-law ($\eta \approx 7 \times 10^{-3}$). The most efficient choice of parameter is now slightly larger than $\mu = 2$, rather than slightly below $\mu = 2$ as with the unbounded distribution. The bounded power-law distribution for the non-destructive case also seems to be relatively sensitive to the discretisation size.

To understand why the upper bound on the step-length distribution results in a worse efficiency, recall that for the unbounded power-law strategy, the peak efficiency was around $\mu = 2$, which is a Lévy walk. This strategy would have mostly small steps, but every so often would take very large steps, allowing the forager to travel a large distance without any backtracking. However, with a small ℓ_{\max} , the forager is unable to make these large steps,

The efficiency of the unbounded exponential strategy in Figure 5.1.3a is optimal with $\mu \rightarrow 0$, and has a huge drop off immediately as μ increases. For the non-destructive search, the unbounded exponential has an optimal efficiency when $\mu \rightarrow 0$, as seen in Figure 5.1.3b. There is a large drop off in efficiency as μ increases, though not as extreme as in the destructive case. The efficiency in the non-destructive case seems to be relatively sensitive to discretisation size.

Finally, we consider the bounded exponential in Figure 5.1.4a, Figure 5.1.4b for destructive and non-destructive foraging, respectively. The destructive foraging

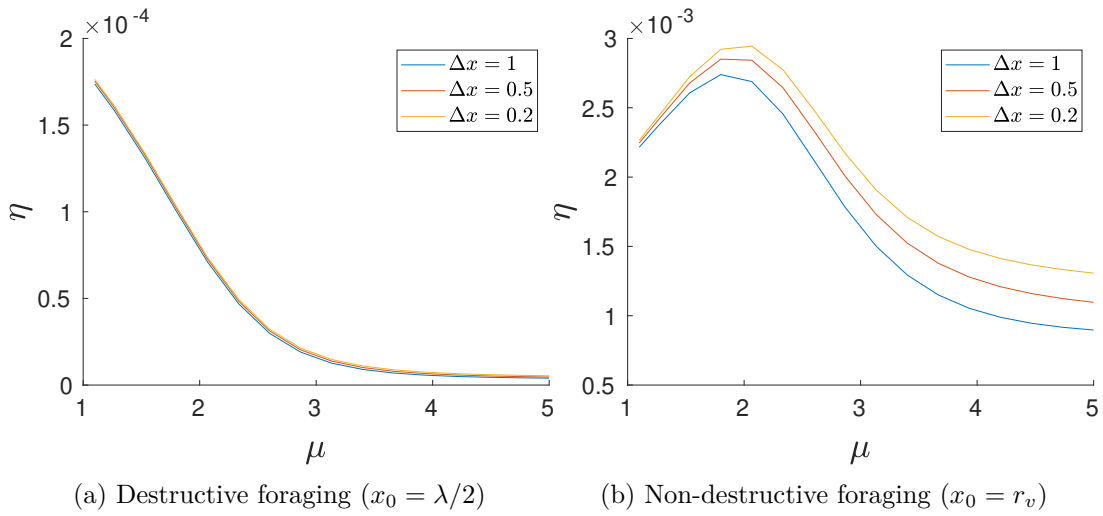


Figure 5.1.2: Search efficiency, η , versus distribution parameter μ , of a bounded power-law step-length distribution strategy with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, $\ell_{\max} = 100$, and simulations run with three different discretisation sizes, $\Delta x = \{1, 0.5, 0.2\}$

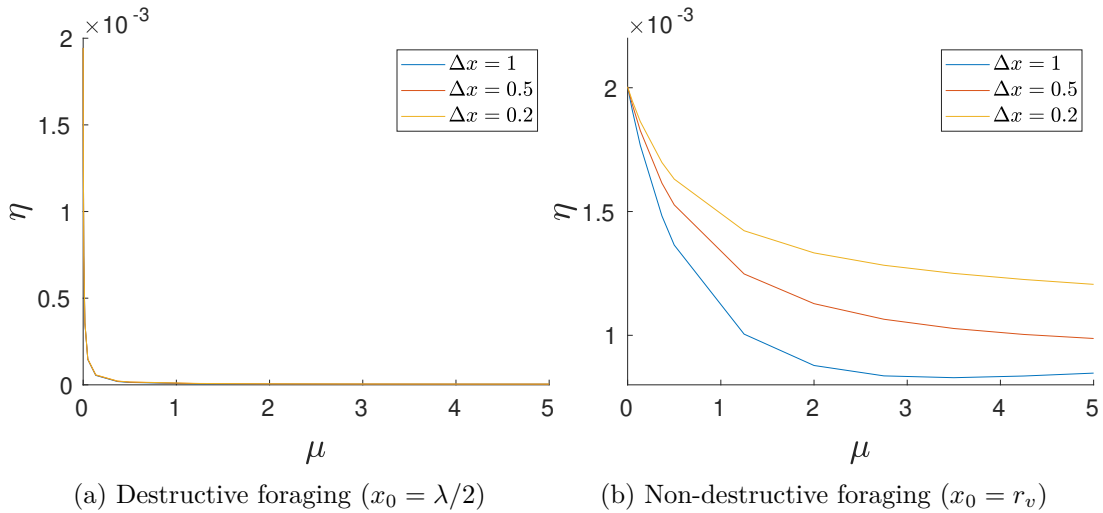


Figure 5.1.3: Search efficiency, η , versus distribution parameter μ , of an exponential step-length distribution strategy with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, and simulations run with three different discretisation sizes, $\Delta x = \{1, 0.5, 0.2\}$.

efficiency looks very similar to that of the unbounded exponential. The non-destructive foraging follows a similar pattern as the unbounded exponential distribution, although the drop off in efficiency as μ increases is not as pronounced. The optimal efficiency for both of these cases is found as $\mu \rightarrow 0$, which corresponds to the steps getting larger.

It is worth noting that a strategy that involved simply choosing a direction at

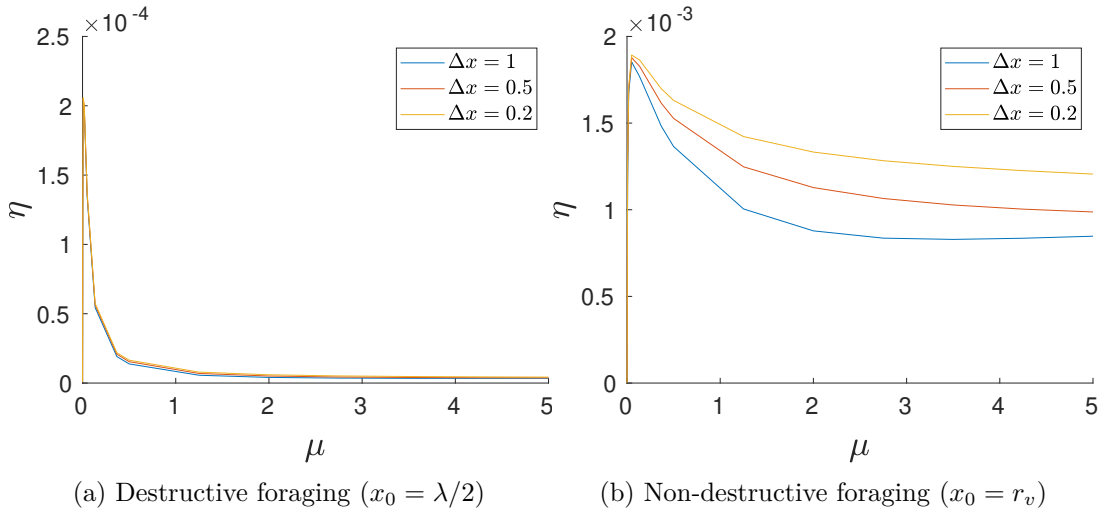


Figure 5.1.4: Search efficiency, η , versus distribution parameter μ , of a bounded exponential step-length distribution strategy with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, $\ell_{\max} = 100$, and simulations run with three different discretisation sizes, $\Delta x = \{1, 0.5, 0.2\}$

random and walking in a straight line (ballistic motion) will have an efficiency of $\eta = 2 \times 10^{-3}$ for $\lambda = 10^{-3}$, since the expected distance travelled will be $\lambda/2$, regardless of starting position. Of the strategies investigated above, the only time an efficiency greater than 2×10^{-3} was achieved was for both the bounded and unbounded power-law distributions, and only for non-destructive foraging. For the unbounded power-law distribution, if the parameter is in the approximate range $1.1 \leq \mu \leq 2.8$, then an efficiency greater than that of ballistic motion will be achieved. Similarly, for the bounded power-law distribution, although the range is slightly smaller, requiring approximately $1.1 \leq \mu \leq 2.6$.

To help summarise the results of comparing these four step-length distributions, we plot the efficiency of all four on a single plot, in Figure 5.1.5a for destructive foraging and Figure 5.1.5b for non-destructive foraging.

We also investigate the effect of the size of the upper bound, by plotting the efficiency of the bounded power-law and bounded exponential distributions for a range of ℓ_{\max} , for both destructive and non-destructive foraging. Based on Figure 5.1.6a, Figure 5.1.6b, Figure 5.1.7a, Figure 5.1.7b, the larger ℓ_{\max} is, the higher the efficiency of a foraging strategy, although for the exponential distribution an effect is only properly noticed for $\mu \rightarrow 0$. It is also worth noting that the graphs of the bounded distributions look the same as the corresponding unbounded distributions as $\ell_{\max} \rightarrow \infty$, which is to be expected since we recover the unbounded distributions as $\ell_{\max} \rightarrow \infty$. When $\ell_{\max} = 1000$, the bounded

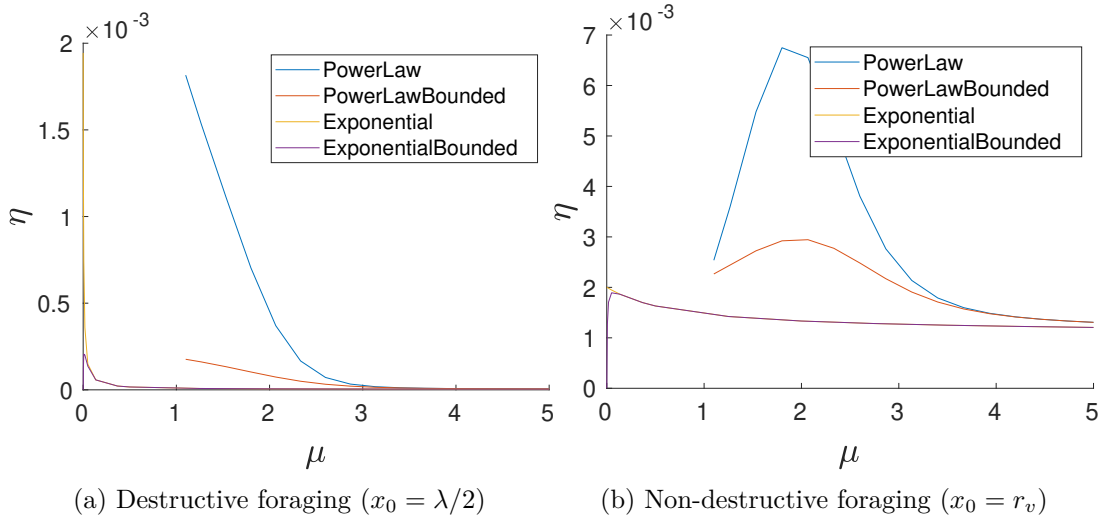


Figure 5.1.5: Search efficiency, η , versus distribution parameter μ , for all four step-length distributions with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, $\ell_{\max} = 100$, and simulations run using discretisation size $\Delta x = 0.2$.

distributions have a different distribution to the unbounded, even though a step can reach the boundary from any point, which seems counter-intuitive. This occurs because the bounded distribution is being normalised over ℓ_{\min} to ℓ_{\max} , where ℓ_{\max} is finite, meaning larger steps are still less likely than they are for the unbounded distributions.

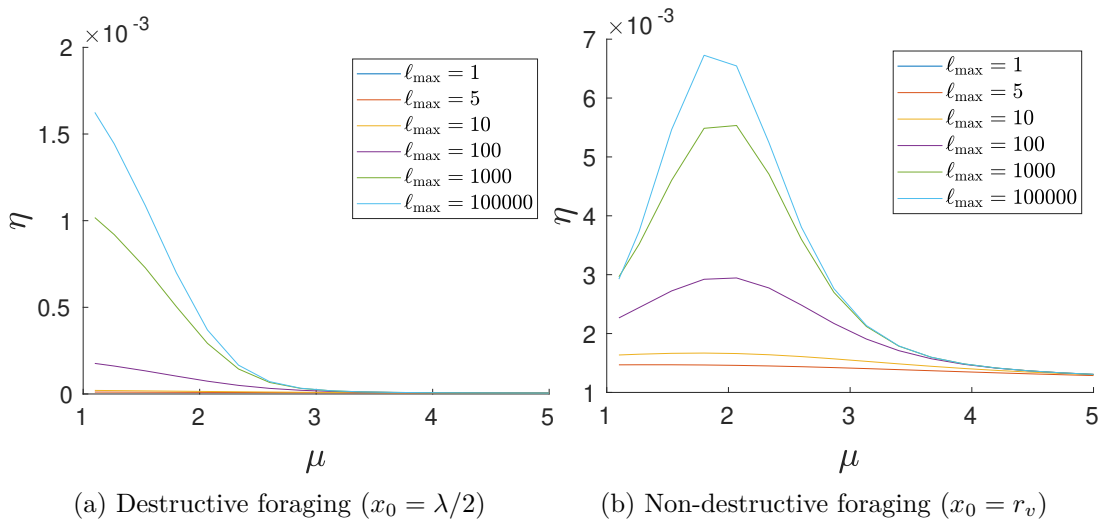


Figure 5.1.6: Search efficiency, η , versus distribution parameter μ , for a bounded power-law distribution with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, $\Delta x = 0.2$, and simulations run across multiple different upper bounds, $\ell_{\max} = \{1, 5, 10, 100, 1000\}$.

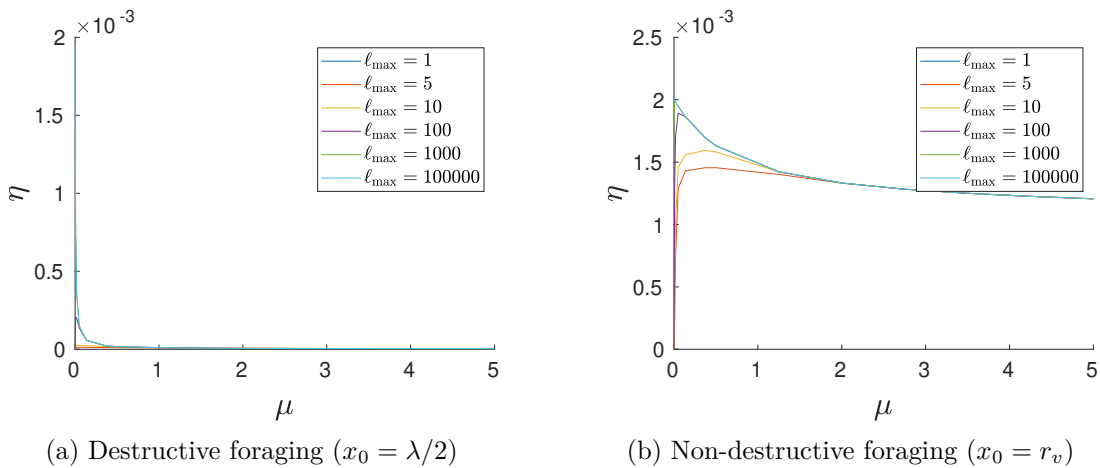


Figure 5.1.7: Search efficiency, η , versus distribution parameter μ , for a bounded exponential distribution with model parameters: $\lambda = 1000$, $r_v = 1$, $\ell_{\min} = 1$, $\Delta x = 0.2$, and simulations run across multiple different upper bounds, $\ell_{\max} = \{1, 5, 10, 100, 1000\}$.

■ 5.1.2 Giving-up-time forager

Another special case of the Markov-modulated random walk is the random walk with “giving-up time”, as discussed in papers by Benhamou [8], Plank and James, [61], and Reynolds [66, 67].

These papers described a giving-up-time forager that undergoes an intensive search until some time, τ , has elapsed, and then gives up its search and switches into an extensive search. The type of motion used in the intensive and extensive searches varies between papers, though most commonly it was Brownian motion and ballistic motion, respectively. Of the giving-up time strategies that have been investigated, the adaptive Lévy walk is the most general, which has Brownian motion for its intensive search, and the extensive search is drawn from a power-law distribution, where μ is allowed to take any value. Reynolds [66] showed that his adaptive Lévy walk model could also model the previously investigated giving-up time models (e.g [61]) by setting $\mu \rightarrow 1$.

We now show that the adaptive Lévy walk can itself be thought of as a special case of a two-state Markov-modulated random walk. Furthermore, we have derived some analytic expressions for the efficiency in Section 3.3.1 and their discretised counterparts Section 4.3, which should provide better accuracy compared to the results found by Reynolds [67] via numerical simulation. We let state 1 represent the intensive search, and so

$$p_1(x) = \frac{\mu_1 - 1}{\ell_{\min}} \left(\frac{\ell}{\ell_{\min}} \right)^{-\mu_1},$$

and state 2 represent the extensive search,

$$p_2(x) = \frac{\mu_2 - 1}{\ell_{\min}} \left(\frac{\ell}{\ell_{\min}} \right)^{-\mu_2}.$$

For a giving-up time model, the forager begins with an intensive search, and so $\mathbf{z}_0 = (1, 0)$. When $\mu_1 \geq 3$, the intensive phase is a Brownian motion and hence we have the adaptive Lévy walk model of Reynolds [66]. Further, when $\mu_1 \geq 3$ and $\mu_2 \rightarrow 1$, we recover the giving-up time strategy of Plank and James [61]. One of the earliest giving-up time strategies was the composite Brownian walk, which was introduced by Benhamou [8]. This involved exponentially distributed step-sizes for both the intensive and extensive search, though with different parameters for each. We may also model this using a two-state Markov-modulated random walk, by choosing exponential distributions for both states, although we do not do this since adaptive Lévy walks were found to have a better performance.

How we construct the transition matrix, P , will depend on how the giving-up-time, τ , is defined. For example, a geometrically distributed giving-up-time with expected value $1/p$ can be modelled with the transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 0 & 1 \end{bmatrix}. \quad (5.1.1)$$

This geometric giving-up time will correspond to an exponential giving-up time in the continuous limit. If instead we have a deterministic giving-up-time, say N steps before giving up, we can model this with the transition matrix

$$P = \begin{bmatrix} \mathbf{0}^T & \mathbb{I}_N \\ 0 & \mathbf{e}_N \end{bmatrix}, \quad (5.1.2)$$

where $\mathbf{0} = (0, \dots, 0)$ and \mathbf{e}_N is a vector with the N th element a 1 and all other elements are 0. In this case, the first N states of the Markov chain correspond to the same step-length distribution, $p_1(x)$, and the final state, state $N + 1$, corresponds to the step-length distribution $p_2(x)$.

By reordering the states of our Markov chain, we can write Equation 5.1.1, Equation 5.1.2 in the form

$$P = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{t} & T \end{bmatrix},$$

matching Definition 2.1.46. That is, the giving-up time for both of these cases can be thought of as a discrete phase-type distribution. In fact, we can construct our states and transition matrix in such a way that we can consider any phase-type

distribution for the giving-up time. Not only does this allow us to consider distributions such as geometric and negative-binomial, but using Theorem 2.1.47 we can approximate any possible distribution for the giving-up time using our Markov-modulated random walk.

Optimal giving-up time for Brownian motion giving up into ballistic motion

Plank and James [61] described a forager that used a Brownian motion for its intensive search before giving-up into a ballistic motion for the extensive search. They were able to determine an approximate value for the optimal choice of giving-up time:

$$\tau^* = \frac{d}{4v_I} \left(1 - \frac{4}{3}\varepsilon - \frac{2}{5}\varepsilon^2 \right),$$

where d is the distance between targets, v_I is the velocity of the forager during the intensive search, and $\varepsilon = \frac{2x_0^2}{\pi v_I d} < 0.19$. If $\varepsilon > 0.19$ then the efficiency is optimised at $\tau = 0$, meaning the search is comprised entirely of ballistic motion. The average distance between patches, d , is equivalent to λ in our model, and we are considering non-destructive foraging, which implies $x_0 = r_v$. Since we are considering the distance as opposed to the time required to find food, we have not yet discussed the velocity of a searcher. In the model of Plank and James [61], a Brownian motion with variance σ^2 has velocity $v_I = \sqrt{2/\pi}\sigma$. Recalling Example 2.1.1, the variance of a Brownian motion which arises as the limit of a random walk with unbounded power-law distributed steps is $\sigma^2 = (\mu - 1)/(\mu - 3)\ell_{\min}^2$, and we are choosing $\ell_{\min} = 1$. Taking these differences into account, under our model we would expect the optimal giving-up time to be

$$\tau^* = \frac{\lambda\sqrt{\pi}}{4\sqrt{2}\sigma} \left(1 - \frac{4}{3} \frac{2r_v^2}{\pi\lambda} - \frac{2}{5} \frac{4r_v^4}{\pi^2\lambda^2} \right),$$

with $\lambda \leq \frac{2r_v^2}{0.19\pi}$ implying that $\tau^* = 0$.

In our case, $\lambda = 1000$, $r_v = 1$, and we choose $\mu = 5$, so $\sigma = \sqrt{2}$ and $\lambda > \frac{2r_v^2}{0.19\pi}$. Thus, the optimal giving-up time is

$$\tau^* = \frac{1000\sqrt{\pi}}{8} \left(1 - \frac{4}{3} \frac{2}{\pi 1000} - \frac{2}{5} \frac{4}{\pi^2 1000^2} \right) \approx 221.3686,$$

and so the optimal parameter $p = 1/\tau^* = 0.0045$.

We consider a Markov-modulated random walk with geometric giving-up time defined as above, with state 1 being a power-law with $\mu = 5$, and state 2 being a power-law with $\mu \rightarrow 1$. We use the transition matrix in Equation 5.1.1 and plot

the efficiency of a non-destructive search for a range of different values of p in Figure 5.1.8. We also use Matlab's *fmincon* solver to find the optimal choice of p , and mark it on the plot.

The optimal efficiency is $\eta = 0.0152$ and occurs at $p = 0.0038$, which corresponds to a mean giving-up time of 263.4465. This is not exactly the same as the value predicted by Plank and James [61], but is a reasonable approximation, given the differences between their model and ours. Firstly, they considered a deterministic giving-up time, whereas we are considering a geometric giving-up time. We could consider a deterministic giving-up time, although to consider the values around this size, the matrix \mathbf{A} would be far too large. Another difference between our model and theirs is that they considered continuous-time processes, whereas we are using discrete-time processes.

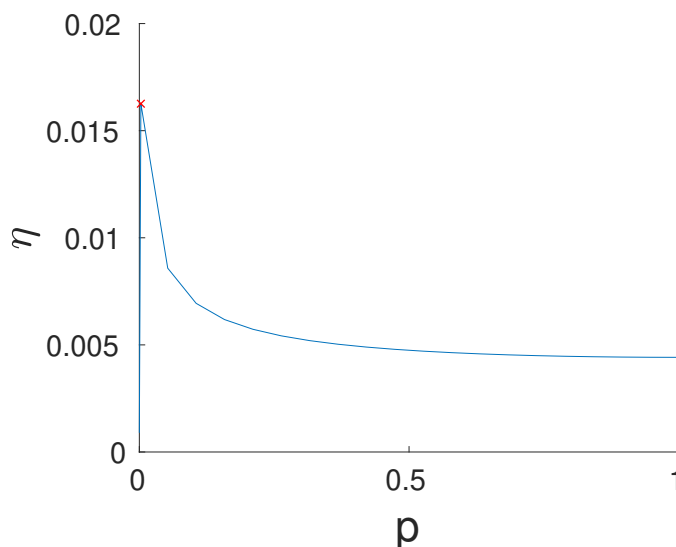


Figure 5.1.8: The efficiency of a non-destructive search ($x_0 = r_v$) against the parameter p for the geometric giving-up time. The forager uses a power-law distribution search with parameter $\mu_1 = 5$ before “giving up” and using a power-law search with $\mu_2 \rightarrow 1$. The peak efficiency (the red x) is $\eta = 0.0152$ which occurs at $p = 0.0038$, corresponding to a mean giving-up time of 263.4465.

Optimal adaptive Lévy walk parameters for fixed giving-up times

As discussed in Section 2.2, Reynolds [66] investigated an adaptive Lévy walk, which involved switching between a power-law distribution with $\mu = 3$, and a power-law distribution with $\mu \rightarrow 1$, called the intensive and extensive phases, respectively. Reynolds also considered changing the values of μ during the extensive phase, and determined the optimal extensive search parameter. However, as correctly pointed out by Plank and James [61], Reynolds only optimised over

fixed giving-up times.

We now use our Markov-modulated random walk to investigate the adaptive Lévy walk model of Reynolds [66], although not only do we consider a range of different μ values for the extensive phase, but we also allow μ to vary in the intensive phase. We plot the efficiency at 40 equispaced points in $1 < \mu_1 \leq 5$ and $1 < \mu_2 \leq 5$, for 5 different choices of p for a geometric giving-up time distribution.

The geometric giving-up time with the highest mean, $p = 0.01$, had an optimal efficiency at $\mu_1 = 5$, and $\mu_2 = 1.6154$, as can be seen in Figure 5.1.9. As the mean giving-up time was reduced, the optimal parameter for the intensive search does not change, but the extensive search parameter does change. For $p = 0.1$ in Figure 5.1.10, the optimal choice for the extensive parameter was $\mu_2 = 1.7179$. For $p = 0.5$, $p = 0.9$, and $p = 0.99$, in Figure 5.1.11, Figure 5.1.12, Figure 5.1.13 the optimal choice for the extensive parameter was $\mu_2 = 1.8205$.

In all of these cases, the optimal extensive strategy is not a ballistic motion, but rather a Lévy flight, and the optimal intensive search is Brownian motion, matching the results of Reynolds [66]. As the giving-up time increases, the extensive search gets closer to a ballistic motion, with a higher probability of very large steps, whereas for short giving-up times, the extensive parameter is closer to $\mu \approx 2$, which is the optimal for unmodulated search strategies.

In Figure 5.1.14, we also plot the efficiency of a giving-up time strategy for destructive foraging. Unsurprisingly, the optimal efficiency occurs at $\mu_1 \rightarrow 1$ and $\mu_2 \rightarrow 1$, which is just a ballistic motion and not a true giving-up time strategy.

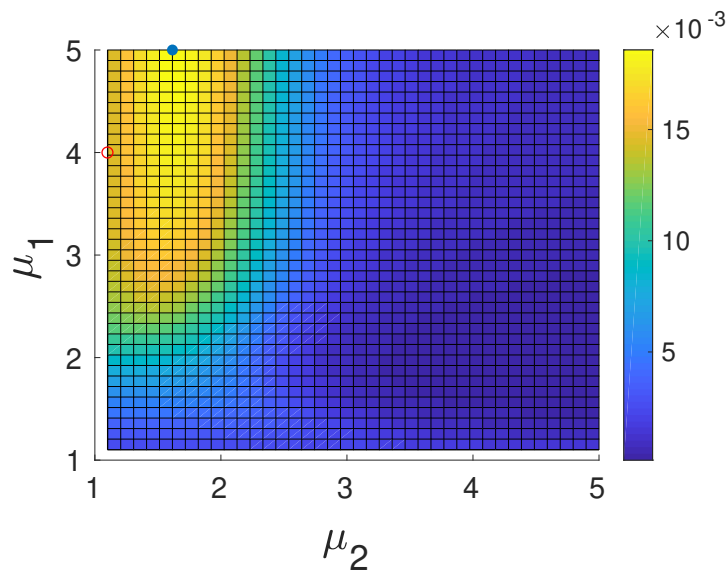


Figure 5.1.9: The efficiency of a non-destructive search ($x_0 = r_v$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.01$) amount of time has elapsed, which has a mean of 100. The optimal efficiency is marked with the blue circle, and the red circle represents a Brownian motion giving-up into ballistic motion, although it is actually at any $\mu \geq 3$ rather than specifically at $\mu = 4$.

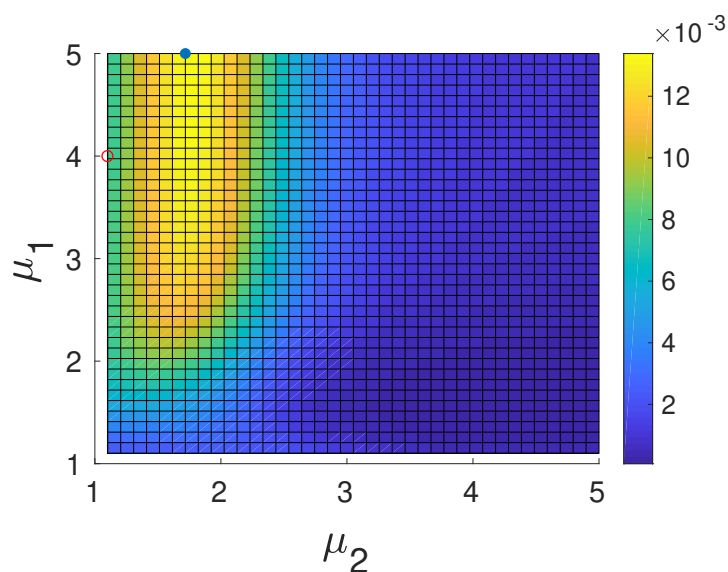


Figure 5.1.10: The efficiency of a non-destructive search ($x_0 = r_v$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.1$) amount of time has elapsed, which has a mean of 10. The optimal efficiency is marked with the blue circle, and the red circle represents a Brownian motion giving-up into ballistic motion, although it is actually at any $\mu \geq 3$ rather than specifically at $\mu = 4$.

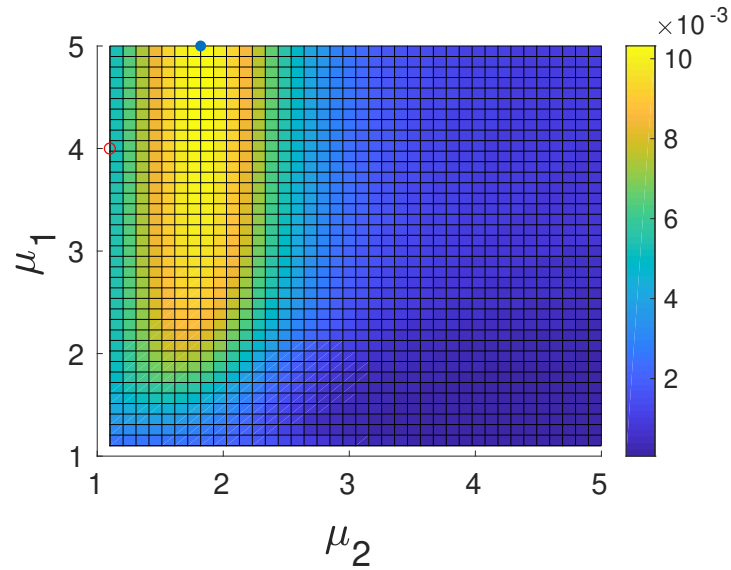


Figure 5.1.11: The efficiency of a non-destructive search ($x_0 = r_v$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.5$) amount of time has elapsed, which has a mean of 2. The optimal efficiency is marked with the blue circle, and the red circle represents a Brownian motion giving-up into ballistic motion, although it is actually at any $\mu \geq 3$ rather than specifically at $\mu = 4$.

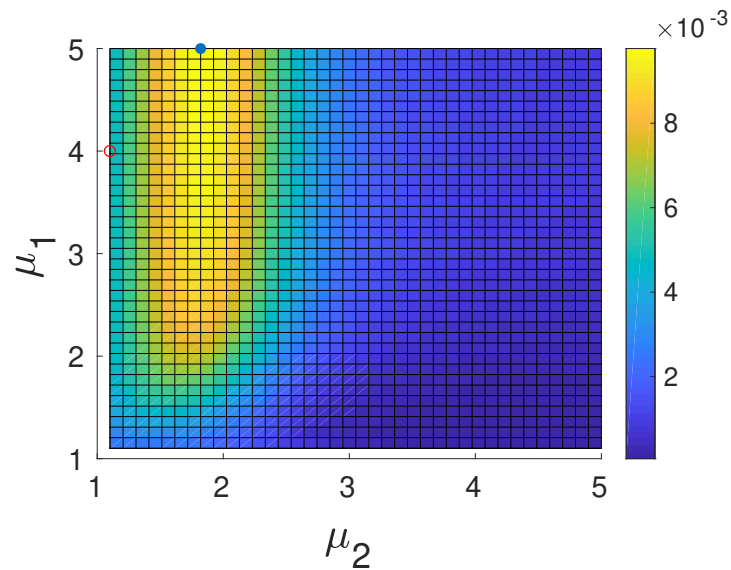


Figure 5.1.12: The efficiency of a non-destructive search ($x_0 = r_v$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.9$) amount of time has elapsed, which has a mean of approximately 1.11. The optimal efficiency is marked with the blue circle, and the red circle represents a Brownian motion giving-up into ballistic motion, although it is actually at any $\mu \geq 3$ rather than specifically at $\mu = 4$.

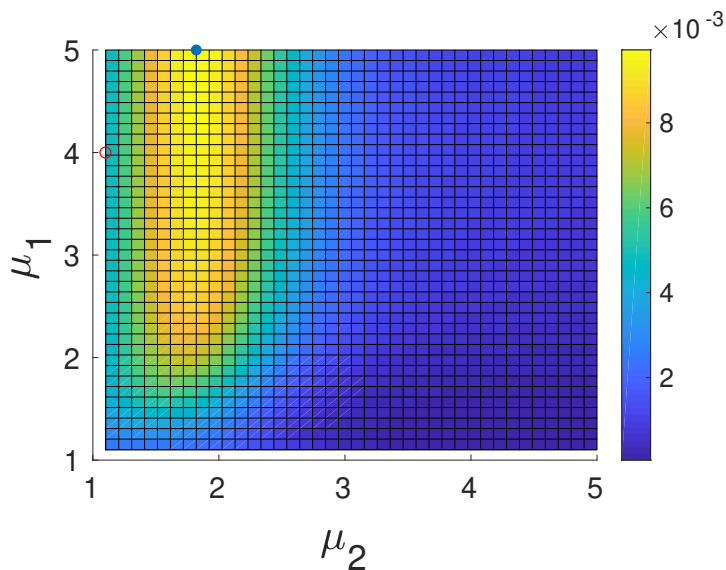


Figure 5.1.13: The efficiency of a non-destructive search ($x_0 = r_v$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.99$) amount of time has elapsed, which has a mean of 1.010101. The optimal efficiency is marked with the blue circle, and the red circle represents a Brownian motion giving-up into ballistic motion, although it is actually at any $\mu \geq 3$ rather than specifically at $\mu = 4$.

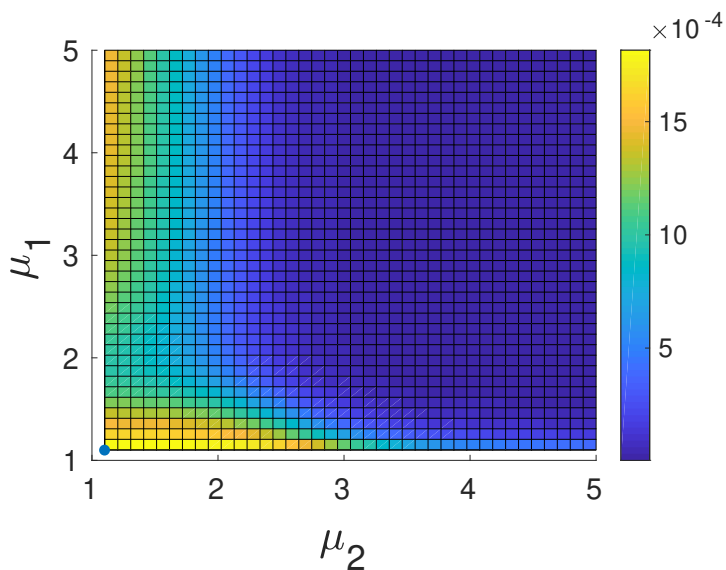


Figure 5.1.14: The efficiency of a destructive search ($x_0 = \lambda/2$) where the forager uses a power-law distribution search with parameter μ_1 before “giving up” and using a power-law search with μ_2 , after a geometrically distributed ($p = 0.01$) amount of time has elapsed, which has a mean of 100. The optimal efficiency is marked with the blue circle.

Optimal geometric giving-up time strategy

In the previous two parts, we considered the giving-up time models of Plank and James [61] and Reynolds [66]. However, both of these have advantages and disadvantages. Plank and James [61] were able to find the optimal giving-up time, but considered only Brownian motion giving-up into ballistic motion. On the other hand, Reynolds [66] found the optimal extensive parameters, but considered only fixed giving-up times. As shown above, we were able to use our Markov-modulated random walk strategy to investigate both of these models. Now, we can combine these ideas and use our model to optimise over both the giving-up time and the parameters μ_1 and μ_2 .

We make use of Matlab's *fmincon* function to minimise the expected total travel length, and hence maximise the efficiency over the three different parameters. We use the constraints that $0 \leq p \leq 1$, $1 < \mu_1 \leq 10$ and $1 < \mu_2 \leq 10$.

For a non-destructive search, *fmincon* finds the optimal efficiency at $p = 0.0060$, $\mu_1 = 10.0000$, and $\mu_2 = 1.5868$. This value of p corresponds to a mean giving-up time of approximately 167, and the optimal values of μ correspond to a Brownian motion for the intensive search that gives up into a Lévy walk for the extensive search. The optimal value of $\mu_1 = 10$ was the maximum allowed value of μ_1 according to our constraints. Increasing this upper bound will change the optimal value of μ_1 accordingly, although it only changes the efficiency very slightly.

■ 5.1.3 Vision switching forager

Benichou *et al.* [14] defined a model where the forager switches between two states an unlimited number of times. The time spent in each of the states is exponential. However, while in one of the two states, the forager cannot locate any targets.

Although this scenario cannot be modelled directly with our model, we can model a scenario which should produce very similar results. In the model outlined by Benichou *et al.* [14], when the forager is unable to locate targets, reaching a food target is impossible. However, in our model, even when the radius of vision is 0, it is still possible for a forager to locate a target by running into it, or jumping beyond it and being truncated. To alleviate this issue, we could define the distance between food patches to be very large, and then have the radius of vision switch between a very large value and zero. However, in practice this is not feasible since a Lévy walk has heavy-tailed steps and so there is no choice of distance between targets that is sufficiently large.

For example, say we wanted to investigate food patches that are a distance λ apart, and have the animal unable to locate a food patch during one of the two search modes, and have a radius of vision, r_v , during the other mode. Then, we can model this approximately using an interval size of λ^* , with $\lambda^* \gg \lambda$, and choosing $r_1 = r_v + (\lambda^* - \lambda)$ and $r_2 = 0$. Then, as long as λ^* is sufficiently large the animal will never locate a patch while in state 2, and while in state 1 the scenario is equivalent to a search with food at the endpoints of $[r_v, \lambda - r_v]$. The downside to this approximation is that the size of the matrix that must be solved will be very large when λ^* is very large, although the increase with λ^* is only linear. Also, we could only realistically investigate search strategies that are not heavy-tailed.

Although we cannot easily investigate hidden targets with our Markov-modulated random walk model, we can consider some difficult-to-detect targets and models where adjustments to the forager's radius of vision occur. A difficult-to-detect target may be harder to find in a certain mode, but it does not necessarily have to be impossible to find targets in this mode. Thus, we can define λ^* , λ , r_1 and r_2 as above, although we no longer required λ^* to be large enough to prevent a certain step from reaching this boundary. Other models that involve a change in the forager's radius of vision are easily modelled using our Markov-modulated random walk. A model like this may occur in nature, for example, when the weather changes from sunny to foggy, resulting in a degradation of a forager's perceptive ability.

One possible way to extend our Markov-modulated random walk to allow us to investigate hidden and hard-to-detect targets is by defining some kind of periodic boundary conditions on the interval $[0, \lambda]$. Making this extension will be fairly involved, and is not something we do in this thesis, though we discuss some further details in Chapter 7.

We now consider two different strategies for non-destructive searching, on a search space of length $\lambda = 1000$. We plot the efficiency of a two-state power-law strategy against the parameters of μ in Figure 5.1.15, where the Markov chain is given by

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$

and $\mathbf{z}_0 = (0.5, 0.5)$. In Figure 5.1.15a, we plot the efficiency of a strategy with $r_v = 50$ in both states, and in Figure 5.1.15b, we plot the efficiency of a strategy that switches between $r_v = 50$ and $r_v = 0$. One important thing to note about Figure 5.1.15b is that the efficiency is not longer symmetric about the $\mu_1 = \mu_2$ line, as it is for Figure 5.1.15a. For both scenarios, large efficiencies occur for

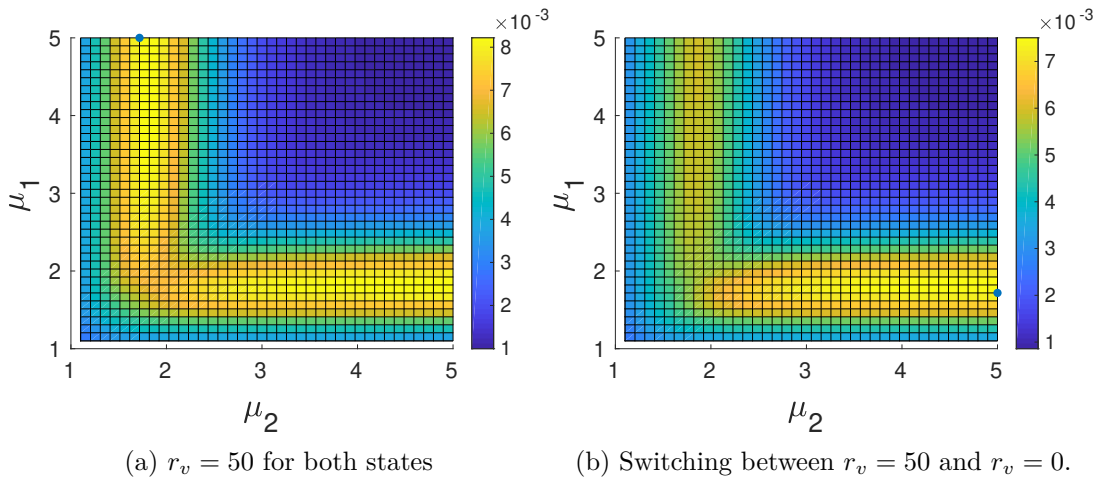


Figure 5.1.15: The efficiency of a non-destructive search ($x_0 = 50$) where the forager switches between two power-law distributions with parameters μ_1 and μ_2 , with equal probability of switching between states. The radius of vision is different for each subplot. The optimal efficiency is marked with a blue circle.

strategies that switch between a Brownian motion and a Lévy walk, though for the vision switching strategy the efficiency is much worse, as expected, if the Brownian motion occurs in state 1 while the vision is low.

■ 5.2 Optimal Markov-modulated search strategies for two-states, three states, and higher

In Section 5.1.2, we used Markov-modulated random walk strategies to consider a giving-up time forager. These giving-up time strategies are a special case of Markov-modulated random walks in which there is an absorbing state. In this section, we consider more general Markov-modulated random walk strategies that do not necessarily have an absorbing state. For a strategy with two states, the Markov chain will have a transition matrix

$$P = \begin{bmatrix} 1 - a & a \\ b & 1 - b \end{bmatrix}, \quad (5.2.1)$$

where $a, b \in [0, 1]$. For three states, the transition matrix is

$$P = \begin{bmatrix} 1 - a - b & a & b \\ c & 1 - c - d & d \\ e & f & 1 - e - f \end{bmatrix},$$

where $a, b, c, d, e, f \in [0, 1]$. We can continue in this manner, defining the transition matrix for higher-state Markov-modulated random walk strategies. A J -state

Markov-modulated random walk will have $J^2 - J$ free variables in its transition matrix. For a J -state Markov-modulated random walk, we also have J different step-length distributions. We choose these step-length distributions to all be unbounded power-law distributions, since they offered the highest efficiency in previous sections, although there is nothing preventing us from considering any other combination of distributions. Each step-length distribution may have a different parameter μ , and we choose $\ell_{\min} = 1$ and $r_v = 1$ for all distributions, for simplicity. Finally, the initial distribution of the Markov chain has $J - 1$ parameters, since it must sum to one. Thus, we have J different parameters for the distributions, $J^2 - J$ for the transition matrix, and $J - 1$ for the initial distribution, totalling $J^2 + J - 1$ parameters. The evaluation of the efficiency for any set of parameters also becomes slower for larger J , since the matrix \mathbf{A} has dimensions $J(M - 1) \times J(M - 1)$. Because of this, we only consider search strategies with up to $J = 3$ states.

We once again use Matlab's *fmincon* to minimise the expected total travel length and hence maximise the efficiency. We begin with a strategy where $J = 1$ with an initial guess of $\mu = 2$. As we increase the number of states to J , we use the optimal solution for $J - 1$ states to determine an initial guess for *fmincon*. The transition matrix gets an extra column, with each element being $1/J$. To compensate and ensure the row sums equal 1, the transition matrix elements from the previous solution are all scaled by $(J - 1)/J$ so they sum to $1 - (1/J)$ rather than 1. We also add an extra row to the bottom of the previous transition matrix, with each element being $1/J$. For example, if the optimal solution for the $J = 2$ strategy was given by Equation 5.2.1, then the initial guess for the $J = 3$ strategy is

$$P = \begin{bmatrix} 2(1 - a)/3 & 2a/3 & 1/3 \\ 2b/3 & 2(1 - b)/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

Our initial guess for the new parameter μ is the mean of the other μ values. For the J strategy, our initial guess for the initial distribution, we take the optimal choice for the $J - 1$ strategy and add an element $1/J$ for the final state, and renormalise the other states by multiplying them by $(J - 1)/J$. For example, if the $J = 2$ had an optimal initial distribution of

$$\mathbf{z}_0 = (a, 1 - a),$$

with $a \in [0, 1]$, then our initial guess for the initial distribution is

$$\mathbf{z}_0 = (2a/3, 2(1 - a)/3, 1/3).$$

Results of *fmincon*

For the 1-state forager, the optimal strategy is $P = [1]$, $\mathbf{z}_0 = [1]$, and $\mu = 1.8699$. This is a Lévy walk, matching the results we found in Section 5.1.1,

For the 2-state forager, the optimal strategy is a transition matrix given by

$$P = \begin{bmatrix} 0.9940 & 0.0060 \\ 0 & 1 \end{bmatrix},$$

an initial probability vector $\mathbf{z}_0 = (1, 0)$ and distribution parameters $\boldsymbol{\mu} = (10, 1.5868)$. This result is interesting, since the optimal general 2-state strategy is, in fact, a giving-up time strategy, with the same parameters as found in Section 5.1.2. This corresponds to a forager that begins using Brownian motion, before giving up after a geometrically distributed amount of time with an approximate mean of 166.66 steps, before performing a Lévy flight with $\mu = 1.5868$.

For the 3-state forager, the optimal strategy is a transition matrix given by

$$P = \begin{bmatrix} 0.9849 & 0 & 0.0151 \\ 0 & 1 & 0 \\ 0 & 0.0152 & 0.9848 \end{bmatrix},$$

an initial probability vector $\mathbf{z}_0 = (1, 0, 0)$, and distribution parameters $\boldsymbol{\mu} = (10, 1.3908, 2.4368)$. This strategy once again corresponds to a giving-up time strategy that begins following a Brownian motion, before giving-up into a Lévy flight, with parameter $\mu = 2.4368$. Then, after some further time, the forager gives up again, and switches to a Lévy flight with $\mu = 1.3908$, which will somewhat resemble a ballistic path, due to the small μ parameter. The mean of both geometrically distributed giving-up time are approximately 66 steps.

Two-dimensional Foraging Models

■ 6.1 General assumptions

We consider a forager in two dimensions, searching for targets that are scattered throughout a boundaryless search space \mathbb{R}^2 . Targets are represented by points that are randomly distributed in the search space according to a homogeneous spatial Poisson point process.

We only consider destructive foraging in this chapter, since non-destructive foraging is more complicated. We discuss how non-destructive foraging may be investigated using this model in Chapter 7.

The forager begins its search at some point which we define to be the origin. Each step in the forager's search strategy involves two parts. Firstly, the forager selects a search direction, θ_i , measured in radians about a unit circle centred at the current location. Secondly, the forager selects a step length, ℓ_i , in which to travel in the chosen direction. We can also define the *turning angle*, $\beta_i := \theta_i - \theta_{i-1}$, which is the change in direction from the previous step's direction, measured in radians. The location of a forager after taking i steps is then denoted as (x_i, y_i) . The forager also has a constant radius of vision, r_v , and whenever a target is within this range the forager will be able to detect it and move directly towards it. As a forager takes a step it can still detect targets while moving, and so each step will have an area of vision associated with it.

The possible search strategies that we consider are random walks with any choice of step-length distribution, and with uniform turning angles. Some extensions to the model are possible, such as allowing for arbitrary distributions for the turning angle, as well as considering Markov-modulated random walks, or even

non-destructive targets, although these may make the model intractable. We discuss some of these further in Chapter 7.

We begin by making a simplifying assumption: upon revisiting a previously explored area, a forager may still find targets within this area. This is an equivalent scenario to the spatial Poisson process being redrawn in an already explored area, after a forager has already found it to be empty. Under the simplest model, which we call the *zeroth-order* approximation, the targets may have respawned in an area immediately following a search through that area. Thus, a search that involved a lot of backtracking over the same area is not penalised for repeated searching of the same area. In the *first-order* approximation, the area covered in the previous step may not have food in it, although all area that was explored two or more steps ago may now have food. This means that the points within an explored area are redrawn, although with one step of lag. This corresponds to some penalisation of backtracking, although not fully.

In general, the *n*th order approximation means that a forager may not find food in any area already covered in the previous *n* steps. In the next section we solve the model analytically for both the zeroth-order and first-order models, and show why second-order and higher models are too complicated to consider analytically. In Section 6.3.2, we use simulations to model the same scenario, and can deduce the difference in results between different order approximations, including the case of no approximation (∞ -order).

As with the one-dimensional model, in two dimensions our primary aim is to determine the most efficient search strategy. We maintain the same notion of efficiency as in the one-dimensional case. That is, the efficiency is defined as

$$\eta = \frac{1}{\mathbb{E}[L]}, \quad (6.1.1)$$

where L is the total distance travelled to find a target. Based on the assumptions of our model we are able to show that this definition of efficiency is equivalent under the zeroth-order model to the efficiency as defined by Viswanathan *et al.* [86], which is

$$\eta = \frac{1}{\mathbb{E}[|\ell|] \mathbb{E}[N]}, \quad (6.1.2)$$

where $\mathbb{E}[|\ell|]$ is the expected distance travelled in a single step, and $\mathbb{E}[N]$ is the expected number of steps to find a target.

Due to the complete spatial randomness of the target distribution, we can make a few simplifications to the model. The probability of a target being within some searched area depends on the size of the area and the density of the targets.

Since the targets have a homogeneous spatial Poisson distribution, the density is constant everywhere. This means that the start and end locations of a step make no difference to the probability of finding food, and what really matters is the search area covered by a step. Thus, we do not actually need to keep track of the forager's location, and can without loss of generality assume that each step begins at the origin. For the first-order approximation, we do need to track the search area of the previous step and determine the overlapping area between that and the current step.

We make one final assumption; that there are no targets within the initial radius of vision of the forager before taking any steps. This is reasonable since if there was, a forager would immediately move to this food target and then begin searching again, repeating this process until there was no target within its initial radius of vision.

■ 6.2 Solving the model analytically

In deriving an analytic expression for the efficiency, much of the process depends on the order of model we are considering. We initially begin deriving expressions for the probability of finding a target on any given step, and for the expected length of a single step, for any order of model. Then, in Section 6.2.1 we use these expressions to find an analytic expression for the efficiency for the zeroth-order model. In Section 6.2.2, we consider the first-order model, and are unable to derive an exact expression for the efficiency, and we discuss why this is the case.

In Section 3.2 we found an expression for the total cost to find a single target, which we simplified in Equation 3.2.3 to

$$Q(x_0) = \sum_{n=0}^{\infty} \mathbb{1}_{(\tau \geq n+1)} q(X_n, X_{n+1}),$$

where $Q(x_0)$ was the total cost for a search beginning at x_0 , and $q(X_n, X_{n+1})$ was the cost attributed to a step that begins at X_n and ends at X_{n+1} , and the random variable τ represented the time until food was found. In the case of our two-dimensional model, we are not interested in finding a general cost function, but rather, the total length travelled to find food. Thus, we get

$$L = \sum_{n=0}^{\infty} \mathbb{1}_{(\tau \geq n+1)} |\ell(X_n, X_{n+1})|,$$

where we can also drop the dependence on the starting location x_0 , since all starting locations are considered equivalent.

$$\mathbb{E}[L] = \sum_{n=0}^{\infty} \mathbb{E} \left[\mathbb{1}_{(\tau \geq n+1)} |\ell_n(X_n, X_{n+1})| \right],$$

which we can rewrite as

$$\mathbb{E}[L] = \sum_{n=0}^{\infty} \mathbb{E} \left[|\ell(X_n, X_{n+1})| \mid \mathbb{1}_{(\tau \geq n+1)} \right] \Pr(\tau \geq n+1).$$

Since our step-length distribution has no dependence on the start or end locations, we may rewrite this without the dependence on X_n and X_{n+1} . However, we include a subscript on the ℓ to denote the step number, since for first-order and higher models we need to keep the step number in mind. Thus,

$$\mathbb{E}[L] = \sum_{n=1}^{\infty} \mathbb{E} [|\ell_n|] \Pr(\tau \geq n).$$

where we have also shifted the summation. Thus, we now have two expressions — the expected length of a single step, and the probability that food is located within n steps — that must be found in order to solve for the efficiency.

Both of these expressions will depend on whether we are considering the zeroth-order or the first-order approximation. Since the targets have a homogeneous spatial Poisson distribution, the starting point of a step will not affect the probability that food is located. The probability of finding exactly k points within productive area a , with food density ρ , is

$$\Pr(k, \rho, a) = \frac{(a\rho)^k e^{-(a\rho)}}{k!}.$$

Let $f_A(a)$ represent the PDF of the area covered in a single step. The probability of not finding a target on a given step is then

$$\begin{aligned} \Pr(\text{not found}) &= \int_{A_{\min}}^{\infty} \Pr(\text{not found} \mid A = a) f_A(a) da \\ &= \int_{A_{\min}}^{\infty} \Pr(0, \rho, a) f_A(a) da \\ &= \int_{A_{\min}}^{\infty} e^{-a\rho} f_A(a) da. \end{aligned} \tag{6.2.1}$$

Therefore, the probability of finding a target on any given step is

$$\Pr(\text{found}) = 1 - \int_{A_{\min}}^{\infty} e^{-a\rho} f_A(a) da. \tag{6.2.2}$$

We use Equation 6.2.2 to solve for $\mathbb{E}[N]$, although how we now proceed depends on which order of the model we are considering. We derive $\mathbb{E}[N]$ for the zeroth-order model in Section 6.2.1.

Next, we derive an expression for the expected length of a single step that is valid for any order of model. This expression is similar to the expectation of the

step-length distribution, except we must take into account the possibility of a step being truncated:

$$\mathbb{E}[|\ell|] = \int_{\ell_{\min}}^{\infty} \mathbb{E}[\ell_T | \ell] f(\ell) d\ell, \quad (6.2.3)$$

where $\mathbb{E}[\ell_T | \ell]$ is the expected truncated length of a step which would have had full length ℓ . Let $A(u, \beta)$ represent the new area covered by a step of length u and with turning-angle β . Since the targets are distributed according to an homogeneous spatial Poisson process, the probability of finding a target after moving certain distance is exponentially distributed with parameter $\rho A(u, \beta)$. Then, the truncated length will be given by

$$\mathbb{E}[\ell_T | \ell] = \int_{\beta=0}^{2\pi} \int_{u=0}^{\ell} \rho A(u, \beta) e^{-\rho A(u, \beta)} \frac{1}{2\pi} du d\beta + \ell \int_{\beta=0}^{2\pi} \left(e^{-\rho A(u, \beta)} \right) \frac{1}{2\pi} d\beta, \quad (6.2.4)$$

where the first term represents when the forager finds a target, and hence truncates the distance to u , whereas for the second integral the target is not found and the forager travels a full distance ℓ . The $1/(2\pi)$ in both terms comes from the distribution of the turning angle, which is uniform between 0 and 2π . Note also that the integral does not begin at ℓ_{\min} but rather at 0 since a step may be truncated after less than ℓ_{\min} has been travelled. To solve this further we need an expression for $A(u, \beta)$, which depends on which order approximation we are considering.

■ 6.2.1 Zeroth-order approximation

To solve for the efficiency of the zeroth-order approximation, we first need an expression for the amount of new area travelled, $A(\ell, \beta)$, for a step of length ℓ , and a turning-angle β . Consider a single step of a forager's search, where the forager begins at a point that is not within vision range of a target. While travelling a distance ℓ , the forager can see r_v distance around itself in all directions, which results in a rectangular shaped area (of dimension $2r_v$ by ℓ) that is searched while travelling. Once landing, the forager can also see r_v in all directions around itself, resulting in a circle shaped area of searching (with radius r_v). We do not need to include the initial circle of radius r_v since this cannot have a target in since the previous step did not find food. As mentioned earlier, we can without loss of generality assume that this circle on the initial step does not contain food, as if it did, the forager immediately finds and destroys it, and the search begins again at this new origin. The overlap between these three regions must also be accounted for, to avoid double counting some of the area. Figure 6.2.1 shows the area in question, which totals

$$A(\ell, \beta) = 2\ell r_v - \frac{1}{2}\pi r_v^2 + \frac{1}{2}\pi r_v^2 = 2\ell r_v,$$

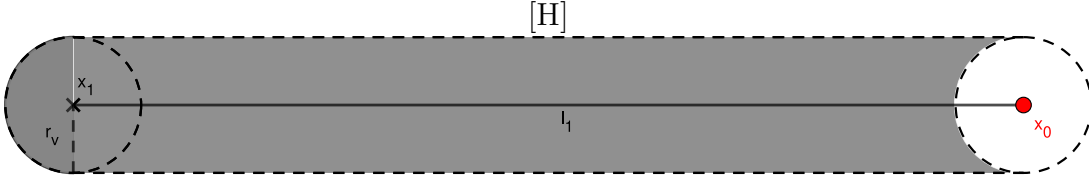


Figure 6.2.1: A step from location x_0 to x_1 . The search area that is explored by the forager is shaded in grey. The initial circle of radius r_v around the forager is known to be empty, otherwise the forager's search would have ended during the previous step.

and is therefore not a function of the turning-angle. Further, we can approximate the area searched in each step by a forager as a rectangle, which for the zeroth-order model is still $A(\ell, \beta) = 2\ell r_v$, as in Figure 6.2.2.



Figure 6.2.2: A step from location x_0 to x_1 . The search area that is explored by the forager is shaded in grey, and has now been simplified into a rectangle by not taking into account the radius of vision around the forager at either end of a step. The total area is still the same.

Since ℓ is a random variable, A is also a random variable, which in our case is

$$f_A(a) = \frac{f\left(\frac{a}{2r_v}\right)}{2r_v}.$$

The probability of not finding food, from Equation 6.2.1, was

$$\begin{aligned} \Pr(\text{not found}) &= \int_{A_{\min}}^{\infty} e^{-a\rho} f_A(a) da \\ &= \int_{A_{\min}}^{\infty} \frac{e^{-2r_v\ell\rho}}{2r_v} f\left(\frac{\ell}{2r_v}\right) d\ell, \end{aligned}$$

for which integration by substitution yields

$$\Pr(\text{not found}) = \int_{\ell_{\min}}^{\infty} e^{-2\ell r_v \rho} f(\ell) d\ell, \quad (6.2.5)$$

which can now be solved exactly, depending on the step-length distribution.

Since the probability of finding food is the same for every step under our zeroth-order model, we can consider τ as a geometric random variable.

The fact that τ is geometrically distributed also tells us that the expectation of the total number of steps is

$$\mathbb{E}[N] = \frac{1}{\Pr(\text{found})} = \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2\ell r_v \rho} f(\ell) d\ell\right)^{-1}, \quad (6.2.6)$$

which is used in evaluating the efficiency.

Recall Equation 6.2.4,

$$\mathbb{E}[\ell_T | \ell] = \int_{\beta=0}^{2\pi} \int_{u=0}^{\ell} \rho A(u, \beta) e^{-\rho A(u, \beta)} \frac{1}{2\pi} du d\beta + \ell \int_{\beta=0}^{2\pi} \left(e^{-\rho A(u, \beta)} \right) \frac{1}{2\pi} d\beta,$$

which gives us an expression for the expected step-length of a single step. For the zeroth-order approximation, this becomes

$$\begin{aligned} \mathbb{E}[\ell_T | \ell] &= 2r_v \rho \int_{u=0}^{\ell} u e^{-2r_v \rho u} du + \ell e^{-2r_v \rho \ell} \\ &= 2r_v \rho \left[\frac{-e^{-2r_v \rho u} (2r_v \rho u + 1)}{(2r_v \rho)^2} \right]_0^{\ell} + \ell e^{-2r_v \rho \ell} \\ &= \frac{1 - e^{-2r_v \rho \ell} (2r_v \rho \ell + 1) + 2r_v \rho \ell e^{-2r_v \rho \ell}}{2r_v \rho} \\ &= \frac{1 - e^{-2r_v \rho \ell}}{2r_v \rho}. \end{aligned}$$

Substituting this back into Equation 6.2.3, we get

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{\ell_{\min}}^{\infty} \left(\frac{1 - e^{-2r_v \rho \ell}}{2r_v \rho} \right) f(\ell) d\ell \\ &= \frac{1}{2r_v \rho} \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2r_v \rho \ell} f(\ell) d\ell \right). \end{aligned}$$

Now, we have expressions for both the expected number of steps, and the length travelled in a single step, so we can find the efficiency. Note that,

$$\begin{aligned} \mathbb{E}[N] \mathbb{E}[|\ell|] &= \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2r_v \rho \ell} f(\ell) d\ell \right)^{-1} \frac{1}{2r_v \rho} \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2r_v \rho \ell} f(\ell) d\ell \right) \\ &= \frac{1}{2r_v \rho}, \end{aligned}$$

and hence

$$\eta = \frac{1}{\mathbb{E}[N] \mathbb{E}[|\ell|]} = 2r_v \rho, \quad (6.2.7)$$

meaning, as expected, that the efficiency of the zeroth-order approximation does not depend on the step-length distribution at all.

With analytic expressions for the expected length of a step, as well as both the expected number of steps and the probability of finding a food within n steps, we can now show that the two definitions of efficiency, Equation 6.1.1, Equation 6.1.2, are equivalent under the zeroth-order model.

Recalling our expression for the expected total travel length,

$$\mathbb{E}[L] = \sum_{n=1}^{\infty} \mathbb{E}[|\ell_n|] (\Pr(\tau \geq n)).$$

Now, since the expected length of a single step, $\mathbb{E}[|\ell|]$ does not depend on n , we may take it outside of the summation.

$$\mathbb{E}[L] = \mathbb{E}[|\ell|] \sum_{n=1}^{\infty} \Pr(\tau \geq n),$$

which is equivalent to

$$\mathbb{E}[L] = \mathbb{E}[|\ell|] \mathbb{E}[N].$$

Thus, our definition of efficiency is

$$\eta = \frac{1}{\mathbb{E}[L]} = \frac{1}{\mathbb{E}[|\ell|] \mathbb{E}[N]},$$

which is equivalent to the definition of Viswanathan *et al.* [86].

Example 6.2.1. We consider a forager with an exponential step-length distribution,

$$p(\ell) = \mu e^{-\mu(\ell - \ell_{\min})}, \quad \ell \in [\ell_{\min}, \infty), \quad \mu > 0.$$

Then, the probability that a step does not find a food target is

$$\begin{aligned} \Pr(\text{not found}) &= \int_{\ell_{\min}}^{\infty} e^{-2r_v \rho} p(\ell) d\ell \\ &= \int_{\ell_{\min}}^{\infty} \mu e^{-(2r_v \rho + \mu)\ell + \mu \ell_{\min}} d\ell \\ &= \mu e^{\mu \ell_{\min}} \left[\frac{e^{-(2r_v \rho + \mu)\ell}}{-(2r_v \rho + \mu)} \right]_{\ell_{\min}}^{\infty} \\ &= \mu e^{\mu \ell_{\min}} \frac{e^{-(2r_v \rho + \mu)\ell_{\min}}}{2r_v \rho + \mu} \\ &= \frac{\mu e^{-2r_v \rho \ell_{\min}}}{2r_v \rho + \mu}, \end{aligned}$$

and hence

$$(\Pr(\text{not found}))^n = \left(\frac{\mu e^{-2r_v \rho \ell_{\min}}}{2r_v \rho + \mu} \right)^n.$$

Then, the probability of a food target being found on any given step is given by

$$\Pr(\text{found}) = 1 - \frac{\mu e^{-2r_v \rho \ell_{\min}}}{2r_v \rho + \mu} = \frac{2r_v \rho + \mu - \mu e^{-2r_v \rho \ell_{\min}}}{2r_v \rho + \mu}.$$

Finally, the expected total number of steps to find a food target is

$$\mathbb{E}[N] = \frac{1}{\Pr(\text{found})} = \frac{2r_v \rho + \mu}{2r_v \rho + \mu - \mu e^{-2r_v \rho \ell_{\min}}}.$$

Consider the case of $\mu \gg r_v \rho$, which corresponds to an average step size that is relatively small compared to the radius of vision and the food target density. For

this case, $\Pr(\text{not found}) \approx e^{-2r_v\rho\ell_{\min}}$, and hence $\mathbb{E}[N] \approx \frac{1}{1-e^{-2r_v\rho\ell_{\min}}}$. Then, the smaller r_v , ρ and ℓ_{\min} are, the greater the total number of expected steps, which makes intuitive sense.

The expected length of a single step is

$$\begin{aligned} \mathbb{E}[|\ell|] &= \frac{1}{2r_v\rho} \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2r_v\rho\ell} p(\ell) d\ell \right) \\ &= \frac{1}{2r_v\rho} \left(1 - \int_{\ell_{\min}}^{\infty} e^{-2r_v\rho\ell} \mu e^{-\mu(\ell-\ell_{\min})} d\ell \right) \\ &= \frac{1}{2r_v\rho} \left(1 - \mu e^{\mu\ell_{\min}} \int_{\ell_{\min}}^{\infty} e^{-(2r_v\rho+\mu)\ell} d\ell \right) \\ &= \frac{1}{2r_v\rho} \left(1 - \mu e^{\mu\ell_{\min}} \left[\frac{e^{-(2r_v\rho+\mu)\ell}}{-(2r_v\rho+\mu)} \right]_{\ell_{\min}}^{\infty} \right) \\ &= \frac{1}{2r_v\rho} \left(1 - \frac{\mu e^{-2r_v\rho\ell_{\min}}}{2r_v\rho+\mu} \right). \end{aligned}$$

Then, unsurprisingly we get

$$\eta = \frac{1}{\mathbb{E}[N] \mathbb{E}[|\ell|]} = 2r_v\rho.$$

■ 6.2.2 First-order approximation

We now consider the first-order approximation, which is considerably more difficult to solve. Since the previous area must be taken into account overlap will occur, and thus the area will now also be a function of the turning angle. As can be seen in Figure 6.2.3, there are four separate cases to consider when determining the area of overlap, for a turning-angle of $0 \leq \beta \leq \pi$. There are a further four cases for when the turning-angle is $\pi \leq \beta \leq 2\pi$, although these are mirrors of the first four cases, and only need a minor adjustment to ensure the sign of the trigonometric identities in the expressions are correct.

Throughout the derivation of the area, we are forced to make some further simplifications in order to be able to realistically solve the problems. For example, there is actually more than 4 cases to consider, since there is a separate case in between case 1 and case 2, although this only occurs at very specific angles and is similar in size to both of these two, so we exclude this. There are also times where we assume that very small triangles of area are right-angled, or else the expressions for the area would be far more complicated for little difference in accuracy. We use Monte Carlo simulations in Section 6.3.1 to ensure our expressions for the area are suitably accurate.

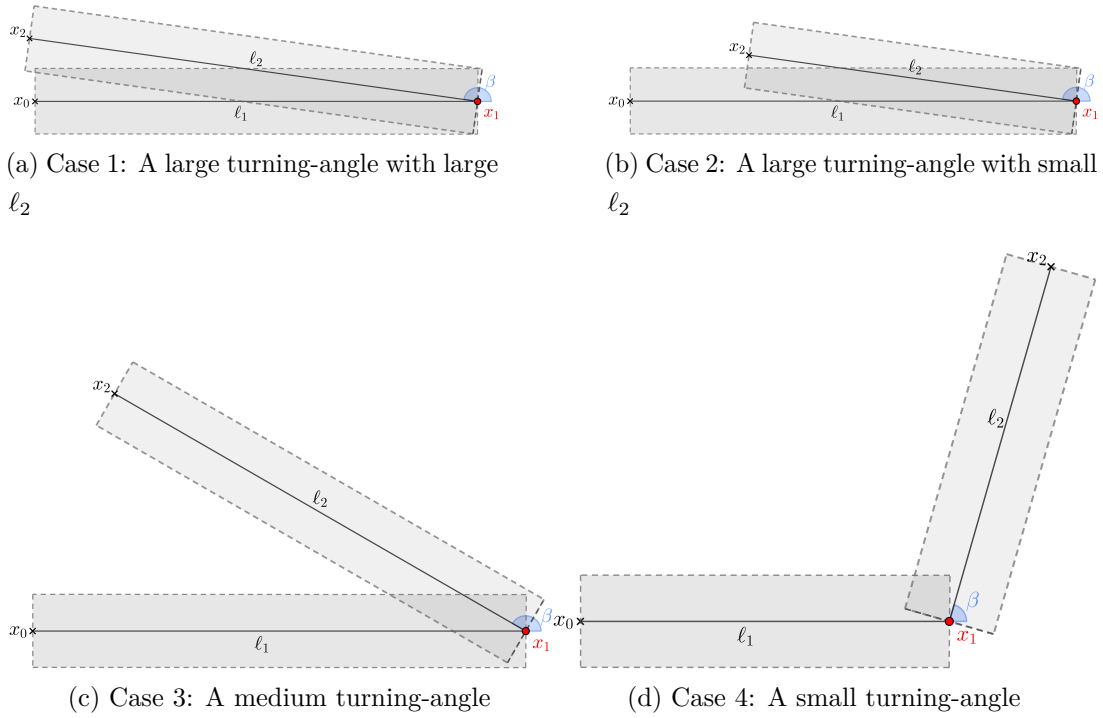


Figure 6.2.3: The four cases that need to be considered to determine the overlapping area between two steps, with different turning-angles giving rise to different cases. When the turning-angle is large, which of case 1 or case 2 arise depends on the relative sizes of ℓ_1 and ℓ_2 . When ℓ_2 is sufficiently small, case 3 becomes case 2. Four more cases exist for turning-angles greater than π , although these are mirrors of the four cases above.

Even with these simplifications, deriving expressions for the area covered by a step is quite tedious and results in complicated expressions. For this reason, we relegate the derivation of the first-order area to Chapter B. The resulting expressions for the area of overlap between consecutive steps are

$$A_{\text{overlap}} = \begin{cases} r_v^2 |\tan(\beta/2)| & \text{if } \beta \leq \pi/2 \text{ or } \beta \geq 3\pi/2, \\ \frac{r_v^2 |\sin(\beta)|}{1 - |\cos(\beta)|} & \text{if } \pi/2 \leq \beta \leq \pi - \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right), \\ & \text{or } \pi + \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right) \leq \beta \leq 3\pi/2 \\ \frac{1}{2} \left[-\min(\ell_1, \ell_2)^2 |\tan(\beta)| \right. & \text{if } \pi - \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right) \leq \beta \\ \left. + 2r_v \min(\ell_1, \ell_2)(2 + \tan^2(\beta)) \right] & \text{and } \beta \leq \pi + \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right). \\ -r_v^2 |\tan^3(\beta)| & \end{cases}$$

Then, the total area covered by a step of length ℓ_2 with turning-angle β and

previous step-length ℓ_1 is

$$A = 2\ell_2 r_v - A_{\text{overlap}}. \quad (6.2.8)$$

In Section 6.3.1, we use Monte Carlo simulations to show Equation 6.2.8 is close to the true value of A .

Probability density function of the new area covered

Let $f_A(x)$ be the PDF representing a forager taking a step that covers x amount of new area. We can express this as

$$f_A(x) = \int_{\ell_1=\ell_{\min}}^{\infty} \int_{\beta=0}^{2\pi} f_A(x, \beta, \ell_1) d\beta d\ell_1,$$

where ℓ_1 is the length of the forager's previous step, β is the forager's turning angle, and x is a function of ℓ_1 , ℓ_2 and β . Since the turning angle and the length of a step are drawn independently, we can further rearrange this to get

$$f_A(x) = \int_{\ell_1=\ell_{\min}}^{\infty} \int_{\beta=0}^{2\pi} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1, \quad (6.2.9)$$

where f_{TA} is the density of the forager's turning angle, and $f(\ell_1)$ is the PDF of the forager's previous step. The area covered by a step is a piecewise function that depends on whether $\ell_1 < -\ell_2 \cos(\beta)$ or $\ell_1 \geq -\ell_2 \cos(\beta)$, so we could write it as two separate integrals:

$$\begin{aligned} f_A(x) = & \int_{\ell_1=\ell_{\min}}^{\ell_2} \int_{\beta=0}^{2\pi} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) f(\ell_2) d\beta d\ell_1 \\ & + \int_{\ell_1=\ell_2}^{\infty} \int_{\beta=0}^{2\pi} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) f(\ell_2) d\beta d\ell_1. \end{aligned}$$

Then, we can further split up the integrals to get

$$\begin{aligned} f_A(x) = & 2 \int_{\ell_1=\ell_{\min}}^{\ell_2} \int_{\beta=0}^{\pi/2} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1 \\ & + 2 \int_{\ell_1=\ell_{\min}}^{\ell_2} \int_{\beta=\pi/2}^{\pi-\arcsin(2r_v/\ell_1)} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1 \\ & + 2 \int_{\ell_1=\ell_{\min}}^{\ell_2} \int_{\beta=\pi-\arcsin(2r_v/\ell_1)}^{\pi} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1 \\ & + 2 \int_{\ell_1=\ell_2}^{\infty} \int_{\beta=0}^{\pi/2} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1 \\ & + 2 \int_{\ell_1=\ell_2}^{\infty} \int_{\beta=\pi/2}^{\pi-\arcsin(2r_v/\ell_2)} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1 \\ & + 2 \int_{\ell_1=\ell_2}^{\infty} \int_{\beta=\pi-\arcsin(2r_v/\ell_2)}^{\pi} f_A(x | \beta, \ell_1) f_{TA}(\beta) f(\ell_1) d\beta d\ell_1, \end{aligned}$$

where we have 6 rather than 12 different terms since we have multiplied each term by 2 due to symmetry. For the first term, note that $f_A(x | \beta, \ell_1) = 2r_v\ell_1 - r_v^2 \tan(\beta/2)$ by Equation 6.2.8. We can rearrange this expression to find the inverse

$$x = 2r_v\ell_2 - r_v^2 \tan(\beta/2) \implies \ell_2 = \frac{x + r_v^2 \tan(\beta/2)}{2r_v}.$$

Thus, we can replace $f_A(x | \beta, \ell_1)$ with $f\left(\frac{x+r_v^2 \tan(\beta/2)}{2r_v}\right)$. We can do similarly for the other integrals. However, note that the third and the sixth integrals have an expression for A which cannot be inverted easily, meaning we must either make further simplifications or solve the integrals numerically.

After solving for $f_A(a)$, we then must substitute the result, along with our piecewise function A into

$$\Pr(\text{not found}) = \int_{A_{\min}}^{\infty} e^{-a\rho} f_A(a) da,$$

which is then used to find $\mathbb{E}[N]$. For the first-order model, every step except the first step has the same probability of finding food, and so if we account for this we can also consider τ as a geometric random variable. We do not proceed any further with this process analytically since the remainder of the work is dependent on the choice of step-length distribution, as well as requiring integration over complicated expressions.

Truncated step-length

As with the zeroth-order model, we need to determine the expected truncated length of a step, given the forager was intending to step a distance ℓ . However, in the first-order model this will depend on the size of the steps as well as the turning-angle and the radius of vision. The probability of finding a target depends on the amount of new area being searched, and so should depend heavily on the geometry formed by the current and previous step. Recall that our expression for A , Equation 6.2.8, required that $\ell_2 \geq r_v$, meaning we do not have a valid expression for the amount of new area covered while the forager is travelling the first r_v distance. We can assume that these values are also valid for the first r_v distance of a step, although this is shown to be fairly inaccurate in Section 6.3. This assumption would allow us to write down some integrals for the truncated step-length, though we would not be able to solve them.

Efficiency

Although we have not been able to solve the first-order model analytically, we can draw some conclusions from the expressions we have found. Consider first the

case when $\ell_1 < \ell_2$. The first two intervals for the area of overlap, for small and medium angles, have no dependence on ℓ_1 or ℓ_2 , apart from changing when an angle is considered large. The area of overlap in the final interval, for large angles, will depend on the length of the previous step, ℓ_1 , since it is the smaller of the two steps. Thus, once ℓ_2 gets larger than ℓ_1 , the area of overlap does not change. The total amount of new area will still increase as $2r_v\ell_2$. When $\ell_2 \leq \ell_1$, the total amount of overlapping area is similar for small and medium angles. The amount of overlapping area for large angles is very large, and increases as ℓ_2 increases. Thus, we can conclude that to maximise the amount of new area covered, a forager should take steps as large as possible, regardless of what the previous step was. This would correspond to, for example, an exponential distribution with $\mu \rightarrow 0$ or a power-law distribution with $\mu \rightarrow 1$, which matches known results for destructive foraging.

■ 6.3 Simulations

■ 6.3.1 Monte Carlo simulations for the first-order area

We use Monte Carlo simulations to estimate the true value of A , and compare with the results of our analytic expression. We first define the rectangle of search area from the previous step, at the following points in clockwise order: $(-\ell_1, r_v)$, $(0, r_v)$, $(0, -r_v)$, and $(-\ell_1, -r_v)$. Then, we can generate points in the area of the second step to determine the area of overlap. For some fixed ℓ_2 and turning-angle β , we choose a uniformly distributed random variable u_1 , between 0 and ℓ_2 , and can then define the point $(u_1 \cos(\beta), u_1 \sin(\beta))$. This point is some random point along the path of the forager's current step. Next, we randomly generate a uniformly distributed random variable u_2 , between $-r_v$ and r_v , to represent any point along the perpendicular radius of vision. Thus, the final location of our point will be at $(u_1 \cos(\beta) + u_2 \sin(\beta), u_1 \sin(\beta) - u_2 \cos(\beta))$.

If the randomly generated point is within the rectangle of the previous step, it is overlapping with the previous step. We randomly generate 50000 points using the above process, and can estimate the proportion of overlapping points as $n/50000$, where n is the number of points that were overlapping. The total area of overlap will therefore be, $(n/50000)2r_v\ell_2$, and so the newly explored area is $A = 2(50000 - n)/nr_v\ell_2$. We take the Monte Carlo simulation to be the true value, and determine the relative error of the analytic expressions based on this. Implicit in this is the assumption that the Monte Carlo simulations produce the exact value for the area, which is not true, although relative to the error

in our analytic expressions the Monte Carlo simulations produce very accurate values. Figure 6.3.1 shows an example of this. We repeat this whole process for various values of ℓ_1 , ℓ_2 , r_v , and β , and compare the found area of overlap with our analytic expressions Equation 6.2.8.

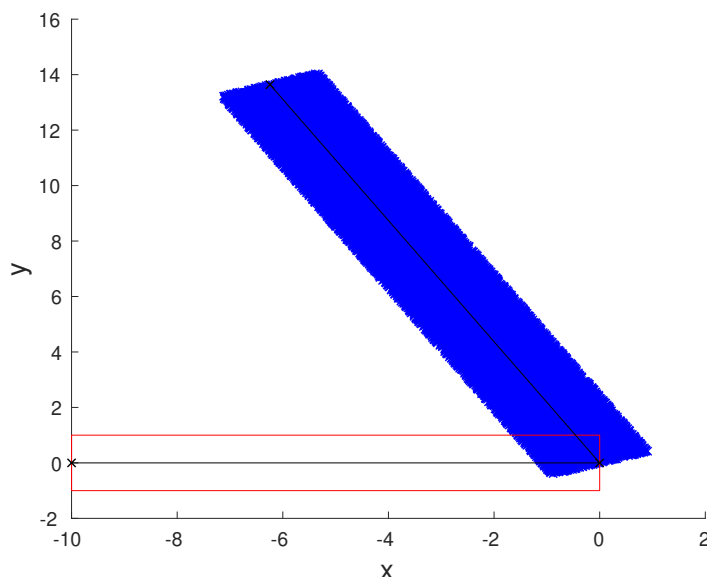


Figure 6.3.1: An example of our Monte Carlo simulation method to determine the amount of overlapping area between two consecutive steps, for $\ell_1 = 10$, $\ell_2 = 15$, $r_v = 1$, and $\beta = 2$. The area covered by the previous step is represented by the red rectangle, and each of the 50000 simulated points are blue crosses. In this case, $n = 243$ points overlap with the red rectangle, and the full area of the current step is $2r_v\ell_2 = 30$, so the area of overlap is $A = 1.458$. The amount of total new area covered is $A = 28.542$ according to this method, and $A = 28.443$ according to our analytic approximation, giving a relative error of 0.0072.

Looking first at Figure 6.3.2a, we see that the relative error in the area is less than 0.05 everywhere except for when $\ell_2 = 2$. When $\ell_2 = 2$, there are two large spikes in the relative error when $\beta \approx 2.2$ and $\beta \approx 4.1$. The reason for the large error at these points are the assumption we made in Chapter B, that $\ell_2 \gg r_v$, and $\ell_1 \approx \ell_2$, both of which are violated at these values, causing the wrong piecewise expression to be considered. This happens for these angles specifically because they are normally close to a threshold angle. Also looking at Figure 6.3.2b, we can draw the same conclusions, with a lower ℓ_2 causing larger errors, as well as large errors occurring at $\beta = 0$, $\beta \approx 2.86$, and $\beta \approx 4.57$.

In Figure 6.3.3 we plot the relative efficiency against the turning-angle as we did in Figure 6.3.2a, except we exclude the lowest two values of ℓ_2 . Now, the two lowest values of $\ell_2 = 5.27$ and $\ell_2 = 6.91$ produce the largest error, at $\beta \approx \pi$. The error at these points is less than 0.04, and the cause for the error is the

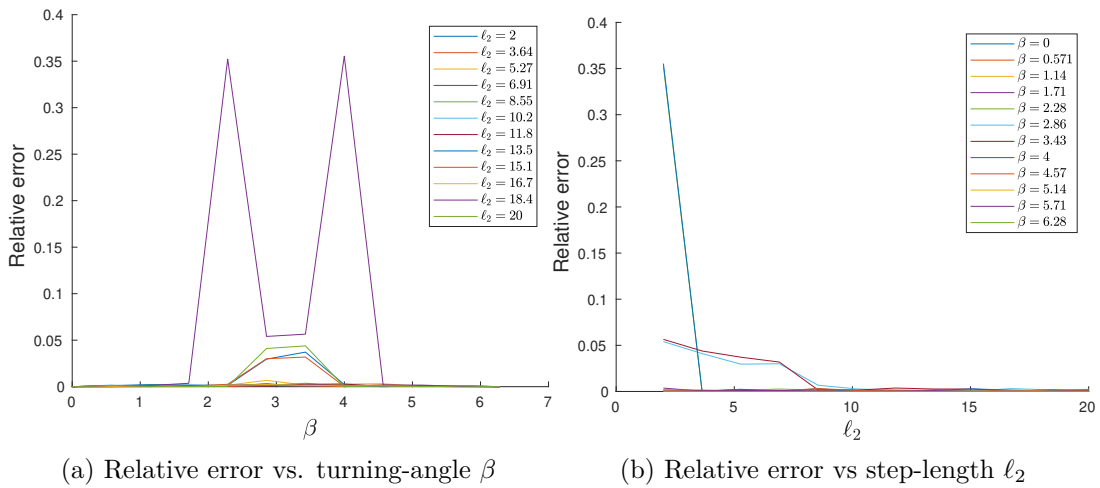


Figure 6.3.2: The relative error plotted against the step-length ℓ_2 , for $2 \leq \ell_2 \leq 20$ at 10 equispaced intervals and against the turning-angle β , for $0 \leq \beta \leq 2\pi$. The radius of vision is $r_v = 1$ and $\ell_1 = 10$. Large relative errors occur at $\beta = 2.86$ and $\beta = 4$ take large values for small values of ℓ_2 , and all other values of ℓ_2 and β produce a relative error less than 0.065.

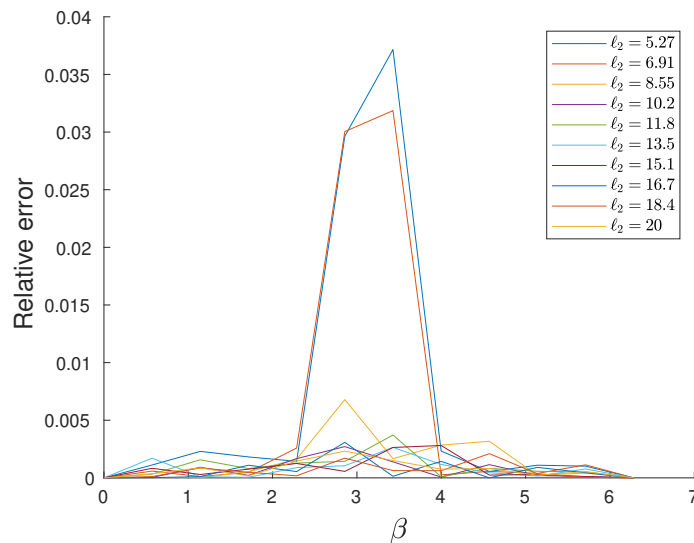


Figure 6.3.3: The relative error plotted against the turning angle β , for $0 \leq \beta \leq 2\pi$ at 10 equispaced intervals. The radius of vision is $r_v = 1$ and $\ell_1 = 10$, and lines are plotted for 10 different values of ℓ_2 , from 5.27 to 20. The lines for $\ell_2 = 5.27$ and $\ell_2 = 6.91$ has large peaks near the centre, for values of β near π , corresponding to a large turning angle, and all other values of ℓ_2 and β produce a relative error less than 0.05.

approximations made for the case of large turning-angles (case 1) in Chapter B. For all other values of β and ℓ_2 , the relative error is lower than 0.01, which is a good level of accuracy.

In summary, the analytic expressions from Equation 6.2.8 are relatively accurate for most values of β and ℓ_2 , with some larger errors occurring when ℓ_2 is small and $\beta \approx \pi$. The effect that these points of inaccuracy will have on the overall calculation will depend on the turning-angle distribution as well as the step-length distribution. For example, a step-length distribution with a very large minimum step-size would always produce large steps and would not suffer from many of these inaccuracies.

■ 6.3.2 Simulating our two-dimensional model

We use Matlab to simulate the model outlined above. We initially use a 50×50 search space, and generate food targets according to a spatial Poisson point process on this search space, with density ρ . We simulate the Poisson process by first drawing a number $N \sim \mathcal{P}(50^2\rho)$, and then drawing N random variables from $x \sim \text{Uni}[0, 50]$ and N random variables from $y \sim \text{Uni}[0, 50]$, and pairing these up to form the coordinates of the targets.

The forager begins at the centre of the search space at $(25, 25)$. To simulate the movement of the forager, we draw an angle with $\theta \sim \text{Uni}[0, 2\pi]$ and a step-length from our step-length distribution using inverse transform sampling. Thus, we are able to simulate a series of coordinates corresponding to the foragers steps. When a forager steps outside of the search area, we extend the search area in multiples of 50 until the foragers new location is within the search area, and generate new targets in the new area in the same way as above.

After each step is taken, we check to see if a target has been located. To do this, we determine the perpendicular distance that each food patch is away from the line segment connecting the start and end points of a step. If we wanted to consider the circle of vision at the start and end of each step, we could also check the distance each point is from these points, in any direction. However, for consistency with our analytic model we exclude these. If any of these distances are less than r_v , then a target is located on that step. The located target is the one that is closest to the starting point of the step.

Figure 6.3.4 shows a single simulation of our model, with the forager using an exponential distribution. This model currently corresponds to the infinite order foraging model since once an area has been explored unsuccessfully, there is no chance of finding points there when reexploring the area. To instead consider our zeroth-order approximation, each time a step fails to locate food we generate targets according to a Poisson distribution as done above, although this time only in the area that was just explored. For the first-order model, we also do this

although there is a lag of 1 step before each regeneration. We can then consider any k th-order approximation by setting the regeneration lag to k .

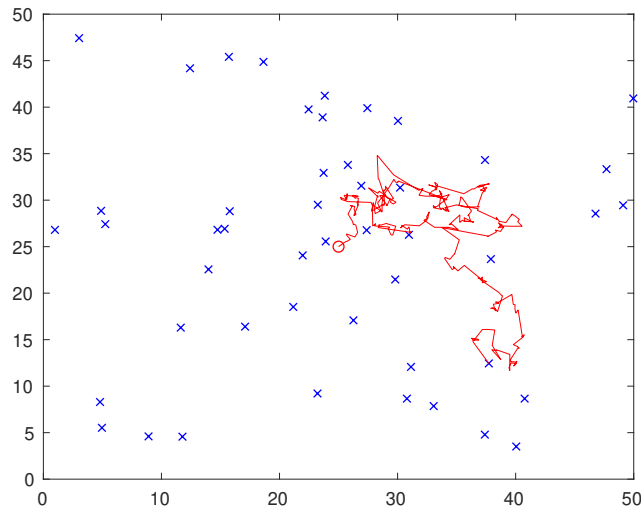


Figure 6.3.4: A simulation of our two-dimensional foraging model, with parameters $r_v = 0.2$, $\rho = 0.02$. The forager takes steps from an exponential distribution with $\ell_{\min} = 0$ and $\mu = 1.8$. No points are regenerated in areas that have already been explored, and thus this corresponds to the infinite order model.

Figure 6.3.5 compares the average step-length, mean number of steps, and efficiency, respectively, for the zeroth-order, first-order and infinite order models for a power-law forager. The efficiency of the simulated zeroth-order model matches the exact expression for the efficiency that we found, Equation 6.2.7. The efficiency of the first-order model decreases slightly as μ increases, and is slightly less efficient than the zeroth-order model, which is to be expected since targets are not being regenerated as quickly. The decrease in efficiency as μ increases also matches our conclusions from the analytic work on the first-order model, where we reasoned that the larger each step was, the less backtracking that would occur and hence more fresh area would be explored. Also, we are considering destructive foraging, for which it is known that ballistic motion offers the best efficiency (e.g. [86]).

The efficiency of the infinite order model decreases much faster than either the zeroth or first-order models. While the zeroth and first-order model seem to be fairly similar to each other, the infinite order model is significantly different to both, indicating that neither the zeroth or first-order models are accurate approximations for the infinite order model, and higher order models are needed. However, the first-order model is already too complicated to fully solve analytically, so an analytic approach to even higher order models is out of the question. Nonetheless, the zeroth-order and the first-order models can be considered as

upper bounds for the infinite-order model.

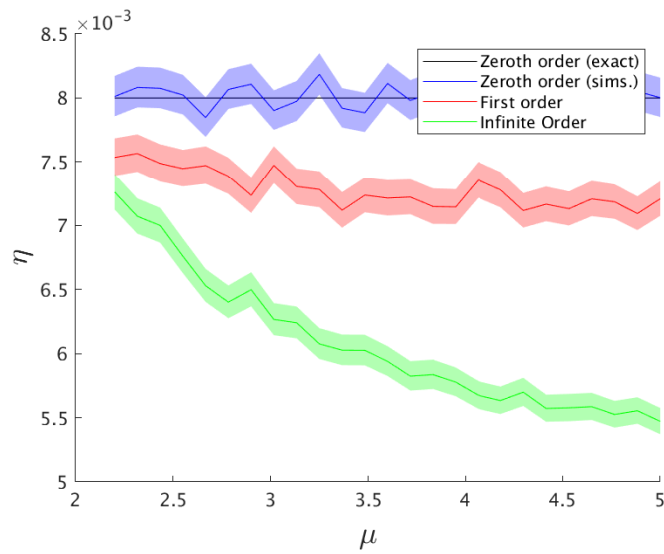


Figure 6.3.5: The mean efficiency for the zeroth, first, and infinite order models with 95% confidence intervals, as well as the exact result found for the zeroth-order model. The forager takes steps from a power-law distribution with $\ell_{\min} = 1$, and for 25 equispaced values of μ between 2.2 and 5. The mean is taken over 10000 repetitions.

Chapter 7

Conclusion

In this thesis, we have considered an existing one-dimensional foraging model and derived expressions for the efficiency of a Markov-modulated random walk strategy. We have shown how these can be discretised and solved numerically. Using these expressions, results for the optimal efficiency were found for a range of different strategies, and we have discussed how some existing strategies can be seen as a special case of our Markov-modulated random walk strategy. We have confirmed the conclusions of these, and in some cases been able to extend the results. We also constructed a simple two-dimensional model and found some basic results about the efficiency of destructive foraging. We now recap the main findings of our thesis, as well as describe some future work.

Chapter 3: One-dimensional foraging model

This chapter was perhaps the most important. We began with an existing one-dimensional foraging model and first considered a random walk search strategy, for which results already exist [6]. We proved a number of these results, such as the expected total distance travelled to find a target. Although already known, we have placed these results on more rigorous grounding. For example, we have shown explicitly that the operator norm of the integral operator we use is less than one, and hence convergence of the Neumann series is guaranteed. Further, we also found a separate expression for the expected total travel length which does not require a known starting location.

We then considered a Markov-modulated random walk search strategy, and derived analogues to the existing expressions for the random walk search strategy, such as the expected total travel length and the expected number of steps to find a target. More than just switching step-length distributions, we also allow

the Markov chain to change the forager's radius of vision. One limitation of our expressions is that we are required to assume that the transition matrix does not depend on the forager's current location. A transition matrix that was a function of the forager's current location would actually allow us to investigate location-aware foragers, such as those of [56]. This would represent a forager that changes strategy depending on where it is relative to a food patch, for which there is empirical evidence of this occurring [58]. Another limitation is that our derivation requires a step-length distribution that is symmetric about zero. If we could consider non-symmetric step-length distributions, we could investigate the effects of an orientational memory by making movements in one direction more likely than the other.

Chapter 4: Discretisation of the one-dimensional search space

In this chapter, we discretised the one-dimensional model from Chapter 3. Following a similar methodology as earlier, we once again use an operator, which we showed could be represented by a matrix A . We find discretised analogues of all of the important results from Chapter 3, and discussed how they could be solved numerically in Matlab. The work of this chapter also helps to demonstrate some of the difficulties that arose in Chapter 3, since we are able to look at the problem in terms of matrices instead of integrals. In particular, the issue with finding the cost incurred in a single state rather than all states is easier to understand from the perspective of a discrete search space.

Chapter 5: Results for the one-dimensional model

In Chapter 5, we implemented the discrete expressions for the expected total area that we derived in Chapter 4, and investigated the efficiency of a range of different strategies. In Section 5.1.1, we considered the unmodulated random walk, which is a special case of the Markov-modulated random walk, when the Markov chain has only a single state. We plotted the efficiency for four different step-length distributions: unbounded power-law, bounded power-law, unbounded exponential, and bounded exponential. For destructive foraging, the optimal strategy was to perform ballistic motion. For non-destructive foraging, ballistic motion was best for the unbounded and bounded exponential distributions, but for the two power-law distributions, a Lévy walk with $\mu \approx 2$ provided a higher efficiency. These results are all consistent with previous results found throughout the literature.

We also investigated the effect that the upper bound has on the search efficiency.

For both the exponential and power-law distributions, the upper bound caused a lower efficiency, and the lower the upper bound was, the worse the efficiency became.

In Section 5.1.2, we showed how giving-up time strategies could be modelled using our Markov-modulated random walk strategy, by constructing the transition matrix to have an absorbing state. Thus, we reasoned that we could use our Markov-modulated random walk model to consider any discrete phase-type distribution for the giving-up time, and could in theory construct our transition matrix in such a way as to approximate any giving-up time distribution, since the phase-type distribution is dense in the field of all positive-valued distributions. However, for accurate approximations, we may require using Markov chains with a large number of states, which quickly becomes computationally infeasible.

We considered a Brownian motion giving-up into a ballistic motion, and found the giving-up time parameter which optimised the efficiency. Our results are relatively close to the existing results by Plank and James [61], given the significantly different approaches taken between us and them.

Next, we fixed the giving-up time and investigated the optimal choice of step-length distribution parameters. We found that a Brownian motion that gives up and performs a Lévy flight is the most efficient, matching the results of Reynolds [66]. The optimal parameter of the Lévy flight depends on the giving-up time, and the faster a forager gives up the larger μ should be, with a maximum being $\mu \approx 2$, corresponding to when a forager instantly gives up which is essentially the same as using a single strategy.

Finally, for the giving-up time strategies, we optimised over both the giving-up time parameter and the step-length distribution parameters, for a two-state strategy. This combines the best of both of the two previously mentioned papers on giving-up time strategies. We found the optimal efficiency occurred when the forager began with a Brownian motion with μ as large as possible, and gave up after a geometrically distributed time ($p = 0.0060$, mean of 166.6) and performed a Lévy flight with $\mu_2 = 1.5868$. As far as we are aware, there is no existing literature that has optimised over both the step-length parameters and the giving-up time at the same time.

We also discussed how we could consider various vision-switching strategies, in Section 5.1.3. We outlined how models with hidden or hard-to-detect targets could be approximated using our model, although this is computationally infeasible for heavy-tailed distributions, making this not very useful. Simpler models in

which the forager's radius of vision switches at various times, which perhaps represents switching between different weather conditions, can easily be modelled with our Markov-modulated random walk strategy. We plotted an example of this, showing how a switching vision skews the symmetrical plot of step-length parameters against searching efficiency for a forager with equal probabilities of switching between each state. Since a model like this is very animal-specific and dependent on the choice of parameters, we do not determine any results for the optimal vision-switching forager.

Finally, in Section 5.2 we investigated the most general models thus far, the Markov-modulated random walk, with any parameters for the transition matrix, initial distribution and step-length parameters. We began with a one-state model, and explain how we can construct a J -state model, showing that this involves $J^2 + J - 1$ parameters to optimise over. We used Matlab's *fmincon* function to determine the parameters that result in the optimal efficiency for the 1-state, 2-state, and 3-state strategies. We also outlined a method for selecting a reasonable starting point for the optimisation, which uses the results of the previous solution.

The optimal solution for the 1-state model simply reiterated the results that we found for the unmodulated random walk, which is $\mu = 1.8699$. For the 2-state model, we found something more interesting. The optimal strategy was to begin with a Brownian motion before switching into a Lévy flight with $\mu = 1.5868$, with the transition matrix having an absorbing state, and initial distribution $(1, 0)$. This is exactly the optimal two-state giving-up time strategy found earlier. Thus, we are able to conclude that giving-up time strategies are optimal for 2-state Markov-modulated random walks. For the 3-state model, we once again find that a giving-up time strategy is optimal. The optimal transition matrix was found to be

$$P = \begin{bmatrix} 0.9849 & 0 & 0.0151 \\ 0 & 1 & 0 \\ 0 & 0.0152 & 0.9848 \end{bmatrix},$$

with an initial probability vector $\mathbf{z}_0 = (1, 0, 0)$, and distribution parameters $\boldsymbol{\mu} = (10, 1.3908, 2.4368)$. This strategy once again corresponds to a giving-up time strategy that begins following a Brownian motion, before giving-up into a Lévy flight, with parameter $\mu = 2.4368$. Then, after some further time, the forager gives up again, and switches to a Lévy flight with $\mu = 1.3908$, which will somewhat resemble a ballistic path, due to the small μ parameter. The mean of the both geometrically distributed giving-up times is approximately 66 steps.

We hypothesise that the optimal intermittent strategies will always be giving-up

time strategies. Further, we suspect that a J -state model will involve $J - 1$ times at which the forager gives up into a different state, with no switching back to previous states. The μ parameters of these states will decrease each time a forager gives up. If this hypothesis is correct, it may perhaps also point to a potential optimal strategy in which a forager begins following a Brownian motion and μ decays continuously over time.

Chapter 6: Two-dimensional model

In Chapter 6 we constructed a simple two-dimensional foraging model with targets distributed with complete spatial randomness. In order to find some analytic results, we introduced a simplifying assumption about target regeneration in areas that have already been explored. When targets can regenerate in areas that were explored previously in the step immediately prior, we call it the zeroth-order model. Similarly, if targets can regenerate after a delay of one step, we call this the first-order model. The case of no regeneration is called the infinite-order model, which involves no approximation.

We found some analytic results for the efficiency of a random walk search strategy in the zeroth-order, which happens to not depend on the choice of step-length distribution at all. This result also corresponds to an upper bound for the efficiency of a strategy. For the first-order model, there is overlap between consecutive steps so we discussed the area of overlap as a function of the turning-angle and the step-lengths. We found approximate expressions for the area of a step, and use Monte Carlo simulations to ensure these are accurate. Ultimately, the expressions for the area are still too complicated to arrive at analytic expressions for the efficiency under the first-order model, though we can conclude that distributions with larger step-lengths will offer the highest efficiency, matching existing results.

Finally, we simulated the zeroth-order, first-order, and infinite-order model numerically, and show that the first-order model does not actually offer a very good approximation of the infinite-order model anyway. We also argued that for analytic results to be found, much simpler two-dimensional models should be considered, such as lattice models.

■ 7.1 Future Work

As discussed above, our work in Chapter 3 required an assumption that the transition matrix did not depend on the forager's current location, preventing us from considering location-aware foragers. However, our model may perhaps

still be able to be used in investigating location-aware foragers, by making some extensions. In the location-aware forager of Nolting [56], the search area is divided into three sections, and the forager would use a different strategy in the left and right sections than the one it used in the middle section. One possible way we could consider this is by actually considering this as multiple separate searches, with the boundaries of the middle section representing a change of strategy rather than a food target. However, this would also require changes to the way our truncation is performed, as well as knowing more about whether it is the left or the right boundary that is found throughout a search.

In terms of considering non-symmetric step-length distributions in order to introduce an orientational memory, there are some possible avenues to explore. There are some key moments throughout Chapter 3 at which the symmetry is needed, in particular, after the introduction of the Dirac delta function. However, recall that we also found expressions that did not require a known starting location, so perhaps the symmetry assumption may be relaxed. This will require more investigation, and will also need to consider how this will affect the discretisation of the step-length distributions, as well as any restrictions introduced to ensure that the operator norm is less than unity.

Recall from Chapter 5 that we outlined a way to approximate the hidden targets model, although it was not adequate. This is because even when our forager has a vision radius of 0, it may still reach the food by running into the target. To alleviate this problem, we can introduce periodic boundary conditions on the interval $[0, \lambda]$. Then, a food target would only be found if the forager landed with the food within its radius of vision, meaning it could skip over targets. To do this, we would have to adjust our expression for $\rho_n(x_n)$, and we now have to consider the previous location $\rho_{n-1}(x_{n-1})$ coming from any other interval, not just the current interval. This will introduce an infinite summation into the $\rho_n(x)$ recursive relation, which will not necessarily converge quickly when we consider heavy-tailed step-length distributions.

In Chapter 5, apart from the unmodulated strategies, we only considered unbounded power-law distributions. This is because the other step-length distributions we considered offered worse performance in the unmodulated case. However, they may still offer better efficiency when considering switching strategies. No further extensions are required to test these out, though there are many possible combinations to consider, which can take a fairly long time to run. Similarly, we can consider 4-state and higher Markov-modulated random walk models, which are also possible using our existing work, though slow computationally.

Further, we could also consider other distributions than the four we have investigated throughout this thesis. Our reasoning in choosing these four was that they are the only four that have been commonly used throughout the literature. However, more recent papers (e.g [17]) have considered a generalized Pareto distribution, of which Normal, exponential, and power-law distributions are special cases.

In terms of two-dimensional models, further work should first focus on simulations. Further investigation should be made into which one-dimensional results hold in higher dimensions. The model we outline in Chapter 6 could also be adjusted to account for non-destructive foraging in a few different ways. For example, rather than have a fully homogeneous spatial Poisson process, there could be different regions, representing areas that are either close or far from a previously visited target. Alternatively, the density of targets may be set high initially and made to decay over time, to represent the chance of reaching the previously visited target diminishing.

Calculating $h(x)$ and A for some step-length distributions

■ A.1 Average step length

As demonstrated in Section 3.2.2, the average step length for any given step-length distribution, where we are taking into account the possibility of truncation, is given by

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell \\ &\quad + (a-r_v) \int_{-\infty}^{-(a-r_v)} p(\ell) d\ell + (\lambda-r_v-a) \int_{(\lambda-r_v-a)}^{\infty} p(\ell) d\ell, \end{aligned} \tag{A.1.1}$$

where the first or second integral become zero for steps beginning near the boundaries at r_v or $\lambda - r_v$, respectively. Similarly, the third and fourth integral's limits are adjusted slightly for steps beginning near either r_v or $\lambda - r_v$, respectively.

Therefore, the average step length for each distribution will have three separate cases to consider, depending on the starting location of a step.

When the distribution also has an upper limit on the step size, ℓ_{\max} , then we will also have to make further adjustments to the limits of integration, which will result in even more cases to consider.

We can rewrite Equation A.1.1 for this situation, as four separate cases, depending on the value of ℓ_{\max} . Firstly, if $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell \\ &\quad + (a-r_v) \int_{-\ell_{\max}}^{-(a-r_v)} p(\ell) d\ell + (\lambda-r_v-a) \int_{(\lambda-r_v-a)}^{\ell_{\max}} p(\ell) d\ell. \end{aligned} \tag{A.1.2}$$

If $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$,

$$\begin{aligned}\mathbb{E}[|\ell|] &= \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\ell_{\max}} |\ell| p(\ell) d\ell \\ &\quad + (a - r_v) \int_{-\ell_{\max}}^{-(a-r_v)} p(\ell) d\ell.\end{aligned}$$

If $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$,

$$\begin{aligned}\mathbb{E}[|\ell|] &= \int_{-\ell_{\max}}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{(\lambda-r_v-a)}^{\ell_{\max}} p(\ell) d\ell.\end{aligned}$$

If $\ell_{\max} < a - r_v$ and $\ell_{\max} < \lambda - a - r_v$,

$$\mathbb{E}[|\ell|] = \int_{-\ell_{\max}}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\ell_{\max}} |\ell| p(\ell) d\ell.$$

We can also note that as $\ell_{\max} \rightarrow \infty$, Equation A.1.2 approaches Equation A.1.1 as expected.

These above expressions are valid when the forager is not near a boundary, that is, when $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$. When the forager is close to the boundaries, we must adjust some of the limits of integration. Since the first two integrals in Equation A.1.2 represent steps that land within the boundary, which aren't necessarily possible when next to the boundary, we must remove some of the integrals too.

For the left boundary, $r_v < a < r_v + \ell_{\min}$, we get when $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$,

$$\begin{aligned}\mathbb{E}[|\ell|] &= \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\ell_{\max}}^{-\ell_{\min}} p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{(\lambda-r_v-a)}^{\ell_{\max}} p(\ell) d\ell.\end{aligned}$$

If $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$,

$$\mathbb{E}[|\ell|] = \int_{\ell_{\min}}^{\ell_{\max}} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\ell_{\max}}^{-\ell_{\min}} p(\ell) d\ell.$$

The case of $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$ does not exist, since it would require $\ell_{\max} < \ell_{\min}$, and likewise for the case of $\ell_{\max} < a - r_v$ and $\ell_{\max} < \lambda - a - r_v$.

Based on these, we may also conclude that for a step-length distribution with no upper bound on the step size, at the left boundary we get,

$$\begin{aligned}\mathbb{E}[|\ell|] &= \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\infty}^{-\ell_{\min}} p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{(\lambda-r_v-a)}^{\infty} p(\ell) d\ell.\end{aligned}\tag{A.1.3}$$

For steps near the right boundary, $\lambda - r_v - \ell_{\min} < a < \lambda - r_v$, we get when $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$,

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\ell_{\max}}^{-(a-r_v)} p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{\ell_{\min}}^{\ell_{\max}} p(\ell) d\ell. \end{aligned}$$

If $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$,

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{-\ell_{\max}}^{-\ell_{\min}} |\ell| p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{\ell_{\min}}^{\ell_{\max}} p(\ell) d\ell. \end{aligned}$$

The case of $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$ does not exist, since it would require $\ell_{\max} < \ell_{\min}$, and likewise for the case of $\ell_{\max} < a - r_v$ and $\ell_{\max} < \lambda - a - r_v$,

Based on these, we may also conclude that for a step-length distribution with no upper bound on the step size, near the right boundary we get,

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{\ell_{\min}}^{\lambda - r_v - a} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\infty}^{-\ell_{\min}} p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{(\lambda - r_v - a)}^{\infty} p(\ell) d\ell. \end{aligned} \tag{A.1.4}$$

Thus, all up we have 8 different cases to consider for each distribution — two near the left boundary, four in the centre, and two near the right boundary. If the distribution doesn't have a maximum step size, then we have only 3 cases to consider. Fortunately, all 10 cases are made up of combinations of integrals which only take 6 different forms, 3 of which are:

$$\int_c^d |\ell| p(\ell) d\ell, \quad c, d \in \mathbb{R}^+, \tag{A.1.5}$$

$$\int_c^d p(\ell) d\ell, \quad c, d \in \mathbb{R}^+, \tag{A.1.6}$$

$$\int_c^{\infty} p(\ell) d\ell, \quad c \in \mathbb{R}^+, \tag{A.1.7}$$

and then, due to the symmetry of $p(\ell)$, the remaining three integrals are equivalent to integrals A.1.5, A.1.6, and A.1.7, respectively:

$$\int_{-d}^{-c} |\ell| p(\ell) d\ell, \quad c, d \in \mathbb{R}^+, \tag{A.1.8}$$

$$\int_{-d}^{-c} p(\ell) d\ell, \quad c, d \in \mathbb{R}^+, \tag{A.1.9}$$

$$\int_{-\infty}^{-c} p(\ell) d\ell, \quad c \in \mathbb{R}^+. \tag{A.1.10}$$

Thus, finding the expected cost of a step for each step-length distribution amounts to not much more than solving integrals A.1.5, A.1.6, and A.1.7 and rearranging. Furthermore, for distributions *without* an upper bound on the step size, integrals of the same form as integral A.1.6 never appear, and distributions *with* an upper bound have no integrals of the same form as integral A.1.7.

As discussed in Chapter 4, to find the discretized equivalents of these expressions, we make the replacements:

$$\ell_{\min} = m_0 \Delta x, \quad \ell_{\max} = m_m \Delta x, \quad r_v = m_r \Delta x, \quad m_0, m_m, m_r \in \mathbb{Z}.$$

We find the exact expressions for the unbounded power-law distribution, and show that there are some small differences between our results and the expressions listed by Bartumeus *et al.* [6], which we explain is most likely due to a typo. We also present the discretized equivalents of the expressions for the unbounded power-law distribution.

For the other step-length distributions, we simply present the solutions without showing any of the working.

■ A.1.1 Unbounded power-law distribution

Recall that an unbounded power-law distribution, where we allow both positive and negative values is:

$$p(\ell) = C |\ell|^{-\mu}, \quad |\ell| \geq \ell_{\min},$$

where

$$C = \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1}.$$

We can represent the support of the distribution by rewriting this as

$$p(\ell) = C |\ell|^{-\mu} \theta(|\ell| - \ell_{\min}),$$

and

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For the unbounded power-law distribution, integral A.1.5 is given by

$$\begin{aligned} \int_c^d |\ell| p(\ell) d\ell &= \int_c^d \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} |\ell|^{1-\mu} d\ell \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \int_c^d (\ell)^{1-\mu} d\ell \end{aligned}$$

for $\mu \neq 2$:

$$\begin{aligned} \int_c^d |\ell| p(\ell) d\ell &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{\ell^{2-\mu}}{2-\mu} \right]_c^d \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{d^{2-\mu} - c^{2-\mu}}{2-\mu} \right] \end{aligned}$$

and for $\mu = 2$:

$$\begin{aligned} \int_c^d |\ell| p(\ell) d\ell &= \frac{\ell_{\min}}{2} [\log(\ell)]_c^d \\ &= \frac{\ell_{\min}}{2} [\log(d) - \log(c)] \\ &= \frac{\ell_{\min}}{2} \log\left(\frac{d}{c}\right). \end{aligned}$$

Integral A.1.7 is given by

$$\begin{aligned} \int_c^\infty p(\ell) d\ell &= \int_c^\infty \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} |\ell|^{-\mu} d\ell \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \int_c^\infty (\ell)^{-\mu} d\ell \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{\ell^{1-\mu}}{1-\mu} \right]_c^\infty \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{-c^{1-\mu}}{1-\mu} \right] \\ &= \frac{1}{2} \left(\frac{c}{\ell_{\min}} \right)^{1-\mu}. \end{aligned}$$

For the unbounded power-law distribution we have no upper bound on the step length, and hence there are only three cases to consider: near the left boundary, in the centre, and near the right boundary, which correspond to Equation A.1.3, Equation A.1.1, Equation A.1.4, respectively. However, we do get two different expressions for each, depending on whether or not $\mu = 2$, making six expressions total.

For steps beginning near the left boundary, specifically for $r_v < a < r_v + \ell_{\min}$, when $\mu \neq 2$ we get

$$\begin{aligned} \mathbb{E}[|\ell|] &= \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\infty}^{-\ell_{\min}} p(\ell) d\ell \\ &\quad + (\lambda - r_v - a) \int_{(\lambda-r_v-a)}^{\infty} p(\ell) d\ell \\ &= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{(\lambda - r_v - a)^{2-\mu} - \ell_{\min}^{2-\mu}}{2-\mu} \right] + (a - r_v) \frac{1}{2} \left(\frac{\ell_{\min}}{\ell_{\min}} \right)^{1-\mu} \end{aligned}$$

$$\begin{aligned}
& + (\lambda - r_v - a) \frac{1}{2} \left(\frac{\lambda - r_v - a}{\ell_{\min}} \right)^{1-\mu} \\
& = \frac{(a - r_v)}{2} + \frac{\ell_{\min}(1 - \mu)}{2(2 - \mu)} \left[1 + \frac{((\lambda - a - r_v)/\ell_{\min})^{2-\mu}}{1 - \mu} \right], \quad (\text{A.1.11})
\end{aligned}$$

and when $\mu = 2$ we get

$$\begin{aligned}
\mathbb{E} [|\ell|] & = \frac{\ell_{\min}}{2} \left[\log \left(\frac{\lambda - a - r_v}{\ell_{\min}} \right) + 1 \right] + \frac{(a - r_v)}{2} \\
& = \frac{(a - r_v)}{2} + \frac{\ell_{\min}}{2} [1 + \log((\lambda - a - r_v)/\ell_{\min})], \quad (\text{A.1.12})
\end{aligned}$$

which agree with the expressions from Bartumeus *et al.* [6].

For the middle of the search space, which is valid for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$, when $\mu \neq 2$ we get

$$\begin{aligned}
\mathbb{E} [|\ell|] & = \int_{-(a-r_v)}^{-\ell_{\min}} |\ell| p(\ell) d\ell + \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell \\
& + (a - r_v) \int_{-\infty}^{-(a-r_v)} p(\ell) d\ell + (\lambda - r_v - a) \int_{(\lambda-r_v-a)}^{\infty} p(\ell) d\ell \\
& = \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{(a - r_v)^{2-\mu} - \ell_{\min}^{2-\mu}}{2 - \mu} \right] + \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{(\lambda - r_v - a)^{2-\mu} - \ell_{\min}^{2-\mu}}{2 - \mu} \right] \\
& + (a - r_v) \frac{1}{2} \left(\frac{a - r_v}{\ell_{\min}} \right)^{1-\mu} + (\lambda - r_v - a) \frac{1}{2} \left(\frac{\lambda - r_v - a}{\ell_{\min}} \right)^{1-\mu} \\
& = \frac{\ell_{\min}(1 - \mu)}{2(2 - \mu)} \left[2 + \frac{((a - r_v)/\ell_{\min})^{2-\mu}}{1 - \mu} + \frac{((\lambda - a - r_v)/\ell_{\min})^{2-\mu}}{1 - \mu} \right], \quad (\text{A.1.13})
\end{aligned}$$

and for $\mu = 2$ we get

$$\mathbb{E} [|\ell|] = \ell_{\min} \left[1 + \log \left(\frac{[(\lambda - a - r_v)(a - r_v)]^{1/2}}{\ell_{\min}} \right) \right]. \quad (\text{A.1.14})$$

These two expressions are not consistent with Bartumeus *et al.* [6].

For steps beginning near the right boundary, specifically for $\lambda - r_v - \ell_{\min} < a < \lambda - r_v$, when $\mu \neq 2$ we get

$$\begin{aligned}
\mathbb{E} [|\ell|] & = \int_{\ell_{\min}}^{\lambda-r_v-a} |\ell| p(\ell) d\ell + (a - r_v) \int_{-\infty}^{-(a-r_v)} p(\ell) d\ell \\
& + (\lambda - r_v - a) \int_{\ell_{\min}}^{\infty} p(\ell) d\ell
\end{aligned}$$

$$\begin{aligned}
&= \frac{(\mu - 1)}{2} \ell_{\min}^{\mu-1} \left[\frac{(\lambda - r_v - a)^{2-\mu} - \ell_{\min}^{2-\mu}}{2 - \mu} \right] + (a - r_v) \frac{1}{2} \left(\frac{a - r_v}{\ell_{\min}} \right)^{1-\mu} \\
&+ (\lambda - r_v - a) \frac{1}{2} \left(\frac{\ell_{\min}}{\ell_{\min}} \right)^{1-\mu} \\
&= \frac{(\lambda - a - r_v)}{2} + \frac{\ell_{\min}(1 - \mu)}{2(2 - \mu)} \left[1 + \frac{((a - r_v)/\ell_{\min})^{2-\mu}}{1 - \mu} \right], \tag{A.1.15}
\end{aligned}$$

and for $\mu = 2$ we get

$$\mathbb{E} [|\ell|] = \frac{(\lambda - a - r_v)}{2} + \frac{\ell_{\min}}{2} [1 + \log((a - r_v)/\ell_{\min})]. \tag{A.1.16}$$

These two expressions are also not consistent with Bartumeus *et al.* [6]. However, after solving all three cases, we can now see that they have not made an error with their mathematics, but rather, have listed the middle expression as the left expression, and vice versa.

The discretized versions of Equation A.1.11, Equation A.1.12, Equation A.1.13, Equation A.1.14, Equation A.1.15, Equation A.1.16 are listed below.

When $\mu \neq 2$, for the left, middle, and right sections, respectively:

$$\begin{aligned}
\mathbb{E} [|\ell|]_{i_a} &= \frac{(i_a - m_r)\Delta x}{2} + \frac{m_0\Delta x(1 - \mu)}{2(2 - \mu)} \left[1 + \frac{((M - i_a - m_r)/m_0)^{2-\mu}}{1 - \mu} \right], \\
\mathbb{E} [|\ell|]_{i_a} &= \frac{m_0\Delta x(1 - \mu)}{2(2 - \mu)} \left[2 + \frac{((i_a - m_r)/m_0)^{2-\mu}}{1 - \mu} + \frac{((M - i_a - m_r)/m_0)^{2-\mu}}{1 - \mu} \right], \\
\mathbb{E} [|\ell|]_{i_a} &= \frac{(M - i_a - m_r)\Delta x}{2} + \frac{m_0\Delta x(1 - \mu)}{2(2 - \mu)} \left[1 + \frac{((i_a - m_r)/m_0)^{2-\mu}}{1 - \mu} \right].
\end{aligned}$$

When $\mu = 2$, for the left, middle, and right sections, respectively:

$$\begin{aligned}
\mathbb{E} [|\ell|]_{i_a} &= \frac{(i_a - m_r)\Delta x}{2} + \frac{m_0\Delta x}{2} [1 + \log((M - i_a - m_r)/m_0)], \\
\mathbb{E} [|\ell|]_{i_a} &= m_0\Delta x \left[1 + \log \left(\frac{[(M - i_a - m_r)(i_a - m_r)]^{1/2}}{m_0} \right) \right], \\
\mathbb{E} [|\ell|]_{i_a} &= \frac{(M - i_a - m_r)\Delta x}{2} + \frac{m_0\Delta x}{2} [1 + \log((i_a - m_r)/m_0)].
\end{aligned}$$

■ A.1.2 Bounded power-law distribution

A bounded power-law distribution, where we allow both positive and negative values is given by:

$$p(\ell) = C|\ell|^{-\mu}, \quad |\ell| \in [\ell_{\min}, \ell_{\max}],$$

where

$$C = \frac{(\mu - 1)}{2(\ell_{\min}^{1-\mu} - \ell_{\max}^{1-\mu})}.$$

As with the unbounded power-law distribution, we may represent the support of the distribution using the function θ , giving

$$p(\ell) = C|\ell|^{-\mu}\theta(|\ell| - \ell_{\min})\theta(\ell_{\max} - |\ell|),$$

where

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For the bounded power-law distribution, we will have 10 different cases to consider. Recall that to solve all of these expressions, we essentially only need to solve two different integrals.

For steps beginning near the left boundary, specifically for $r_v < a < r_v + \ell_{\min}$ and when $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get for $\mu \neq 2$:

$$\mathbb{E}[|\ell|] = \frac{(\mu - 1)}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} \left(\frac{(\lambda - r_v - a)^{2-\mu}}{\mu - 1} - \ell_{\min}^{2-\mu} + \frac{(\mu - 2)}{(\mu - 1)}(\lambda - r_v - a)\ell_{\max}^{1-\mu} \right) + \frac{a - r_v}{2},$$

and if $\mu = 2$:

$$\mathbb{E}[|\ell|] = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \left(\log\left(\frac{\lambda - r_v - a}{\ell_{\min}}\right) + 1 - \frac{\lambda - r_v - a}{\ell_{\max}} \right) + \frac{a - r_v}{2}$$

For steps beginning near the left boundary and for $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$, when $\mu \neq 2$:

$$\mathbb{E}[|\ell|] = \frac{(\mu - 2)(\ell_{\max}^{2-\mu} - \ell_{\min}^{2-\mu})}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} + \frac{a - r_v}{2}$$

and when $\mu = 2$:

$$E|\ell| = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \log\left(\frac{\ell_{\max}}{\ell_{\min}}\right) + \frac{a - r_v}{2}$$

Both cases when $\ell_{\max} < a - r_v$ do not exist for steps beginning near the left boundary.

Next, we consider steps in the middle of the search space, not near either boundary, specifically for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$. When $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get for $\mu \neq 2$:

$$\begin{aligned} \mathbb{E} [|\ell|] &= \frac{(a - r_v)^{2-\mu} + (\lambda - r_v - a)^{2-\mu} - 2(\mu - 1)\ell_{\min}^{2-\mu}}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} \\ &\quad + \frac{\ell_{\max}^{1-\mu}}{2(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} ((a - r_v) + (\lambda - r_v - a)), \end{aligned}$$

and if $\mu = 2$:

$$\mathbb{E} [|\ell|] = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \left(\log \left(\frac{a - r_v}{\ell_{\min}} \right) + \log \left(\frac{\lambda - r_v - a}{\ell_{\min}} \right) - \frac{\lambda - 2r_v}{\ell_{\max}} + 2 \right).$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, when $\mu \neq 2$:

$$\mathbb{E} [|\ell|] = \frac{(\mu - 1)}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} \left(\ell_{\max}^{2-\mu} + \frac{(\lambda - r_v - a)^{2-\mu}}{\mu - 1} - 2\ell_{\min}^{2-\mu} + \frac{\mu - 2}{\mu - 1}(\lambda - r_v - a)\ell_{\max}^{1-\mu} \right)$$

and when $\mu = 2$:

$$\mathbb{E} [|\ell|] = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \left(\log \left(\frac{\ell_{\max}}{\ell_{\min}} \right) + \log \left(\frac{\lambda - r_v - a}{\ell_{\min}} \right) - \frac{\lambda - r_v - a}{\ell_{\max}} + 1 \right).$$

For $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$, when $\mu \neq 2$:

$$\mathbb{E} [|\ell|] = \frac{(\mu - 1)}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} \left(\ell_{\max}^{2-\mu} + \frac{(a - r_v)^{2-\mu}}{\mu - 1} - 2\ell_{\min}^{2-\mu} + \frac{(\mu - 2)\ell_{\max}^{1-\mu}(a - r_v)}{\mu - 1} \right),$$

and when $\mu = 2$:

$$\mathbb{E} [|\ell|] = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \left(\log \left(\frac{\ell_{\max}(a - r_v)}{\ell_{\min}^2} \right) - \frac{a - r_v}{\ell_{\max}} + 1 \right)$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} < \lambda - a - r_v$, when $\mu \neq 2$:

$$\mathbb{E} [|\ell|] = \frac{(\mu - 1)(\ell_{\max}^{2-\mu} - \ell_{\min}^{2-\mu})}{(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})},$$

and when $\mu = 2$:

$$\mathbb{E} [|\ell|] = \frac{\ell_{\min}\ell_{\max}}{(\ell_{\max} - \ell_{\min})} \log \left(\frac{\ell_{\max}}{\ell_{\min}} \right).$$

Finally, we consider steps beginning near the right boundary, specifically for $\lambda - r_v - \ell_{\min} < a < \lambda - r_v$. When $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get for $\mu \neq 2$:

$$\mathbb{E} [|\ell|] = \frac{(\mu - 1)}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} \left(\frac{(a - r_v)^{2-\mu}}{\mu - 1} + \frac{(\mu - 2)}{(\mu - 1)}(a - r_v)\ell_{\max}^{1-\mu} - \ell_{\min}^{2-\mu} \right) + \frac{\lambda - r_v - a}{2}$$

and if $\mu = 2$:

$$E|\ell| = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \left(\log \left(\frac{a - r_v}{\ell_{\min}} \right) + 1 - \frac{a - r_v}{\ell_{\max}} \right) + \frac{\lambda - r_v - a}{2}$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, when $\mu \neq 2$:

$$\mathbb{E}[|\ell|] = \frac{(\mu - 2)(\ell_{\max}^{2-\mu} - \ell_{\min}^{2-\mu})}{2(\mu - 2)(\ell_{\max}^{1-\mu} - \ell_{\min}^{1-\mu})} + \frac{\lambda - r_v - a}{2}$$

and when $\mu = 2$:

$$\mathbb{E}[|\ell|] = \frac{\ell_{\min}\ell_{\max}}{2(\ell_{\max} - \ell_{\min})} \log \left(\frac{\ell_{\max}}{\ell_{\min}} \right) + \frac{\lambda - r_v - a}{2}$$

Both cases when $\ell_{\max} < \lambda - a - r_v$ do not exist for steps beginning near the right boundary.

■ A.1.3 Unbounded exponential distribution

The probability density function of an unbounded exponential distribution, where we allow both positive and negative values is:

$$p(\ell) = \frac{\mu e^{-\mu(|\ell| - \ell_{\min})}}{2}, \quad |\ell| \geq \ell_{\min}.$$

We can represent the support of the distribution by rewriting this as

$$p(\ell) = \frac{\mu e^{-\mu(|\ell| - \ell_{\min})} \theta(|\ell| - \ell_{\min})}{2},$$

and

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For the unbounded exponential distribution we have no upper bound on the step length, and hence there are only three cases to consider: near the left boundary, in the centre, and near the right boundary, which correspond to Equation A.1.3, Equation A.1.1, Equation A.1.4, respectively.

For steps beginning near the left boundary, specifically for $r_v < a < r_v + \ell_{\min}$, we get

$$\mathbb{E}[|\ell|] = \frac{1 + \mu(a - r_v + \ell_{\min}) - e^{-\mu(\lambda - r_v - \ell_{\min} - a)}}{2\mu}. \quad (\text{A.1.17})$$

For the middle of the search space, which is valid for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$, we get

$$\mathbb{E}[|\ell|] = \frac{2(\mu\ell_{\min} + 1) - e^{-\mu(a-r_v-\ell_{\min})} - e^{-\mu(\lambda-r_v-\ell_{\min}-a)}}{2\mu}. \quad (\text{A.1.18})$$

For steps beginning near the right boundary, specifically for $\lambda - r_v - \ell_{\min} < a < \lambda - r_v$, we get

$$\mathbb{E}[|\ell|] = \frac{1 + \mu(\lambda - r_v - a + \ell_{\min}) - e^{-\mu(a-r_v-\ell_{\min})}}{2\mu}. \quad (\text{A.1.19})$$

The discretized versions of Equation A.1.17, Equation A.1.18, Equation A.1.19 are listed below.

For the left, middle, and right sections, respectively:

$$\mathbb{E}[|\ell|]_{i_a} = \frac{1 + \mu(i_a - m_r + m_0)\Delta x - e^{-\mu(M-m_r-m_0-i_a)\Delta x}}{2\mu},$$

$$\mathbb{E}[|\ell|]_{i_a} = \frac{2(\mu m_0 \Delta x + 1) - e^{-\mu(i_a - m_r - m_0)\Delta x} - e^{-\mu(M - m_r - m_0 - i_a)\Delta x}}{2\mu},$$

$$\mathbb{E}[|\ell|]_{i_a} = \frac{1 + \mu(M - m_r - i_a + m_0)\Delta x - e^{-\mu(i_a - m_r - m_0)\Delta x}}{2\mu}.$$

■ A.1.4 Bounded exponential distribution

A bounded exponential distribution, where we allow both positive and negative values is given by:

$$p(\ell) = C e^{-\mu|\ell|}, \quad |\ell| \in [\ell_{\min}, \ell_{\max}],$$

where

$$C = \frac{\mu}{2(e^{-\mu\ell_{\min}} - e^{-\mu\ell_{\max}})}.$$

As with the other distributions, we may represent the support of the distribution using the function θ , giving

$$p(\ell) = C e^{-\mu|\ell|} \theta(|\ell| - \ell_{\min}) \theta(\ell_{\max} - |\ell|),$$

where

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We begin by considering steps beginning near the left boundary, specifically for $r_v < a < r_v + \ell_{\min}$. When $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get:

$$\mathbb{E} [|\ell|] = \frac{a - r_v}{2} + \frac{(\mu \ell_{\min} + 1)e^{-\mu \ell_{\min}} - \mu(\lambda - r_v - a)e^{-\mu \ell_{\max}} - e^{-\mu(\lambda - r_v - a)}}{2\mu(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

For $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$:

$$\mathbb{E} [|\ell|] = \frac{a - r_v}{2} + \frac{1}{2\mu} + \frac{\ell_{\min}e^{-\mu \ell_{\min}} - \ell_{\max}e^{-\mu \ell_{\max}}}{2(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

Both cases where $\ell_{\max} < a - r_v$ do not exist.

Next, we consider steps in the middle of the search space, not near either boundary, specifically for $r_v + \ell_{\min} \leq a \leq \lambda - r_v - \ell_{\min}$. When $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get:

$$\mathbb{E} [|\ell|] = \frac{2(\mu \ell_{\min} + 1)e^{-\mu \ell_{\min}} - e^{-\mu(a - r_v)} - e^{-\mu(\lambda - r_v - a)} - \mu(\lambda - 2r_v)e^{-\mu \ell_{\max}}}{2\mu(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$:

$$\mathbb{E} [|\ell|] = \frac{2(\mu \ell_{\min} + 1)e^{-\mu \ell_{\min}} - (\mu(\ell_{\max} + \lambda - r_v - a) + 1)e^{-\mu \ell_{\max}} - e^{-\mu(\lambda - r_v - a)}}{2\mu(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

For $\ell_{\max} \geq a - r_v$ and $\ell_{\max} < \lambda - a - r_v$:

$$\mathbb{E} [|\ell|] = \frac{2(\mu \ell_{\min} + 1)e^{-\mu \ell_{\min}} - (\mu(\ell_{\max} + a - r_v) + 1)e^{-\mu \ell_{\max}} - e^{-\mu(a - r_v)}}{2\mu(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} < \lambda - a - r_v$:

$$\mathbb{E} [|\ell|] = \frac{1}{\mu} + \frac{\ell_{\min}e^{-\mu \ell_{\min}} - \ell_{\max}e^{-\mu \ell_{\max}}}{(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

Finally, we consider steps beginning near the right boundary, specifically for $\lambda - r_v - \ell_{\min} < a < \lambda - r_v$. When $\ell_{\max} \geq a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$, we get:

$$\mathbb{E} [|\ell|] = \frac{\lambda - r_v - a}{2} + \frac{(\mu \ell_{\min} + 1)e^{-\mu \ell_{\min}} - \mu(a - r_v)e^{-\mu \ell_{\max}} - e^{-\mu(a - r_v)}}{2\mu(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

For $\ell_{\max} < a - r_v$ and $\ell_{\max} \geq \lambda - a - r_v$:

$$\mathbb{E} [|\ell|] = \frac{\lambda - r_v - a}{2} + \frac{1}{2\mu} + \frac{\ell_{\min}e^{-\mu \ell_{\min}} - \ell_{\max}e^{-\mu \ell_{\max}}}{2(e^{-\mu \ell_{\min}} - e^{-\mu \ell_{\max}})}.$$

Both cases when $\ell_{\max} < \lambda - a - r_v$ do not exist.

■ A.2 Discretized probability distribution matrix

The elements of the matrix A are derived using

$$[A]_{k,j} = [A]_{j,k} = \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} p(\ell) d\ell, \quad k \neq j.$$

■ A.2.1 Unbounded power-law distribution

For an unbounded power law distribution we get, $[A]_{k,j} = 0$ for $|k-j| < m_0$ and for $|k-j| \geq m_0$:

$$\begin{aligned} [A]_{k,j} &= \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \frac{\mu-1}{2} \ell_{\min}^{\mu-1} |\ell|^{-\mu} d\ell \\ &= \frac{\mu-1}{2} \ell_{\min}^{\mu-1} \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \ell^{-\mu} d\ell \\ &= \frac{\mu-1}{2} \ell_{\min}^{\mu-1} \left[\frac{\ell^{-\mu+1}}{-\mu+1} \right]_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \\ &= \frac{\ell_{\min}^{\mu-1}}{2} \left[\frac{1}{(|k-j|\Delta x)^{\mu-1}} - \frac{1}{((|k-j|+1)\Delta x)^{\mu-1}} \right] \\ &= \frac{m_0^{\mu-1}}{2} \left[\frac{1}{(|k-j|)^{\mu-1}} - \frac{1}{(|k-j|+1)^{\mu-1}} \right], \end{aligned}$$

which differs from the result in [6] by the factor of $m_0^{\mu-1}$ at the front.

■ A.2.2 Bounded power-law distribution

For a bounded power law distribution we get, $[A]_{k,j} = 0$ for $|k-j| < m_0$ or $|k-j| > m_m - 1$, and for $m_0 \leq |k-j| \leq m_m - 1$:

$$\begin{aligned} [A]_{k,j} &= \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \frac{\mu-1}{2(\ell_{\min}^{1-\mu} - \ell_{\max}^{1-\mu})} |\ell|^{-\mu} d\ell \\ &= \frac{\mu-1}{2(\ell_{\min}^{1-\mu} - \ell_{\max}^{1-\mu})} \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \ell^{-\mu} d\ell \\ &= \frac{\mu-1}{2(\ell_{\min}^{1-\mu} - \ell_{\max}^{1-\mu})} \left[\frac{\ell^{-\mu+1}}{-\mu+1} \right]_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \\ &= \frac{1}{2(\ell_{\min}^{1-\mu} - \ell_{\max}^{1-\mu})} \left[\frac{1}{(|k-j|\Delta x)^{\mu-1}} - \frac{1}{((|k-j|+1)\Delta x)^{\mu-1}} \right] \\ &= \frac{(|k-j|)^{1-\mu} - (|k-j|+1)^{1-\mu}}{2(m_0^{1-\mu} - m_m^{1-\mu})}, \end{aligned}$$

■ A.2.3 Unbounded exponential distribution

For the exponential distribution, the matrix A is given by

$$\begin{aligned}
[A]_{k,j} &= \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \frac{\mu e^{-\mu(|\ell|-\ell_{\min})}}{2} d\ell, \quad \text{for } |k-j| \geq m_0 \\
&= \frac{\mu e^{\mu\ell_{\min}}}{2} \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} e^{-\mu\ell} d\ell \\
&= \frac{\mu e^{\mu\ell_{\min}}}{2} \left[\frac{e^{-\mu\ell}}{-\mu} \right]_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \\
&= \frac{e^{\mu\ell_{\min}}}{2} \left[e^{-\mu|k-j|\Delta x} - e^{-\mu(|k-j|+1)\Delta x} \right] \\
&= \frac{e^{\mu(m_0-|k-j|)\Delta x}}{2} \left[1 - e^{-\mu\Delta x} \right],
\end{aligned}$$

and $[A]_{k,j} = 0$ for $|k-j| < m_0$.

■ A.2.4 Bounded exponential distribution

For the bounded exponential distribution, the matrix A is given by

$$\begin{aligned}
[A]_{k,j} &= \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \frac{\mu e^{-\mu|\ell|}}{2(e^{-\mu\ell_{\min}} - e^{-\mu\ell_{\max}})} d\ell, \quad \text{for } |k-j| \geq m_0 \\
&= \frac{\mu}{2(e^{-\mu\ell_{\min}} - e^{-\mu\ell_{\max}})} \int_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} e^{-\mu\ell} d\ell \\
&= \frac{\mu}{2(e^{-\mu\ell_{\min}} - e^{-\mu\ell_{\max}})} \left[\frac{e^{-\mu\ell}}{-\mu} \right]_{|k-j|\Delta x}^{(|k-j|+1)\Delta x} \\
&= \frac{e^{-\mu|k-j|\Delta x} - e^{-\mu(|k-j|+1)\Delta x}}{2(e^{-\mu\ell_{\min}} - e^{-\mu\ell_{\max}})} \\
&= \frac{e^{-\mu|k-j|\Delta x} (1 - e^{-\mu\Delta x})}{2(e^{-\mu m_0 \Delta x} - e^{-\mu m_m \Delta x})}.
\end{aligned}$$

and $[A]_{k,j} = 0$ for $|k-j| < m_0$.

Appendix B

Area covered in the first-order model

Rather than consider the amount of new area covered, we determine the amount of overlap between the two areas, and subtract this from the sum of the two areas at the end.

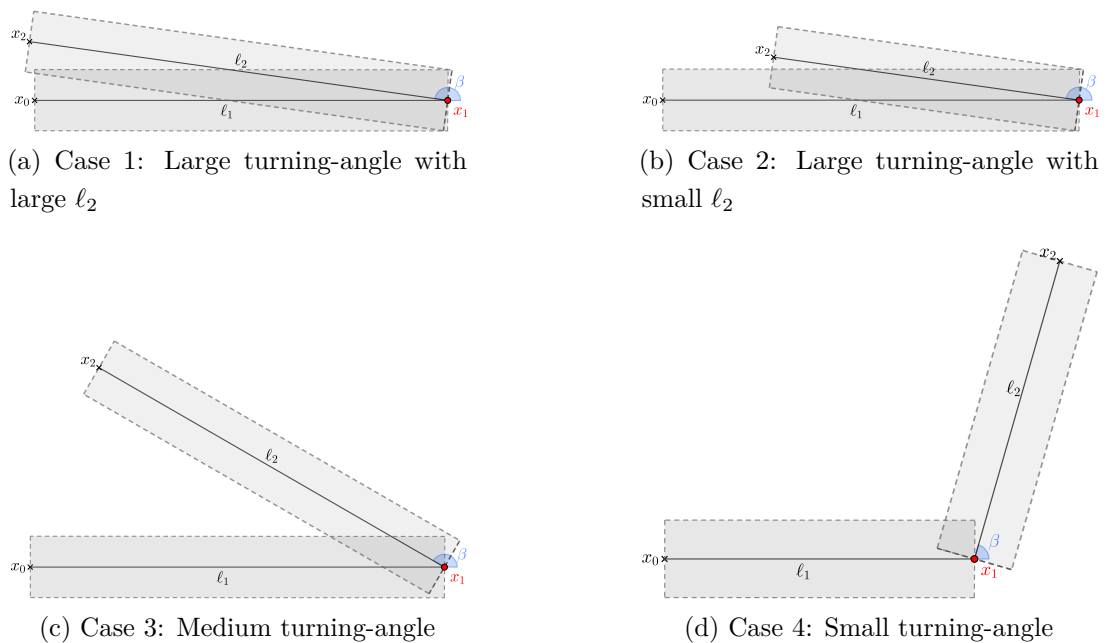


Figure B.0.1: The four cases that need to be considered to determine the overlapping area between two steps, with different turning-angles giving rise to different cases. When the turning-angle is large, which of case 1 or case 2 arise depends on the relative sizes of ℓ_1 and ℓ_2 . When ℓ_2 is sufficiently small, case 3 becomes case 2. Four more cases exist for turning-angles greater than π , although these are mirrors of the four cases above.

Case 1: Large turning-angle, large ℓ_2

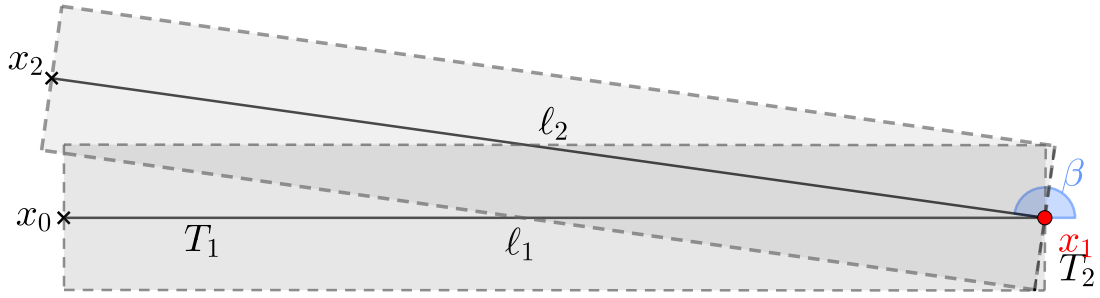


Figure B.0.2: Case 1: A large turning-angle and a large ℓ_2 . The bottom left triangle, T_1 , and the bottom-right triangle, T_2 , are labelled.

For large turning-angles and large ℓ_2 , a large amount of overlap occurs. We consider a turning-angle to be large when the bottom parallel line to ℓ_2 crosses the top parallel line of ℓ_1 . For this case, we require ℓ_2 to be large enough relative to ℓ_1 so that the ℓ_2 rectangle extends out of the end of the ℓ_1 rectangle, as it does in Figure B.0.2. The threshold length at which this occurs is when the horizontal component of ℓ_2 is larger than ℓ_1 , which implies that we require $\ell_1 < -\ell_2 \cos(\beta)$.

Note that there is not a clear transition between case 1 and case 2 at $\ell_1 = -\ell_2 \cos(\beta)$, but rather, there is a separate case entirely, since when ℓ_1 is slightly smaller than $-\ell_2 \cos(\beta)$, geometry will arise that does not fall into either of the two cases. In this transition, both of the left-most corners will be outside of the ℓ_1 rectangle while the line between the corners is within the rectangle, meaning a small triangle shaped area will be counted when it shouldn't. This can be seen in Figure B.0.3. However, since this occurs only in a very specific situation, and the effect is very small, we disregard this and simply assume that case 2 becomes case 1 once $\ell_1 < -\ell_2 \cos(\beta)$.

For case 1, rather than find the area of overlap, it is easier to find the area of the ℓ_1 rectangle that does not have any overlap. This is the large right-angled triangle at the bottom, as well as the small right-angled triangle at the bottom right, which we denote T_1 and T_2 respectively. Note that the non-overlapping area isn't entirely made up of these two triangles since they do not perfectly meet the bottom of the ℓ_1 rectangle, as seen in Figure B.0.4, although we assume that they do to avoid extra complications at the cost of a very small amount of accuracy.

The height of the triangle T_2 is r_v and the angle at x is $\pi - \beta$, and so the base must be $r_v \tan(\pi - \beta)$. Thus we get

$$A_{T_2} = \frac{-r_v^2}{2} \tan(\beta).$$

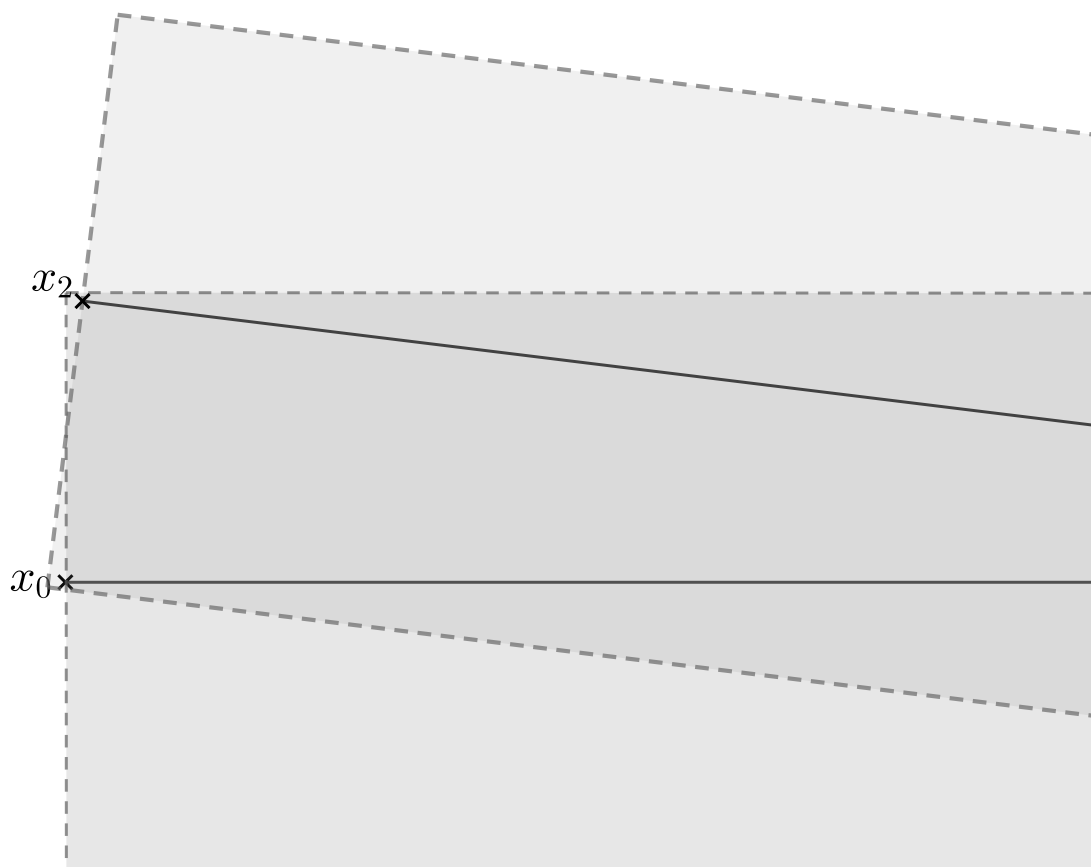


Figure B.0.3: There is not a clear threshold between case 1 and case 2, as can be seen in the above figure, which does not fall into either case. We assume that there is a clear threshold between the two cases, at the point when the bottom left corner of the ℓ_2 rectangle exits the ℓ_1 rectangle. The error caused by doing this will be small since the only difference is the small triangle on the left being counted as overlap when it should not be.

The base of triangle T_1 has length $\ell_1 - r_v \tan(\pi - \beta) = \ell_1 + r_v \tan(\beta)$, and the angle at the bottom-right of T_1 is $\pi - \beta$. The height of T_1 must therefore be

$$\tan(\pi - \beta) = h/(\ell_1 + r_v \tan(\beta)) \implies h = -\ell_1 \tan(\beta) - r_v \tan^2(\beta).$$

The area of triangle T_1 is

$$A_{T_1} = (\ell_1 + r_v \tan(\beta))(-\ell_1 \tan(\beta) - r_v \tan^2(\beta))/2 = \frac{-(\ell_1^2 \tan(\beta) + 2r_v \ell_1 \tan^2(\beta) + r_v^2 \tan^3(\beta))}{2},$$

and so the total area of overlap is

$$\begin{aligned} A_{\text{overlap}} &= 2r_v \ell_1 - A_{T_1} - A_{T_2} \\ &= \frac{\ell_1^2 \tan(\beta) + 2r_v \ell_1 (2 + \tan^2(\beta)) + r_v^2 \tan^3(\beta)}{2}. \end{aligned}$$

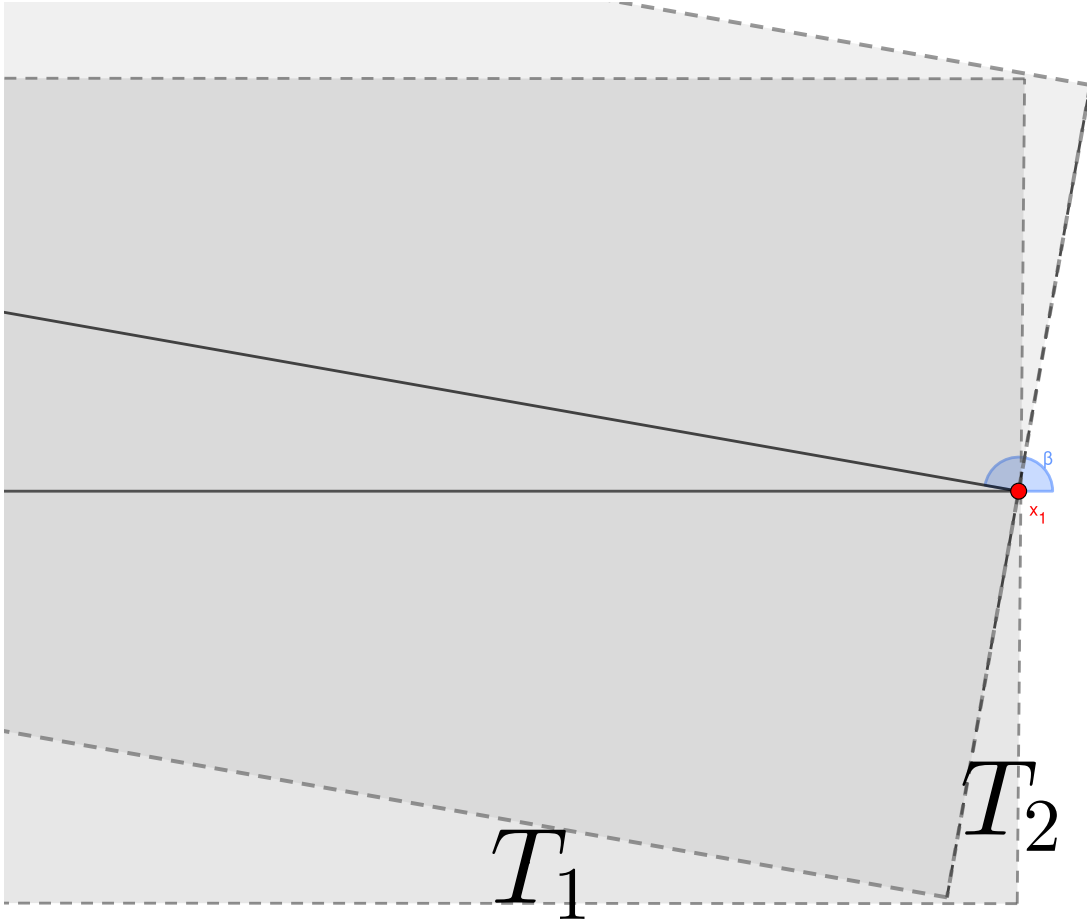


Figure B.0.4: The total unshaded area is not exactly $T_1 + T_2$, since the triangle T_1 does not meet the bottom line. The difference is so small that we assume they do meet, since it simplifies the calculations at the expense of very little error.

We can also use the height of triangle T_1 to determine the threshold angle at which the case of a large turning-angle becomes a medium turning-angle, and vice versa. When the height is greater than the width of the rectangle, $2r_v$, the turning-angle is considered medium. Thus, the threshold angle is

$$-\ell_1 \tan(\beta) - r_v = 2r_v \implies \beta^* = \pi + \arctan\left(\frac{-3r_v}{\ell_1}\right).$$

Therefore, case 1 corresponds to $\pi + \arctan\left(\frac{-3r_v}{\ell_1}\right) \leq \beta \leq \pi$, and $\ell_1 < -\ell_2 \cos(\beta)$.

For turning-angles larger than π , the same geometry will occur, but will be mirrored. The area for this will be

$$A_{\text{overlap}} = \frac{-\ell_1^2 \tan(\beta) + 2r_v \ell_1 (2 + \tan^2(\beta)) - r_v^2 \tan^3(\beta)}{2},$$

since we must flip the sign of $\tan(\beta)$ in this range. This occurs for $\pi \leq \beta \leq \pi - \arctan\left(\frac{-3r_v}{\ell_1}\right)$, and $\ell_1 < -\ell_2 \cos(\beta)$.

We can combine case 1 and its mirror, by taking absolute values of $\tan(\beta)$, giving

$$A_{\text{overlap}} = \frac{-\ell_1^2 |\tan(\beta)| + 2r_v \ell_1 (2 + \tan^2(\beta)) - r_v^2 |\tan^3(\beta)|}{2},$$

for $\pi + \arctan\left(\frac{-3r_v}{\ell_1}\right) \leq \beta \leq \pi - \arctan\left(\frac{-3r_v}{\ell_1}\right)$, and $\ell_1 < -\ell_2 \cos(\beta)$.

Case 2: Large turning-angle, small ℓ_2

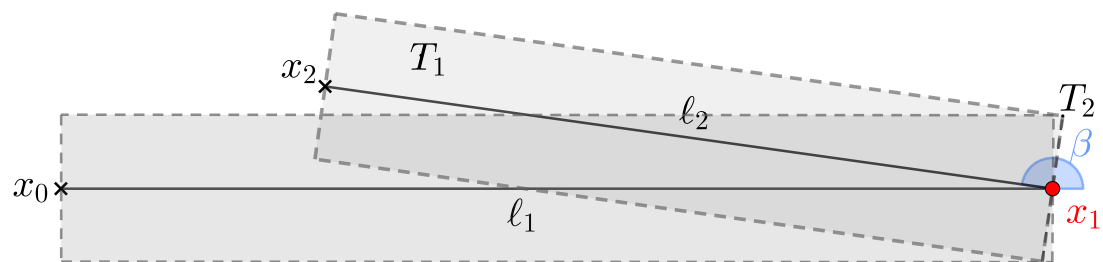


Figure B.0.5: Case 2: A large turning-angle and a relatively small ℓ_2 . The top triangle, T_1 , and the top-right triangle, T_2 , are labelled.

As discussed above, case 2 occurs when $\ell_1 \geq -\ell_2 \cos(\beta)$. We find the area of case 2 in a very similar way to case 1. This time we find the area of the ℓ_2 rectangle that does not have overlap, and this is then the newly searched area. This non-overlapping area is once again made up of two triangles, which we denote T_1 and T_2 . The area of the smaller triangle, T_2 is

$$A_{T_2} = \frac{-r_v^2}{2} \tan(\beta).$$

The base (top edge), of the larger triangle, T_1 , is $\ell_2 + r_v \tan(\beta)$, and so the height is

$$\tan(\pi - \beta) = h/(\ell_2 + r_v \tan(\beta)) \implies h = -\ell_2 \tan(\beta) - r_v \tan^2(\beta).$$

Therefore, the area of T_1 is

$$\begin{aligned} A_{T_1} &= (\ell_2 + r_v \tan(\beta))(-\ell_2 \tan(\beta) - r_v \tan^2(\beta))/2 \\ &= \frac{-(\ell_2^2 \tan(\beta) + 2r_v \ell_2 \tan^2(\beta) + r_v^2 \tan^3(\beta))}{2}. \end{aligned}$$

The total area of overlap is therefore

$$\begin{aligned} A_{\text{overlap}} &= 2r_v \ell_2 - A_{T_1} - A_{T_2} \\ &= \frac{\ell_2^2 \tan(\beta) + 2r_v \ell_2 (2 + \tan^2(\beta)) + r_v^2 \tan^3(\beta)}{2}. \end{aligned}$$

In this case, the threshold angle at which this case changes to case 3 will also be different. As can be seen in Figure B.0.5, when the height of the unshaded

triangle is greater than $2r_v$, case 2 becomes case 3. Since this unshaded section has hypotenuse ℓ_2 and bottom-right angle $\pi - \beta$, we get that the height is $\ell_2 \sin(\beta)$, and so the threshold angle will be

$$\beta^* = \pi - \arcsin\left(\frac{2r_v}{\ell_2}\right).$$

and so case 2 occurs when $\pi - \arcsin\left(\frac{2r_v}{\ell_2}\right) \leq \beta \leq \pi$ and $\ell_1 \geq -\ell_2 \cos(\beta)$.

For angles greater than π , we get a mirror of case 2, which occurs when $\pi \leq \beta \leq \pi + \arcsin\left(\frac{2r_v}{\ell_2}\right)$ and $\ell_1 \geq -\ell_2 \cos(\beta)$. The area for this case will be

$$A_{\text{overlap}} = \frac{-\ell_2^2 \tan(\beta) + 2r_v \ell_2 (2 + \tan^2(\beta)) - r_v^2 \tan^3(\beta)}{2},$$

where we have changed the sign of \tan from the original expression, as we did for case 1.

We can combine case 2 and its mirror, by taking absolute values of $\tan(\beta)$, giving

$$A_{\text{overlap}} = \frac{-\ell_2^2 |\tan(\beta)| + 2r_v \ell_2 (2 + \tan^2(\beta)) - r_v^2 |\tan^3(\beta)|}{2},$$

for $\pi - \arcsin\left(\frac{2r_v}{\ell_2}\right) \leq \beta \leq \pi + \arcsin\left(\frac{2r_v}{\ell_2}\right)$, and $\ell_2 \geq -\ell_1 \cos(\beta)$.

Case 3: Medium turning-angle

For medium-sized turning-angles, substantially less overlap may occur, depending on the size of the angle. Looking at Figure B.0.6, we see that the overlapping region is made up of three separate shapes, a right-angle triangle in the top right, a parallelogram, and a right-angle triangle at the bottom, and we denote these as T_1 , P , and T_2 , respectively. Both triangles have the same area, as we will now show. The angle of T_1 at x_1 is clearly $\beta - \pi/2$, and so is the angle of T_2 at x_1 . The length of the adjacent side to this angle for both triangles must be r_v . Then, the length of the opposite side will be

$$r_v \tan(\beta - \pi/2) = -r_v \cot(\beta).$$

Thus, we get

$$A_{T_1} = A_{T_2} = \frac{-r_v^2}{2} \cot(\beta).$$

For the parallelogram, the bottom and the right edge are the hypotenuse for the bottom and top triangles, respectively. Thus, the edges each have length

$$\sqrt{r_v^2 + r_v^2 \cot^2(\beta)} = r_v \sqrt{1 + \cot^2(\beta)} = r_v \csc(\beta).$$

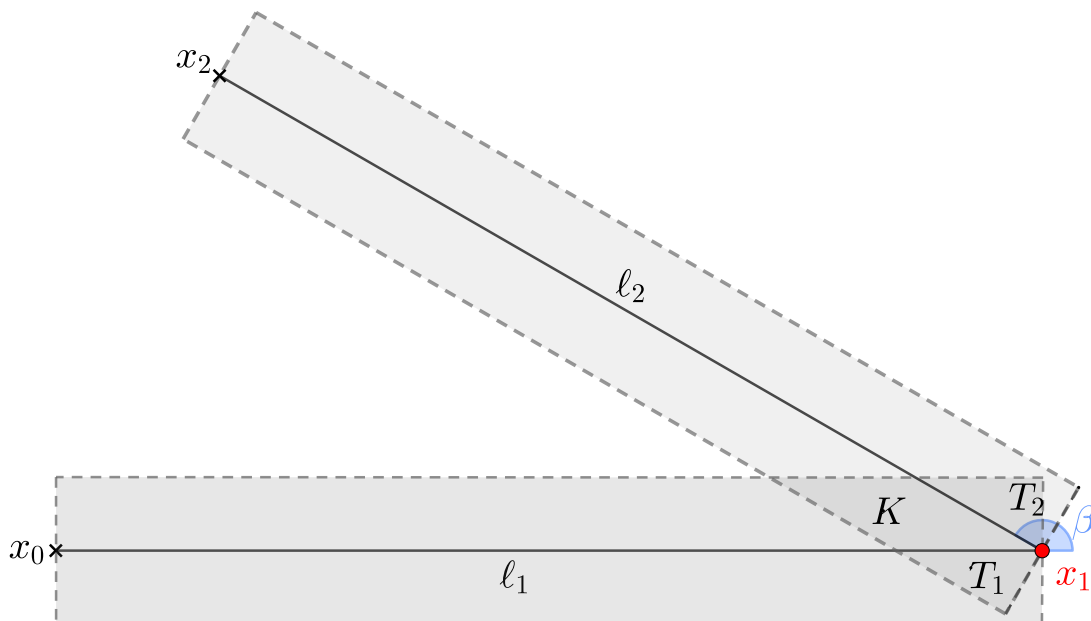


Figure B.0.6: Case 3: A medium turning-angle. The overlapping area is made up of a right-angled triangle at the top, T_1 , a kite, K , and a right-angled triangle at the bottom, T_2 .

The angle between these edges is $\pi - \beta$, and so the total area of the parallelogram is

$$A_k = r_v^2 \csc^2(\beta) \sin(\pi - \beta) = r_v^2 \csc(\beta).$$

Combining the three shapes give a total overlapping area of

$$A_{\text{overlap}} = A_{T_1} + A_k + A_{T_2} = r_v^2 \csc(\beta) - r_v^2 \cot(\beta) = r_v^2 (\csc(\beta) - \cot(\beta)).$$

We can further rearrange this to get

$$A_{\text{overlap}} = \frac{r_v^2 \sin(\beta)}{1 + \cos(\beta)}.$$

However, we need to keep in mind that the sign of the trigonometric functions may change throughout the range of β for case 3. To ensure this doesn't happen, we rewrite our expression as

$$A_{\text{overlap}} = \frac{r_v^2 |\sin(\beta)|}{1 - |\cos(\beta)|}.$$

This expression for the area is valid for the range $\pi/2 \leq \beta \leq \pi + \arctan\left(\frac{-3r_v}{\ell_2}\right)$.

This expression is also valid for the mirrored case when $\pi - \arctan\left(\frac{-3r_v}{\ell_2}\right) \leq \beta \leq 3\pi/2$.

Case 4: Small turning-angle

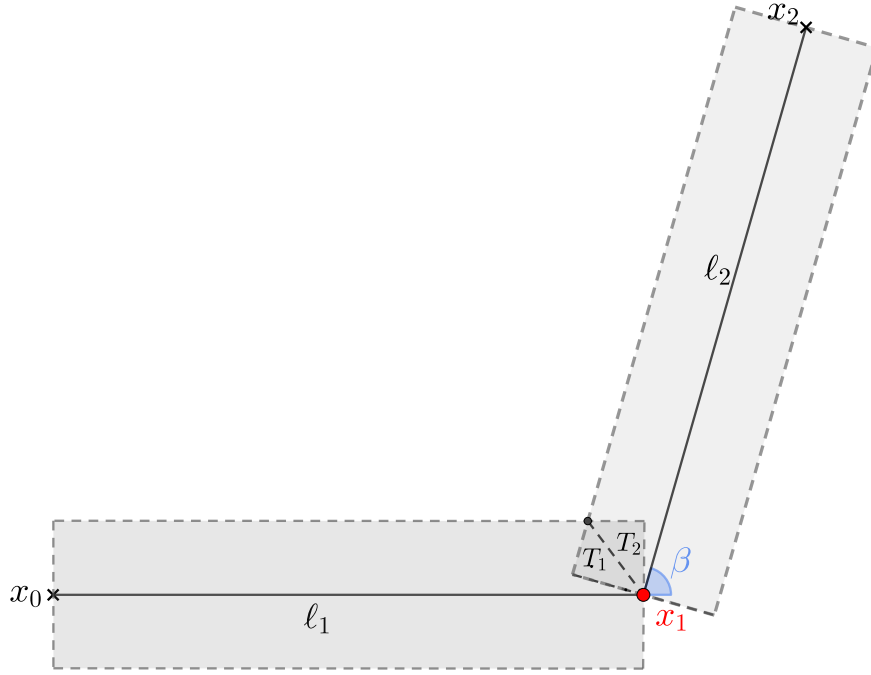


Figure B.0.7: Case 4: A small turning-angle. The overlapping area is made up of a quadrilateral, which we treat as two back-to-back right-angled triangles, T_1 and T_2 .

When the turning angle is small, $0 \leq \beta \leq \pi/2$, only a very small amount of overlap will occur, as can be seen in Figure B.0.7. Since we require that $\ell_{\min} \geq r_v$, the length of ℓ_1 and ℓ_2 will not effect the overlap area for this case. The overlapping area is comprised of a kite, which in this case can be thought of as two right-angled triangles back-to-back. It is easy to see that the top-left angle of the quadrilateral will be $\pi - \beta$, and so the top-left angle for each of the two triangles will be $\pi/2 - \beta/2$. The length of the side opposite to this angle, for both triangles, is r_v . Therefore, the length of the adjacent side for both triangles will be

$$\frac{r_v}{\tan(\pi/2 - \beta/2)} = \frac{r_v}{\cot(\beta/2)} = r_v \tan(\beta/2).$$

Then, the total overlapping area will be the sum of the area of both triangles,

$$A_{\text{overlap}} = r_v^2 \tan(\beta/2).$$

When making a small turn in the opposite direction, $-\pi/2 \leq \beta \leq 0$, we switch the sign of $\tan(\beta)$ to get

$$A_{\text{overlap}} = -r_v^2 \tan(\beta/2).$$

Combining the four cases

Combining each of the four cases, when $\ell_1 \geq -\ell_2 \cos(\beta)$, the total area of overlap is

$$A_{\text{overlap}} = \begin{cases} r_v^2 |\tan(\beta/2)| & \text{if } \beta \leq \pi/2, \text{ or } \beta \geq 3\pi/2, \\ \frac{r_v^2 |\sin(\beta)|}{1 - |\cos(\beta)|} & \text{if } \pi/2 \leq \beta \leq \pi - \arcsin\left(\frac{2r_v}{\ell_2}\right), \\ & \text{or } \pi + \arcsin\left(\frac{2r_v}{\ell_2}\right) \leq \beta \leq 3\pi/2, \\ \frac{1}{2} [-\ell_2^2 |\tan(\beta)| + 2r_v \ell_2 (2 + \tan^2(\beta))] & \text{if } \pi - \arcsin\left(\frac{2r_v}{\ell_2}\right) \leq \beta, \\ -r_v^2 |\tan^3(\beta)| & \text{and } \beta \leq \pi + \arcsin\left(\frac{2r_v}{\ell_2}\right), \end{cases}$$

and when $\ell_1 < -\ell_2 \cos(\beta)$, the overlap is

$$A_{\text{overlap}} = \begin{cases} r_v^2 |\tan(\beta/2)| & \text{if } \beta \leq \pi/2, \text{ or } \beta \geq 3\pi/2, \\ \frac{r_v^2 |\sin(\beta)|}{1 - |\cos(\beta)|} & \text{if } \pi/2 \leq \beta \leq \pi + \arctan\left(\frac{-3r_v}{\ell_1}\right), \\ & \text{or } \pi - \arctan\left(\frac{-3r_v}{\ell_1}\right) \leq \beta \leq 3\pi/2, \\ \frac{1}{2} [-\ell_1^2 |\tan(\beta)| + 2r_v \ell_1 (2 + \tan^2(\beta))] & \text{if } \pi + \arctan\left(\frac{-3r_v}{\ell_1}\right) \leq \beta, \\ -r_v^2 |\tan^3(\beta)| & \text{and } \beta \leq \pi - \arctan\left(\frac{-3r_v}{\ell_1}\right). \end{cases}$$

The angles at which the expressions change depend on the length of the steps, and the lengths at which the expressions change depend on the angle. To get around this, we simplify our expressions by replacing the condition $\ell_1 < -\ell_2 \cos(\beta)$ with $\ell_1 < \ell_2$, introducing another small amount of error. Now, note that when $\ell_1 \approx \ell_2$ and $\ell_2 \gg r_v$, then $\arctan\left(\frac{-3r_v}{\ell_1}\right) \approx -\arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right)$. Thus, the function for

the overlapping area becomes

$$A_{\text{overlap}} = \begin{cases} r_v^2 |\tan(\beta/2)| & \text{if } \beta \leq \pi/2 \text{ or } \beta \geq 3\pi/2, \\ \frac{r_v^2 |\sin(\beta)|}{1 - |\cos(\beta)|} & \text{if } \pi/2 \leq \beta \leq \pi - \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right), \\ & \text{or } \pi + \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right) \leq \beta \leq 3\pi/2 \\ \frac{1}{2} [-\min(\ell_1, \ell_2)^2 |\tan(\beta)| & \text{if } \pi - \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right) \leq \beta \\ + 2r_v \min(\ell_1, \ell_2)(2 + \tan^2(\beta)) & \text{and } \beta \leq \pi + \arcsin\left(\frac{2r_v}{\min(\ell_1, \ell_2)}\right). \\ -r_v^2 |\tan^3(\beta)|] & \end{cases}$$

Then, the total area covered by a step of length ℓ_2 , and with turning-angle β , with previous step-length ℓ_1 will be

$$A = 2\ell_2 r_v - A_{\text{overlap}}.$$

Bibliography

1. R. P. D. Atkinson, C. J. Rhodes, D. W. Macdonald & R. M. Anderson. Scale-free dynamics in the movement patterns of jackals. *Oikos* **98**, 134–140 (July 2002).
2. D. Austin, W. D. Bowen & J. I. McMillan. Intraspecific variation in movement patterns: modeling individual behaviour in a large marine predator. *Oikos* **105**, 15–30 (Apr. 2004).
3. F. Bartumeus, J. Catalan, U. L. Fulco, M. L. Lyra & G. M. Viswanathan. Optimizing the encounter rate in biological interactions: Lévy versus Brownian strategies. *Physical Review Letters* **88** (Feb. 2002).
4. F. Bartumeus, P. Fernández, M. G. E. da Luz, J. Catalan, R. V. Solé & S. A. Levin. Superdiffusion and encounter rates in diluted, low dimensional worlds. *The European Physical Journal Special Topics* **157**, 157–166 (Apr. 2008).
5. F. Bartumeus, F. Peters, S. Pueyo, C. Marrasé & J. Catalan. Helical Lévy walks: Adjusting searching statistics to resource availability in microzooplankton. *Proceedings of the National Academy of Sciences* **100**, 12771–12775 (Oct. 2003).
6. F. Bartumeus, E. P. Raposo, G. M. Viswanathan & M. G. E. da Luz. in *Lecture Notes in Mathematics* 3–32 (Springer Berlin Heidelberg, 2013).
7. W. J. Bell. Searching behavior patterns in insects. *Annual Review of Entomology* **35**, 447–467 (Jan. 1990).
8. S. Benhamou. How many animals really do the Lévy walk? *Ecology* **88**, 1962–1969 (Aug. 2007).
9. S. Benhamou. How many animals really do the Lévy walk? Reply. *Ecology* **89**, 2351–2352 (Aug. 2008).
10. S. Benhamou. Spatial memory and searching efficiency. *Animal Behaviour* **47**, 1423–1433 (June 1994).
11. S. Benhamou & P. Bovet. How animals use their environment: a new look at kinesis. *Animal Behaviour* **38**, 375–383 (Sept. 1989).

12. O. Bénichou, C. Loverdo, M. Moreau & R. Voituriez. A minimal model of intermittent search in dimension two. *Journal of Physics: Condensed Matter* **19**, 065141 (Jan. 2007).
13. O. Bénichou, M. Coppey, M. Moreau & R. Voituriez. Intermittent search strategies: When losing time becomes efficient. *Europhysics Letters (EPL)* **75**, 349–354 (July 2006).
14. O. Bénichou, M. Coppey, M. Moreau, P.-H. Suet & R. Voituriez. Optimal Search Strategies for Hidden Targets. *Physical Review Letters* **94** (May 2005).
15. O. Bénichou, C. Loverdo, M. Moreau & R. Voituriez. Two-dimensional intermittent search processes: an alternative to Lévy flight strategies. *Physical Review E* **74** (Aug. 2006).
16. S. Bertrand, A. Bertrand, R. Guevara-Carrasco & F. Gerlotto. Scale-invariant movements of fishermen: The same foraging strategy as natural predators. *Ecological Applications* **17**, 331–337 (Mar. 2007).
17. S. Bertrand, R. Joo & R. Fablet. Generalized Pareto for Pattern-Oriented Random Walk Modelling of Organisms' Movements. *PLOS ONE* **10** (ed F. Courchamp) e0132231 (July 2015).
18. C. T. Brown, L. S. Liebovitch & R. Glendon. Lévy Flights in Dobe Ju/'hoansi Foraging Patterns. *Human Ecology* **35**, 129–138 (Dec. 2006).
19. E. Browning, M. Bolton, E. Owen, A. Shoji, T. Guilford & R. Freeman. Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Methods in Ecology and Evolution* **9** (ed J. McPherson) 681–692 (Nov. 2017).
20. S. V. Buldyrev, M. Gitterman, S. Havlin, A. Kazakov, M. G. E. da Luz, E. P. Raposo, H. Stanley & G. M. Viswanathan. Properties of Lévy flights on an interval with absorbing boundaries. *Physica A: Statistical Mechanics and its Applications* **302**, 148–161 (Dec. 2001).
21. S. V. Buldyrev, S. Havlin, A. Y. Kazakov, M. G. E. da Luz, E. P. Raposo, H. E. Stanley & G. M. Viswanathan. Average time spent by Lévy flights and walks on an interval with absorbing boundaries. *Physical Review E* **64** (Sept. 2001).
22. E. A. Codling & M. J. Plank. Turn designation, sampling rate and the misidentification of power laws in movement path data using maximum likelihood estimates. *Theoretical Ecology* **4**, 397–406 (June 2010).
23. E. A. Codling, M. J. Plank & S. Benhamou. Random walk models in biology. *Journal of The Royal Society Interface* **5**, 813–834 (Aug. 2008).
24. B. J. Cole. Fractal time in animal behaviour: the movement activity of *Drosophila*. *Animal Behaviour* **50**, 1317–1324 (1995).

25. G. Dalby-Ball & A. Meats. Influence of the odour of fruit, yeast and cue-lure on the flight activity of the Queensland fruit fly, *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae). *Australian Journal of Entomology* **39**, 195–200 (July 2000).
26. M. D. Donsker. Justification and Extension of Doob's Heuristic Approach to the Kolmogorov- Smirnov Theorems. *The Annals of Mathematical Statistics* **23**, 277–281 (June 1952).
27. K. B. Døving, M. Mårstøl, J. R. Andersen & J. A. Knutsen. Experimental evidence of chemokinesis in newly hatched cod larvae (*Gadus morhua* L.) *Marine Biology* **120**, 351–358 (1994).
28. N. J. Dunford & J. T. Schwartz. *Linear Operators, Part 1: General Theory (Pure and Applied Mathematics, Vol. 7)* ISBN: 9780470226056 (John Wiley & Sons Inc, 1958).
29. D. B. Dusenbery. Spatial sensing of stimulus gradients can be superior to temporal sensing for free-swimming bacteria. *Biophysical Journal* **74**, 2272–2277 (May 1998).
30. A. M. Edwards. Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals. *Ecology* **92**, 1247–1257 (June 2011).
31. A. M. Edwards. Using likelihood to test for Lévy flight search patterns and for general power-law distributions in nature. *Journal of Animal Ecology* **77**, 1212–1222 (Nov. 2008).
32. A. M. Edwards, R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley & G. M. Viswanathan. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**, 1044–1048 (Oct. 2007).
33. C. L. Faustino, L. R. da Silva, M. G. E. da Luz, E. P. Raposo & G. M. Viswanathan. Search dynamics at the edge of extinction: Anomalous diffusion as a critical survival state. *Europhysics Letters (EPL)* **77**, 30002 (Jan. 2007).
34. A. Ferran, M. Ettifouri, P. Clement & W. J. Bell. Sources of variability in the transition from extensive to intensive search in coccinellid predators (Homoptera: Coccinellidae). *Journal of Insect Behavior* **7**, 633–647 (Sept. 1994).
35. T. Geisel, J. Nierwetberg & A. Zacherl. Accelerated diffusion in Josephson junctions and related chaotic systems. *Physical Review Letters* **54**, 616–619 (Feb. 1985).
36. G. R. Grimmett & D. R. Stirzaker. *Probability and Random Processes* ISBN: 0198572239 (Oxford University Press, 2001).

37. D. G. Haskell. Experiments and a model examining learning in the area-restricted search behavior of ferrets (*Mustela putorius furo*). *Behavioral Ecology* **8**, 448–449 (1997).
38. P. Hellung-Larsen, V. Leick, N. Tommerup & D. Kronborg. Chemotaxis in *Tetrahymena*. *European Journal of Protistology* **25**, 229–233 (Mar. 1990).
39. T. T. Hills & F. R. Adler. Time's crooked arrow: optimal foraging and rate-biased time perception. *Animal Behaviour* **64**, 589–597 (Oct. 2002).
40. M. Hoyle & J. E. Cresswell. A search theory model of patch-to-patch forager movement with application to pollinator-mediated gene flow. *Journal of Theoretical Biology* **248**, 154–163 (Sept. 2007).
41. A. James, M. J. Plank & R. Brown. Optimizing the encounter rate in biological interactions: Ballistic versus Lévy versus Brownian strategies. *Physical Review E* **78** (Nov. 2008).
42. A. James, M. J. Plank & A. M. Edwards. Assessing Lévy walks as models of animal foraging. *Journal of The Royal Society Interface* **8**, 1233–1247 (June 2011).
43. I. D. Jonsen, R. A. Myers & M. C. James. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology Progress Series* **337**, 255–264 (May 2007).
44. R. Joo, S. Bertrand, J. Tam & R. Fablet. Hidden markov models: The best models for forager movements? *PLoS ONE* **8** (ed G. G. de Polavieja) e71246 (Aug. 2013).
45. R. H. G. Klaassen, B. A. Nolet & D. Bankert. Movement of foraging Tundra Swans explained by spatial pattern in cryptic food densities. *Ecology* **87**, 2244–2254 (Sept. 2006).
46. G. Latouche & V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling* (Society for Industrial and Applied Mathematics, Jan. 1999).
47. T. Latty & M. Beekman. Food quality affects search strategy in the acellular slime mould, *Physarum polycephalum*. *Behavioral Ecology* **20**, 1160–1167 (2009).
48. V. Leick & P. Hellung-Larsen. Chemosensory Responses in *Tetrahymena*: The Involvement of Peptides and other Signal Substances. *The Journal of Protozoology* **32**, 550–553 (Aug. 1985).
49. M. A. Lomholt, K. Tal, R. Metzler & K. Joseph. Lévy strategies in intermittent search processes are advantageous. *Proceedings of the National Academy of Sciences* **105**, 11055–11059 (July 2008).

50. P. Marchal, J.-J. Poos & F. Quirijns. Linkage between fishers' foraging, market and fish stocks density: Examples from some North Sea fisheries. *Fisheries Research* **83**, 33–43 (Jan. 2007).
51. A. Mårell, J. P. Ball & A. Hofgaard. Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology* **80**, 854–865 (May 2002).
52. P. Moore & J. Crimaldi. Odor landscapes and animal behavior: tracking odor plumes in different physical worlds. *Journal of Marine Systems* **49**, 55–64 (Aug. 2004).
53. J. M. Morales, D. T. Haydon, J. Frair, K. E. Holsinger & J. M. Fryxell. Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology* **85**, 2436–2445 (Sept. 2004).
54. G. Nevitt. Olfactory foraging by Antarctic procellariiform seabirds: life at high Reynolds numbers. *The Biological Bulletin* **198**, 245–253 (Apr. 2000).
55. B. A. Nolet & W. M. Mooij. Search paths of swans foraging on spatially autocorrelated tubers. *Journal of Animal Ecology* **71**, 451–462 (May 2002).
56. B. C. Nolting. *Random search models of foraging behavior: theory, simulation, and observation*. PhD thesis (University of Nebraska - Lincoln, 2013).
57. A. Parton & P. G. Blackwell. Bayesian inference for multistate 'step and turn' animal movement in continuous time. *Journal of Agricultural, Biological and Environmental Statistics* **22**, 373–392 (June 2017).
58. T. A. Patterson, A. Parton, R. Langrock, P. G. Blackwell, L. Thomas & R. King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis* **101**, 399–438 (July 2017).
59. M. H. Persons & G. W. Uetz. Foraging patch residence time decisions in wolf spiders: Is perceiving prey as important as eating prey? *Écoscience* **4**, 1–5 (Jan. 1997).
60. M. J. Plank & E. A. Codling. Sampling rate and misidentification of Lévy and non-Lévy movement paths. *Ecology* **90**, 3546–3553 (Dec. 2009).
61. M. J. Plank & A. James. Optimal foraging: Lévy pattern or process? *Journal of The Royal Society Interface* **5**, 1077–1086 (Sept. 2008).
62. G. Ramos-Fernández, J. L. Mateos, O. Miramontes, H. L. G. Cocho & B. Ayala-Orozco. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behavioral Ecology and Sociobiology* **55**, 223–230 (Jan. 2004).
63. E. P. Raposo, S. V. Buldyrev, M. G. E. da Luz, M. C. Santos, H. E. Stanley & G. M. Viswanathan. Dynamical Robustness of Lévy Search Strategies. *Physical Review Letters* **91** (Dec. 2003).

64. E. P. Raposo, S. V. Buldyrev, M. G. E. da Luz, G. M. Viswanathan & H. E. Stanley. Lévy flights and random searches. *Journal of Physics A: Mathematical and Theoretical* **42**, 434003 (Oct. 2009).
65. D. Raubenheimer & S. J. Simpson. Nutritional ecology and foraging theory. *Current Opinion in Insect Science* **27**, 38–45 (June 2018).
66. A. M. Reynolds. Adaptive Lévy walks can outperform composite Brownian walks in non-destructive random searching scenarios. *Physica A: Statistical Mechanics and its Applications* **388**, 561–564 (Mar. 2009).
67. A. M. Reynolds. How many animals really do the Lévy walk? Comment. *Ecology* **89**, 2347–2351 (Aug. 2008).
68. A. M. Reynolds. Lévy flight patterns are predicted to be an emergent property of a bumblebees' foraging strategy. *Behavioral Ecology and Sociobiology* **64**, 19–23 (June 2009).
69. A. M. Reynolds. On the intermittent behaviour of foraging animals. *Europhysics Letters (EPL)* **75**, 517–520 (Aug. 2006).
70. A. M. Reynolds. Scale-free animal movement patterns: Lévy walks outperform fractional Brownian motions and fractional Lévy motions in random search scenarios. *Journal of Physics A: Mathematical and Theoretical* **42**, 434006 (Oct. 2009).
71. A. M. Reynolds & M. A. Frye. Free-flight odor tracking in *Drosophila* is consistent with an optimal intermittent scale-free search. *PLoS ONE* **2** (ed S. Rands) e354 (Apr. 2007).
72. A. M. Reynolds, D. Reynolds, A. Smith, G. Svensson & C. Löfstedt. Appetitive flight patterns of male *Agrotis segetum* moths over landscape scales. *Journal of Theoretical Biology* **245**, 141–149 (Mar. 2007).
73. A. M. Reynolds, A. D. Smith, R. Menzel, U. Greggers, D. R. Reynolds & J. R. Riley. Displaced honey bees perform optimal scale-free search flights. *Ecology* **88**, 1955–1961 (Aug. 2007).
74. G. Richard, C. Guinet, J. Bonnel, N. Gasco & P. Tixier. Do commercial fisheries display optimal foraging? The case of longline fishers in competition with odontocetes. *Canadian Journal of Fisheries and Aquatic Sciences* **75**, 964–976 (June 2018).
75. E. Şahin. in *Swarm Robotics* 10–20 (Springer Berlin Heidelberg, 2005).
76. M. C. Santos, E. P. Raposo, G. M. Viswanathan & M. G. E. da Luz. Optimal random searches of revisitable targets: Crossover from superdiffusive to ballistic random walks. *Europhysics Letters (EPL)* **67**, 734–740 (Sept. 2004).
77. M. C. Santos, G. M. Viswanathan, E. P. Raposo & M. G. E. da Luz. Optimization of random searches on defective lattice networks. *Physical Review E* **77** (Apr. 2008).

78. M. C. Santos, G. M. Viswanathan, E. P. Raposo & M. G. E. da Luz. Optimization of random searches on regular lattices. *Physical Review E* **72** (Oct. 2005).
79. M. Shaw, T. Harwood, M. Wilkinson & L. Elliott. Assembling spatially explicit landscape models of pollen and spore dispersal by wind for risk assessment. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1705–1713 (July 2006).
80. D. W. Sims, E. J. Southall, N. E. Humphries, G. C. Hays, C. J. A. Bradshaw, J. W. Pitchford, A. James, M. Z. Ahmed, A. S. Brierley, M. A. Hindell, D. Morritt, M. K. Musyl, D. Righton, E. L. C. Shepard, V. J. Wearmouth, R. P. Wilson, M. J. Witt & J. D. Metcalfe. Scaling laws of marine predator search behaviour. *Nature* **451**, 1098–1102 (Feb. 2008).
81. M. R. Strand & S. B. Vinson. Behaviour response of the parasitoid *Cardiochiles nigriceps* to a kairomone. *Entomologia Experimentalis et Applicata* **31**, 308–315 (Mar. 1982).
82. T. Sugimoto, H. Murakami & R. Yamazaki. Foraging for patchily-distributed leaf-miners by the parasitoid, *Dapsilarthra rufiventris* (Hymenoptera: Braconidae). II. Stopping rule for host search. *Journal of Ethology* **5**, 95–103 (Dec. 1987).
83. N. Suzuki. On the convergence of Neumann series in Banach space. *Mathematische Annalen* **220**, 143–146 (June 1976).
84. R. Tyson, J. Wilson & W. Lane. Beyond diffusion: Modelling local and long-distance dispersal for organisms exhibiting intensive and extensive search modes. *Theoretical Population Biology* **79**, 70–81 (May 2011).
85. G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince & H. E. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature* **381**, 413–415 (May 1996).
86. G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo & H. E. Stanley. Optimizing the success of random searches. *Nature* **401**, 911–914 (Oct. 1999).
87. G. M. Viswanathan, E. P. Raposo, F. Bartumeus, J. Catalan & M. G. E. da Luz. Necessary criterion for distinguishing true superdiffusion from correlated random walk processes. *Physical Review E* **72** (July 2005).
88. G. M. Viswanathan, M. G. E. da Luz, E. P. Raposo & H. E. Stanley. *Physics of Foraging* (Cambridge University Press, 2009).
89. E. P. White, B. J. Enquist & J. L. Green. On estimating the exponent of power-law frequency distributions. *Ecology* **89**, 905–912 (Apr. 2008).