The Traveling Salesman Problem Under Dynamic Constraints

by

Aviv Adler

B.A., Princeton University (2014)
S.M., Massachusetts Institute of Technology (2017)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2023

Author .....
Department of Electrical Engineering and Computer Science
January 25, 2023

Certified by .....
Sertac Karaman
Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

The Traveling Salesman Problem Under Dynamic Constraints

by

Aviv Adler

Submitted to the Department of Electrical Engineering and Computer Science
on January 25, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The Traveling Salesman Problem (TSP) is a foundational problem in the fields of theoretical computer science and optimization in which an agent is tasked with visiting a set of n target locations (in any order) in the shortest amount of time, either on a graph or in a space. As this problem is well-known to be NP-hard, it is usually solved using heuristics or approximation algorithms. An important variant of the TSP is the Dynamic TSP (DTSP), in which the targets exist in a space in which the agent's trajectory must satisfy dynamic constraints (for instance, limited ability to accelerate). The DTSP arises naturally in many robotic motion planning problems, particularly in exploration, surveillance and reconnaissance, and is generally not amenable to the standard TSP approximation algorithms. An interesting and important question, known as the Dynamic Stochastic TSP (DSTSP), asks: if the target points are distributed randomly, how does the length of the shortest tour (either in expectation or with high probability) grow with the number n of targets? This problem has been studied for a variety of common vehicle models, as well as certain broader classes of dynamic control systems.

In this thesis, we present a novel proof that extends known DSTSP order-of-growth results to a wider variety of dynamic systems, in particular to manifold workspaces, as well as two novel algorithms which achieve a constant-factor approximation of the optimal tour with high probability. These new proofs and algorithms furthermore allow us to study not only the order-of-growth of the tour length but also, for the important subset of 'symmetric' dynamics, to give explicit constant factors and to tightly characterize the relationship between the dynamics, the target point distribution, and the optimal tour length. Finally, we extend these results to the non-stochastic adversarial case, in which the target points are chosen to maximize the length of the optimal tour.

Thesis Supervisor: Sertac Karaman
Title: Professor of Aeronautics and Astronautics

Acknowledgments

I would like to thank the following people for making my time at MIT an incredible period of my life:

My advisor, Professor Sertac Karaman. During my (rather lengthy!) time at MIT, Sertac has always treated me with great generosity and patience. It was additionally an amazing insight of Sertac's which kicked off this work: that a relatively overlooked applied probability paper from 2005 held the key to cracking a problem which had been studied extensively (both for specific cases and in general) for the past two decades, and arguably since at least 1959. His insight, wisdom, guidance, and support were all absolutely essential to bringing this work to fruition.

Professor Yury Polyanskiy, for our work together on quantization, for helping guide me through this thesis, and for the many ways he helped me out over the years. Professor Erik Demaine for his delightful 'Fun with Hardness Proofs' class, which introduced me to the world of complexity and led to many amazing projects with him and others. Professor Mardavij Roozbehani, for being on my committee and giving me great feedback that sharpened this work.

The admins in the EECS department, CSAIL, and LIDS, whose patience and help were invaluable in keeping me on track throughout my time at MIT.

My colleagues and friends from LIDS, the Theory of Computation group, and elsewhere, with whom I worked and played chess and games and solved puzzles and talked and who helped make my time at MIT some of the best in my life. These include David Miculescu, Igor Spasojevic, Flora Meng, Omer Tanovic, Chenyang Yuan, Oscar Mickelin, Jayson Lynch, Michael Coulombe, Gautam Kamath, Jerry Li, Shalev Ben-David, Quanquan Liu, Govind Ramnarayan, Saeed Mehraban, Mohammed Bavarian, Luke Schaefer, Daniel Grier, Nicole Wein, Josh Alman, Dylan Mckay, Enric Boix, Srinivasan Raghuraman, Siddhartha Jayanti, Alex Lombardi, Adam Sealfon, Will Leiserson, Michael Cohen, Jon Gjengset, Max Wolf, Ragesh Ramachandran, Lennart Baardman, Atalay Ileri, Akshay Agarwal, and many many others.

My old friends from high school: Jeremy Dolinko, Ilyas Khan, Bryan Ransil,

Wesley Scoble. Can you believe it's been almost twenty years?

Mama and Aba, for giving me everything, teaching me everything, and whose love and support fills my life with happiness.

And finally, ♡ Jennifer Tang ♡ who for five years has been my constant companion through it all.

Contents

List of Figures

Chapter 1

Introduction and Preliminaries

1.1 Introduction and Motivation

1.1.1 The Dynamic Stochastic Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is a classic problem in computer science and optimization, in which an agent is tasked with visiting a set of target locations, in any order, in the minimum amount of time. In its most common form, the target locations are modeled as a graph with weighted edges indicating the distance between their endpoints. In addition to having an astounding range of practical applications, it attracts a great deal of theoretical interest in many settings.

The TSP is well-known to be computationally challenging – specifically, finding the shortest tour through a given set of points is NP-complete [1], even for Euclidean paths through points in $\mathbb{R}^2$ [2]. Nevertheless, there are a number of successful algorithmic methods for this problem, including approximation algorithms and heuristics.

Although the TSP originated in operations research, it has found numerous applications in the context of robotics as well. Most notably, a number of robot motion planning and routing algorithms employ TSP algorithms at their core [3]. The applications of TSP in the robotics domain are far reaching, including persistent monitoring, surveillance, reconnaissance, exploration, among other important problems.

However, in most cases in robotics, the vehicles we want to visit the target points

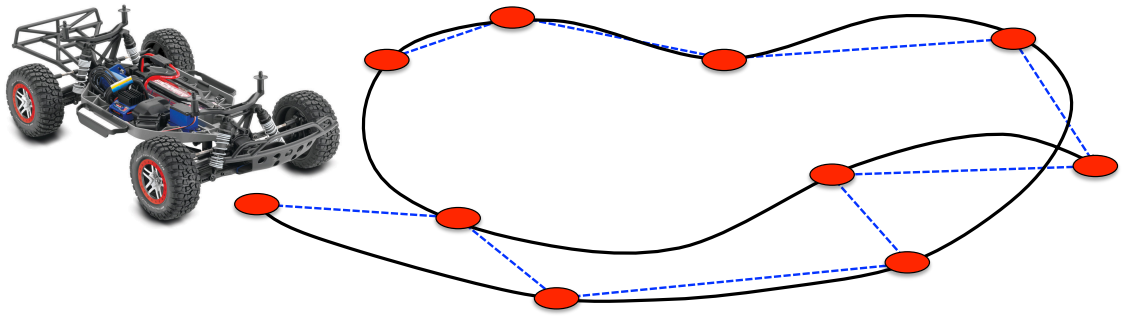

Figure 1-1: A Dynamic TSP instance. The dashed blue lines represent the shortest tour for an agent without dynamics, but the robot shown must obey dynamics and hence must follow a trajectory like the one pictured.

are subject to non-trivial differential constraints which have a substantial impact on the optimal tour. Also, the addition of dynamic constraints means there is no fixed notion of 'distance' between any given pair of target points, as the point-to-point travel time depends on the specific configurations (e.g. heading for a Dubins car) one wishes to have at the two target points in question. This means that the classic approximation algorithms do not apply to this setting, and new algorithms must be used to deal with the TSP in these instances [4]. Finally, understanding the impact of the dynamics on the length of the tour would allow a system designer to pick the best robot for the task at hand. We refer to this as the *Dynamic TSP* (DTSP).

An important variant of the TSP is the case in which the targets are independently and identically distributed (iid), in which case the goal is to understand the behavior of the shortest tour length as a random variable dependent on the targets, and in particular to understand how the length of the shortest tour grows as the number n of targets is increased. Since the length of the shortest tour is random, this is generally studied either in terms of expected value or with high-probability bounds. We refer to this as the *Stochastic TSP*; when combined with dynamic constraints, it is the *Dynamic Stochastic TSP* (DSTSP).

Remark 1. *While the dynamic constraints might describe an agent in many different settings, since our inspiration is the TSP for autonomous robotic vehicles we will generally refer to the agent as the 'vehicle' in this thesis.*

1.1.2 Previous Work

The importance of the Stochastic TSP has not gone unnoticed [5], inspiring extensive study (mainly under Euclidean distance, which we will call the Euclidean Stochastic TSP or ESTSP) from an algorithmic perspective, in particular on discovering algorithms which on average achieve high-quality approximations in polynomial time [6]. In addition, it was found to have connections to a variety of practical and theoretical problems, such as vehicle routing [7] and matching on the Euclidean plane [8].

The aysmptotic properties of the tour length of the Stochastic TSP as $n \to \infty$ was also extensively studied, both in the Euclidean case (notably by Beardwood et al. [9]) and for a variety of specific vehicles, culminating in results concerning the general class of *translation-invariant vehicles* on $\mathbb{R}^d$ by Itani [10]. Finally, concurrent to the DSTSP work, the asymptotic properties of the related *Dynamic Stochastic Orienteering* problem was studied using a powerful general method [11]; the Orienteering problem is in some sense the converse of the TSP, in which an agent (dynamically constrained or otherwise) must visit as many target points out of n as possible given a fixed limit on the time or distance allowed. This allows a new angle for tackling the DSTSP, as Orienteering and the TSP are closely related.

Euclidean Stochastic TSP The general asymptotic behavior of the Euclidean Stochastic TSP was characterized by Beardwood, Halton, and Hammersley in 1959 [9]:

Beardwood-Halton-Hammersley Theorem. For any (integer) dimension $d \geqslant 2$, there is a constant $\beta_d > 0$ such that for any continuous probability distribution with density function f on $\mathbb{R}^d$ with bounded support, $X_1, \ldots, X_n \overset{iid}{\sim} f$ implies:

$$\lim_{n\to\infty} \frac{\mathrm{TSP}(X_1, \ldots, X_n)}{n^{1-\frac{1}{d}}} = \beta_d \int f(x)^{1-\frac{1}{d}}\, dx \quad \text{almost surely.} \tag{1.1}$$

where $\mathrm{TSP}(X_1, \ldots, X_n)$ is the length of the optimal tour (under Euclidean distance) with targets $X_1, \ldots, X_n$.

Note that while the constants β_d can be empirically estimated to a high degree of accuracy, they are not known precisely.

In this theorem we see the general form of the asymptotic tour length which we will try to achieve: an order-of-growth (in this case $n^{1-\frac{1}{d}}$) attached to a constant factor with a clear dependence on the target point distribution's density f, which takes the form of an integral over $f(x)$ taken to a particular power.

Dynamic Stochastic TSP In the second half of the 2000's, the orders-of-growth of the DSTSP were found for a number of vehicles, notably the Dubins car, Reeds-Shepp car, differential-drive vehicles, and double integrators in 2 and 3 dimensions (which are commonly used to model quadcopter dynamics) [12, 13, 14, 15, 16]. The DSTSP for symmetric, translation-invariant dynamics in 2 dimensions was considered by Itani et al. [17], using a technique based on subadditive functionals, which was then expanded in Itani's Ph.D thesis [10] to include a wide class of translation-invariant dynamics in any Euclidean space $\mathbb{R}^d$. These results are summarized in Table 1.1; the general form $\Theta(n^{1-\frac{1}{\gamma}})$ of the order-of-growth is apparent, where the *small-time constraint factor* γ is a parameter dependent on the dynamics whose meaning will be discussed at length in Sections 1.2 and 2.1.

Vehicle	Reeds-Shepp Car Diff Drive [12]	Dubins Car [13, 15, 16]	Double Integrator [14]	Translation-invariant Vehicles [17, 10]
Space	$\mathbb{R}^2$	$\mathbb{R}^2$	$\mathbb{R}^2, \mathbb{R}^3$	$\mathbb{R}^2, \mathbb{R}^d$
Length	$\Theta(n^{\frac{2}{3}})$	$\Theta(n^{\frac{2}{3}})$	$\Theta(n^{\frac{2}{3}}), \Theta(n^{\frac{4}{5}})$	$\Theta(n^{1-\frac{1}{\gamma}})$

Table 1.1: Previous work on the DSTSP

These works contain important ideas which we use; in particular, the *Recursive Bead-Tiling* and *Recursive Cylinder-Covering* algorithms given by Savla et al. [14] (and a similar algorithm in Itani's thesis [10]) form the basis for our more general *Pass-and-Merge* algorithm discussed in Chapter 5. However, we note that unlike the Beardwood-Halton-Hammersley Theorem, these works do not consider what happens when the target point distribution f changes, instead generally assuming that the targets are uniformly distributed over the unit cube; while changing f generally does not affect the order of growth (given simple assumptions on f, such as being a con-

tinuous probability distribution over some bounded full-dimensional support)[1] , as seen in the Beardwood-Halton-Hammersley Theorem it has a serious effect on the constant factor attached to the length of the tour.

Dynamic Stochastic Orienteering The Orienteering problem, as described above, is a sort of converse of the TSP which asks: given an agent, a set of n targets, and some length or time bound $\lambda > 0$, how many target points can be visited by the agent in the given length or time, from any starting position? When the agent is dynamically constrained we call this *Dynamic Orienteering*, and when the targets are distributed iid at random we call this *Dynamic Stochastic Orienteering*. Upper bounds for Orienteering naturally correspond to lower bounds for the TSP in the following way: if at most $L(n, \lambda)$ targets can be reached with a trajectory of length or time λ, then to visit all n points requires (up to rounding) at least a trajectory of length $\lambda n / L(n, \lambda)$. In addition to this correspondence to the TSP, the Orienteering problem and in particular the Stochastic Orienteering and Dynamic Stochastic Orienteering problems have interesting additional applications to problems such as filament detection in image analysis [18].

This problem was studied at length by Arias-Castro, Donoho, Huo, and Tovey [11] using a powerful and general method which discretizes the space into appropriately-sized cells (dependent on the dynamics and the number of targets n) and analyzes the distribution of the target points into the cells based on the classic probability theory combination of the union bound and the Chernoff bound. Their method can not only handle a theoretically infinite range of different dynamically-constrained agents, but also yields a means to strongly characterize how the probability that their bound is violated goes to zero as the number of targets n goes to ∞.

However, their work assumes that an appropriate discretization is supplied externally, which is often highly nontrivial; for instance, in their work they give the appropriate discretizations for Euclidean Stochastic Orienteering and the Dynamic

[1]This is because any such f can be 'bounded above' by distributing (a constant factor) more targets on a cube containing its support, which can then be divided into a fixed number of unit cubes. Each cube can then be handled individually, with the extra time needed to travel between them not dependent on the number of targets n.

Stochastic Orienteering with a Dubins car agent (a simple and extremely versatile car model used in many applications [3]).

1.1.3 Contribution

We bring together the three main strands of work on the DSTSP and the related Dynamic Stochastic Orienteering problem. In doing so, we achieve the following using new and elementary methods:

- We extend the known DSTSP order-of-growth results (which depend on the *small-time constraint factor* γ, see Definition 4) to a broader class of dynamics, specifically non-translation-invariant dynamics, which includes dynamics on manifolds.

- We find a clear lower bound which not only captures the order-of-growth of the optimal tour length but also captures an interesting interaction between the target point density function f and a newly-defined parameter which we refer to as the *agility function* g (Definition 5), which can also be viewed as an extension of the Beardwood-Halton-Hammersley Theorem's characterization of the dependence of the ESTSP tour length on f. We do this by converting the problem to a novel version of the Dynamic Stochastic Orienteering problem (which we refer to as *Cost-Balanced Orienteering* and which has a tighter correspondence to the DSTSP) and applying the general method in [11], showing an 'automatic' way of using the geometry of the dynamics at small scales to produce the necessary discretization.

- For the upper bound, we consider two (well-known) different types of dynamic constraints: symmetric and nonsymmetric, distinguished by whether the agent can reverse their trajectory without loss of time. For each case, we design a DSTSP algorithm by constructing a recursive tiling (which we refer to as *Hierarchical Cell Structures*, see Sections 2.3 and 2.4) and using it to convert the DSTSP analysis problem into a discrete probability problem (one for each case) which we then solve.

  - Symmetric case: our upper bound matches the lower bound up to a factor constant in both number of targets n and target probability density function f (but allowed to depend on the dynamics), thus describing how the tour length depends on the number of targets n and the target point density function f as $n \to \infty$.
  - Nonsymmetric case: our upper bound matches the lower bound up to a factor constant in the number of targets n (but allowed to depend on the dynamics and the target density function f) thus describing how the tour length depends on the number of targets n as $n \to \infty$.

  Both algorithms are probabilistic approximation algorithms: given n targets iid distributed according to f, with very high probability the given algorithm finds a tour within a specified constant multiplicative factor of the optimal tour, where the constant can depend on the dynamics in both cases but can only depend on f in the nonsymmetric case (see Appendix A.2 for more).

- We show how the probability that our upper and lower bounds are violated go to zero as n goes to infinity. In particular, we define a new and stronger notion of 'high probability', which we call *very high probability* (Definition 7) and show that our bounds hold with very high probability.

- We extend our DSTSP lower bound and the symmetric dynamics upper bound to the *adversarial* case, where the targets are not distributed randomly but by an adversary whose goal is to maximize the length of the shortest tour. In doing so we obtain strong deterministic bounds on the longest possible length for the shortest tour through n targets when the dynamics are symmetric.

We also clearly articulate a set of assumptions from which the results are derived; to apply our results, these assumptions can either be shown on their own (e.g. when the geometry of the reachable sets (Definition 3) is clearly known) or by considering deeper geometric properties of the dynamics. All of our work translates cleanly to upper and lower bounds for the Dynamic Stochastic Orienteering problem.

1.1.4 Thesis organization

The rest of this work is organized as follows.

- In the remainder of Chapter 1, we introduce the basic preliminaries needed to understand the results (Section 1.2) and then state our main results (Section 1.3). Note that the concepts introduced in Section 1.2 are sometimes not rigorously defined there, and instead meant only to give the means to understand Section 1.3; the full, rigorous treatment of these concepts is given in Chapter 2.

- In Chapter 2 we state formally all the assumptions needed; we also show how many of these assumptions follow from the underlying geometry of dynamic control systems, as long as they have a given (very common and general) form and particular result (the Ball-Box Theorem) holds for the system.

- In Chapter 3 we show our DSTSP lower bounds, which we achieve via an Orienteering upper bound using the method developed by Arias-Castro et al. [11].

- In Chapter 4 we give our TSP algorithm for symmetric dynamics and derive from it a very-high-probability upper bound on the shortest tour length which is within a constant factor (with regard to the number of targets n and the target density function f) of the lower bound.

- In Chapter 5 we give our TSP algorithm for nonsymmetric dynamics and derive from it a very-high-probability upper bound on the shortest tour length which achieves the optimal order of growth (i.e. is within a constant factor of the lower bound with regard to the number of targets n).

- In Chapter 6 we discuss the extension of our results to the case of worst-case (nonrandom) target points which are placed to maximize the length of the shortest tour.

- In Chapter 7 we conclude and discuss possible future directions of work.

1.1.5 Notation and color-coding

We make use of a few standard notational conventions:

- Calligraphic font is typically used for spaces and sets; e.g. $\mathcal{X}$ for the workspace (the space that the vehicle moves and visits target points in).
- Random values are typically represented by upper-case letters, using the same letter as their non-random counterparts in lower case; e.g. $x \in \mathcal{X}$ is a point in the workspace, while X would represent a random point in the workspace.
- The time derivative of a process $a(t)$ (say, a vehicle's position) on a manifold $\mathcal{A}$ is given by $\dot{a}(t)$, which represents a vector tangent to $\mathcal{A}$ at a; we do not use a' to indicate derivatives. Instead, a' will often be used to denote another point in the same space, e.g. 'for any a and a' in $\mathcal{A}$, the following condition holds'.

For ease-of-reading we also color-code our results:

- Our main theorems are presented in dark blue.
- Supporting (but still crucial) propositions are presented in light blue.
- Corollaries are presented in green.
- Definitions are presented in red.
- Assumptions are presented in yellow.
- Meta-Assumptions (geometric assumptions from which other assumptions follow) are presented in orange.
- Known results we use are presented in gray (for example, see the Beardwood-Halton-Hammersley Theorem above).
- Conjectures are presented in brown.

To avoid clutter, lemmas are not colored.

1.2 Preliminaries

In this work we study the Dynamic Stochastic TSP, in which a vehicle controlled by a system of dynamic constraints on a manifold $\mathcal{X}$, which we call the *workspace*, must visit a set of independently and identically-distributed (iid) target points $X_1, \ldots, X_n \in \mathcal{X}$, in any order, in as short a time as possible. Common examples of such vehicles include the Dubins Car, the Differential-Drive Vehicle, and the Double Integrator (either in $\mathbb{R}^2$ or $\mathbb{R}^3$, commonly used as a model of quadcopter motion).

1.2.1 The dynamic constraints

Although we are primarily concerned about the vehicles's trajectory through the workspace $\mathcal{X}$ (which determines if and when each target point is visited), the dynamic constraints mean the state of the vehicle at any given time is more naturally represented as a point $\boldsymbol{q}$ in the *configuration space* (or *phase space*) $\mathcal{Q}$. Its position in $\mathcal{X}$ is then a function of its position in $\mathcal{Q}$, denoted as $[\boldsymbol{q}]_{\mathcal{X}} = x \in \mathcal{X}$; [1] we denote the preimage of $x \in \mathcal{X}$ as $[x]^{\mathcal{Q}} = \{\boldsymbol{q} \in \mathcal{Q} : [\boldsymbol{q}]_{\mathcal{X}} = x\}$. We assume this function works locally as a projection from a $\dim(\mathcal{Q})$-dimensional Euclidean space onto a $\dim(\mathcal{X})$-dimensional subspace, i.e. for any $\boldsymbol{q}, x$ such that $[\boldsymbol{q}]_{\mathcal{X}} = x$, there are some neighborhoods around $\boldsymbol{q}$ and x and coordinate maps on these neighborhoods so that $[\cdot]_{\mathcal{X}}$ behaves linearly with respect to these maps.

For convenience we assume this mapping is well-behaved:

Assumption 1. $[\cdot]_{\mathcal{X}}$ is smooth.

This doesn't need to hold completely everywhere for our results to still hold (see Section 2.5) but we assume it for what follows.

Given a position $\boldsymbol{q} \in \mathcal{Q}$ and a control input $\boldsymbol{u}$ from some control set $\mathcal{U}$, the vehicle's position evolves according to a *control law* $\dot{\boldsymbol{q}} = h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u})$, where h^{cont} takes a position in $\mathcal{Q}$ and a control and returns a tangent vector indicating the direction in

[1] Typically $\mathcal{Q} = \mathcal{X} \times \mathcal{C}$ for some $\mathcal{C}$, and $[\cdot]_{\mathcal{X}}$ is the projection function. It is possible that $\mathcal{Q} = \mathcal{X}$, but this is a special case.

which the vehicle moves [3]. We call a trajectory $\pi : [0, T] \to \mathcal{Q}$ *valid* if there is some control function $\boldsymbol{u} : [0, T] \to \mathcal{U}$ such that $\dot{\pi}(t) = h^{\text{cont}}(\pi(t), \boldsymbol{u}(t))$ for all $t \in [0, T]$.

We now introduce some notation which we will use for this work. We denote the set of all valid trajectories as Π (we will also refer to the dynamics by the set Π of valid trajectories it induces). Interpreting t as time, T is then the amount of time it takes to execute trajectory $\pi \in \Pi$ via the control function $\boldsymbol{u}$, which we refer to as the *length* of the trajectory π and denote as $\ell(\pi)$, so that $\pi : [0, \ell(\pi)] \to \mathcal{Q}$. Since the targets are points in $\mathcal{X}$ but the dynamics are specified over $\mathcal{Q}$, we will be interested in the projection of trajectory $\pi \in \Pi$ onto $\mathcal{X}$. We will denote:

$$\bar{\pi}(t) = [\pi(t)]_{\mathcal{X}} \tag{1.2}$$

In general $\bar{\cdot}$ will denote the workspace projection of a set or function. We will also slightly abuse notation by writing "$x \in \bar{\pi}$" to mean "$\exists\, t$ s.t. $x = \bar{\pi}(t)$" i.e. that π visits x at some point, and similarly "$\boldsymbol{q} \in \pi$" if there exists t such that $\boldsymbol{q} = \pi(t)$.

1.2.2 Symmetric dynamics

A special class of dynamic constraints are those which allow the vehicle to reverse direction and 'backtrack' with no loss of time. This property, which we call *symmetry*, allows the system to use a simplified algorithm as compared to vehicles that don't have it. Formally:

Definition 1. We say that dynamics Π defined by control law h^{cont} and control set $\mathcal{U}$ are *symmetric* if for every $\boldsymbol{q} \in \mathcal{Q}$ and $\boldsymbol{u} \in \mathcal{U}$:

$$h^{\text{cont}}(\boldsymbol{q}, -\boldsymbol{u}) = -h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u}) \tag{1.3}$$

$$\text{and } \boldsymbol{u} \in \mathcal{U} \iff -\boldsymbol{u} \in \mathcal{U} \tag{1.4}$$

i.e. any possible motion can also be reversed by changing the control.

This then means that for any valid trajectory $\pi : [0, \ell(\pi)] \to \mathcal{Q}$, the reverse trajectory $\overleftarrow{\pi} : [0, \ell(\pi)] \to \mathcal{Q}$ where

$$\overleftarrow{\pi}(t) := \pi(\ell(\pi) - t) \tag{1.5}$$

is also valid. This allows the vehicle to maneuver indefinitely in arbitrarily small regions without having to leave and come back, and allowing a simplified algorithm (with more precise bounds) when it holds.

1.2.3 Control-affine dynamics

A very powerful, general, and useful class of control systems are those which are governed by a set of vector fields $h_0^{\text{cont}}, h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$ over $\mathcal{Q}$: given a location $\boldsymbol{q} \in \mathcal{Q}$, the vehicle's instantaneous motion is a linear combination of the tangent vectors $h_0^{\text{cont}}(\boldsymbol{q}), h_1^{\text{cont}}(\boldsymbol{q}), \ldots, h_m^{\text{cont}}(\boldsymbol{q})$ produced by these fields at $\boldsymbol{q}$, with the control inputs being the coefficients for the various vector fields; the vector field h_0^{cont} is special and always applies with a coefficient of 1 (called the drift), while the rest can be scaled up or down by the controls. Formally, we define:

Definition 2. A *control-affine system* (also called *affine-in-control*) is a system of dynamic constrains in which the control law takes the form:

$$h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u}) = h_0^{\text{cont}}(\boldsymbol{q}) + \sum_{i=1}^{m} h_i^{\text{cont}}(\boldsymbol{q}) u_i \tag{1.6}$$

where $h_0^{\text{cont}}, h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$ are smooth vector fields on $\mathcal{Q}$, and the control is some $\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathcal{U} \subseteq \mathbb{R}^m$.

We assume that the following apply to the control set $\mathcal{U}$:

- $\mathbf{0}$ is in the interior of $\mathcal{U}$;
- $\mathcal{U}$ is bounded, convex, and closed (therefore compact);
- $\mathcal{U}$ is radially symmetric, i.e. $\boldsymbol{u} \in \mathcal{U} \iff -\boldsymbol{u} \in \mathcal{U}$.

The uncontrolled vector field h_0^{cont} is called the *drift*; if $h_0^{\text{cont}}(\boldsymbol{q}) = \mathbf{0}$ (the trivial tangent vector at $\boldsymbol{q}$) for all $\boldsymbol{q} \in \mathcal{Q}$, the vehicles is symmetric, which for control-affine dynamics is also called *driftless*.

In some control-affine systems, $h_0^{\text{cont}}, \ldots, h_m^{\text{cont}}$ span the tangent space at $\boldsymbol{q}$, allowing linear combinations to move in any direction (subject to the drift term h_0^{cont}); however, the vehicles we are interested in typically do not have this property, making them *nonholonomic* [19]. In this case, the direction of movement is restricted to the subspace of the tangent plane spanned by the vector fields h_i^{cont}.

Nevertheless, in symmetric control-affine systems, it is often possible to produce motion in a direction not directly allowed by making small forward and backward motions using commutators. Formally, the Lie algebra of the vector fields at $\boldsymbol{q}$ yields a set of vectors in which small motions can be made. A control-affine system is called *controllable* at $\boldsymbol{q}$ if this Lie algebra spans the tangent space. In this case, for any $\varepsilon > 0$, the ε-ball in the dynamics around $\boldsymbol{q}$ will contain $\boldsymbol{q}$ in its interior.

Example 1. *Any nonsymmetric control-affine system $h_0^{\text{cont}}, h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$ has a symmetric analogue. For example, the Dubins car in $\mathbb{R}^2$, whose dynamics can be written (for $\boldsymbol{q} = (x, y, \theta)$ denoting (x, y) position in $\mathbb{R}^2$ and heading θ) as*

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u_1 \tag{1.7}$$

where $u_1 \in [-1, 1]$; its symmetric analogue is the Reeds-Shepp car*, whose dynamics are*

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix} u_1 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u_2 \,. \tag{1.8}$$

Note that to produce the symmetric analogue, we don't delete the drift vector field but rather add a control to it.

1.2.4 Reachable sets and related parameters

We first define the distance function induced by dynamics Π:

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') = \inf_{\pi \in \Pi} (\ell(\pi) : \pi(0) = \boldsymbol{q} \text{ and } \pi(\ell(\pi)) = \boldsymbol{q}') \tag{1.9}$$

Note that the minimum time required to go from $\boldsymbol{q}$ through $\boldsymbol{q}'$ and end at $\boldsymbol{q}''$ for any $\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'' \in \mathcal{Q}$ is $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') + d_\Pi(\boldsymbol{q}', \boldsymbol{q}'')$. Furthermore, making a useful equivalent definition of the distance between $x, x' \in \mathcal{X}$ poses difficulties as the time required to go $x \to x' \to x''$ generally does not add in the same way it does for $\boldsymbol{q} \to \boldsymbol{q}' \to \boldsymbol{q}''$ since the shortest valid path $x \to x'$ might require a much different configuration at x' from the shortest valid path $x' \to x''$.

When Π is symmetric, d_Π, but when Π is not symmetric it is a *quasimetric* since $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \neq d_\Pi(\boldsymbol{q}', \boldsymbol{q})$ in general.

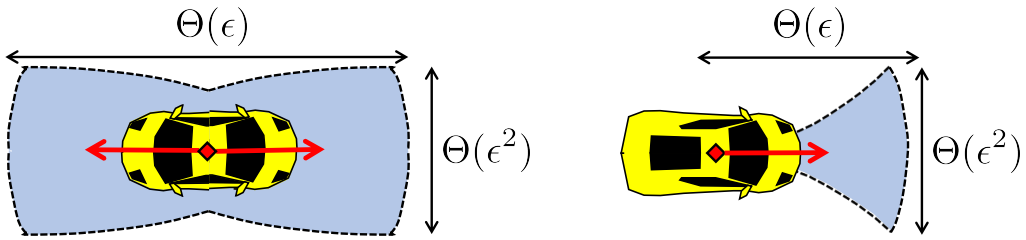

Figure 1-2: Illustration of a symmetric vehicle (left, Reeds-Shepp car) and a non-symmetric vehicle (right, Dubins car) with similar dynamics; both cannot follow trajectories whose curvature exceeds a certain bound. Workspace ε-reachable sets are shown for each, note that the Reeds-Shepp car workspace reachable set is not simply two copies of the Dubins car workspace reachable set. The volumes of the workspace reachable sets are $\Theta(\varepsilon^3)$, indicating that $\gamma = 3$ for the vehicles involved.

Definition 3. Given a system Π, $\boldsymbol{q} \in \mathcal{Q}$ and $\varepsilon > 0$, the *ε-reachable set* and the *workspace ε-reachable set* from $\boldsymbol{q}$ are defined as

$$R_\varepsilon(\boldsymbol{q}) := \{\boldsymbol{q}' \in \mathcal{Q} : d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon\} \quad \text{and} \quad \bar{R}_\varepsilon(\boldsymbol{q}) := [R_\varepsilon(\boldsymbol{q})]_\mathcal{X} . \tag{1.10}$$

We refer to $\boldsymbol{q}$ as the *anchor* of $R_\varepsilon(\boldsymbol{q})$ and $\bar{R}_\varepsilon(\boldsymbol{q})$.

A key aspect of our results is the volume of these reachable sets, particularly the workspace reachable sets, which determines in a sense how maneuverable the vehicle is. In general, this volume will scale polynomially with ε, i.e. $\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})) \propto \varepsilon^\gamma$ for some appropriate γ, which we call (following [10]) the *small-time constraint factor*, and which will ultimately determine how the TSP tour length scales as the number n of targets grows to ∞. Note that we are assuming that γ is constant over the space.

However, in order to obtain more precise bounds (particularly for the case of symmetric vehicles) we are also interested in the constant factor attached to ε^γ, which can vary depending on the configuration $\boldsymbol{q}$ in question. We call this the *configuration agility function* $g : \mathcal{Q} \to \mathbb{R}_{>0}$ and broadly speaking for small $\varepsilon > 0$ we have

$$\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})) \approx g(\boldsymbol{q})\varepsilon^\gamma \tag{1.11}$$

Formally we can define these as follows:

Definition 4. The *small-time constraint factor* of Π is γ such that at all $\boldsymbol{q}$

$$\gamma := \lim_{\varepsilon \to 0} \frac{\log(\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})))}{\log(\varepsilon)} \tag{1.12}$$

and the *agility function* $g : \mathcal{Q} \to \mathbb{R}_{>0}$ is

$$g(\boldsymbol{q}) := \lim_{\varepsilon \to 0} \frac{\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma}\,. \tag{1.13}$$

Note that the configuration agility function takes inputs from $\mathcal{Q}$. Thus, we need to define the *workspace agility function* which roughly measures the maximum possible agility the vehicle can have when at a configuration projecting to $x \in \mathcal{X}$. This is important because if the vehicle can be in many different configurations corresponding to $x \in \mathcal{X}$, it will generally be best to use the configuration maximizing the agility. For simplicity we will also denote this by g; whether g refers to the configuration agility function or the workspace agility function can be determined by whether its input is

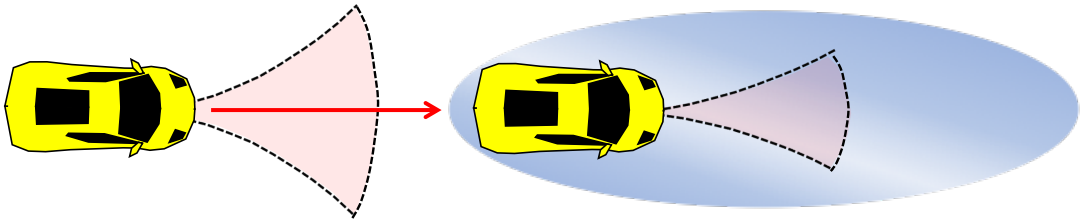

Figure 1-3: Illustration of a vehicle with nonconstant agility function. A Dubins car drives onto a patch of ice, reducing its turn rate and thus reducing its agility (as seen by the smaller workspace reachable set).

in $\mathcal{Q}$ or $\mathcal{X}$. We will use the term *agility function* to refer to either the configuration or workspace agility function depending on context.

Definition 5. The *workspace agility function* is $g : \mathcal{X} \to \mathbb{R}_{>0}$ defined as

$$g(x) := \sup_{\boldsymbol{q} \in [x]^{\mathcal{Q}}} g(\boldsymbol{q}) \tag{1.14}$$

1.2.5 The Dynamic TSP and Dynamic Stochastic TSP

The Dynamic TSP (DTSP) is then the following problem:

Definition 6. Given control system Π on $\mathcal{Q}$, and given n targets $x_1, \ldots, x_n \in \mathcal{X}$, the *TSP trajectory* (or DTSP trajector) is the shortest valid path through all the targets, denoted by:

$$\mathrm{TSP}_\Pi(\{x_i\}) := \mathrm{TSP}_\Pi(x_1, \ldots, x_n) := \inf_{\pi \in \Pi} (\ell(\pi) : x_i \in \bar{\pi} \text{ for all } i)\,. \tag{1.15}$$

Note that we may assume WLOG that any DTSP trajectory through $x_1, \ldots, x_n$ starts and ends in the set $\{x_1, \ldots, x_n\}$, since additional movement before visiting the first target, or after visiting the last target, is wasteful.

Remark 2. *Some variants of the TSP require the agent to visit all the target points and then return to their original location, thus forming a loop. While Definition 6 does not require the vehicle to return to its starting location or starting configuration,*

we note that our results hold equally for the case of loops. This is because we will make the assumption that the targets cannot be arbitrarily far apart (Assumption 4) but must be confined within bounded space; thus, requiring a loop cannot make the shortest tour shorter, and can add only at most a constant extra time to the tour.

The Dynamic Stochastic TSP (DSTSP) is the variant of the above problem in which the targets $X_1, \ldots, X_n$ are randomly chosen; in this work we are concerned with $X_1, \ldots, X_n$ chosen independently and identically distributed (iid) according to some distribution f over $\mathcal{X}$; we write this as $X_1, \ldots, X_n \overset{iid}{\sim} f$. While f can in principle be any probability distribution, we will be concerned with continuous probability distributions where f represents a probability density function. Since f represents a continuous probability distribution, it satisfies $f : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ such that for any subset $\mathcal{X}^* \subseteq \mathcal{X}$, we have $\mathbb{E}_{X \sim f}[X \in \mathcal{X}^*] = \int_{\mathcal{X}^*} f(x)\, dx$. We denote the support of f as

$$\mathcal{X}_f := \{x \in \mathcal{X} : f(x) > 0\} \tag{1.16}$$

which we will assume is bounded (Assumption 5).

1.2.6 With Very High Probability

Finally, in order to completely state our results, we define the notion of *with very high probability*, which is a stronger version of the common notion of *with high probability*.

Definition 7. Let $\{A_n\}_{n \in \mathbb{Z}_{\geqslant 0}}$ be an infinite sequence of events parameterized by an integer n. Then we say that A_n happens *with very high probability* (wvhp) if there are constants $c_1, c_2, c_3 > 0$ such that, for all sufficiently large n,

$$\mathbb{P}[A_n] \geqslant 1 - c_1 e^{-c_2 n^{c_3}} \tag{1.17}$$

This implies that $\lim_{n \to \infty} \mathbb{P}[A_n] = 1$, and converges faster than any inverse polynomial (depending on definition it can even be said to converge to 1 'exponentially'). The rapid convergence rate, aside from being of interest itself, allows bounds on the

expected tour length to follow immediately from our probabilistic bounds as well as other results to immediately follow; for more, see Appendix A.1.1.

1.3 Main Results

In this section we state our main results. While our main results depend on a number of conditions, for clarity we will state the results first. The general conditions needed are given in Section 2.1, while additional assumptions for the upper bounds (for the symmetric and nonsymmetric case, respectively) are given in Section 2.3 and Section 2.4.

1.3.1 Parameters

We first define some parameters needed to state our theorems. Since the formal definitions can be fairly involved, we will sketch the meaning of certain parameters which will be formally defined later in Section 2.1; wherever we do we will refer to the specific place it is defined.

First, recall that γ is the small time constraint factor and $g(x)$ is the agility function (Definition 4), satisfying the rough relation

$$\sup_{\boldsymbol{q}\in[x]^{\mathcal{Q}}} \mathrm{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon}(\boldsymbol{q})) \approx g(x)\varepsilon^{\gamma} \tag{1.18}$$

Then we have the integer *branching factor* b which generally denotes how many ε-configuration reachable sets it takes to cover a (2ε)-configuration reachable set (see Assumption 8). We then use this to define a parameter β which will be used in our lower bound:

Definition 8. Let b be the branching factor and γ be the small-time constraint factor of Π, and r be a constant where $r = 2$ when Π is nonsymmetric and $r = 3/2$

when Π is symmetric. Then we let the *lower constant* be:

$$\beta = (1+\xi)r^\gamma \quad \text{where} \quad \xi = \begin{cases} 3(\log(b)/r^\gamma) & \text{if } \log(b) > r^\gamma \\ 3\sqrt{\log(b)/r^\gamma} & \text{if } \log(b) \leqslant r^\gamma \end{cases} \tag{1.19}$$

Note that $\beta = (1+\xi)2^\gamma \propto \max(\log(b), r^\gamma)$.

For the upper bound for symmetric vehicles, we need to consider *Symmetric Hierarchical Cell Structures* (SHCS) (see Definition 11), which we sketch here. A SHCS is a recursively-defined structure of nested cells, having an integer *scaling parameter* s and an *efficiency parameter* $\alpha \leqslant 1$. An SHCS at scale ε rooted at some $\boldsymbol{q} \in \mathcal{Q}$ is defined by a cell contained in some $\bar{R}_\varepsilon(\boldsymbol{q})$ whose volume is at least $\alpha g([\boldsymbol{q}]_\mathcal{X})\varepsilon^\gamma$ (roughly speaking, as there is an additional approximation term in Definition 11), which can covered by s^γ SHCS's at scale ε/s (the next 'level' down). The scaling parameter s thus denotes how much smaller the SHCS's get when going down a level, and the efficiency parameter α denotes how large the cells are relative to the largest reachable sets at the same scale. Typically, SHCS's can be constructed with scaling factor $s = 2$; for instance, control-affine systems satisfying common regularity conditions have SHCS's with $s = 2$ (see Proposition 4 in Section 2.3). However, we make a more general definition which can be extended to the Euclidean TSP with target points distributed on a set of fractal dimension, which in certain cases (like the Menger Sponge) are naturally scaled using $s \neq 2$; see Appendix C for more.

For the upper bound for nonsymmetric vehicles, we have to consider Nonsymmetric Hierarchical Cell Structures (NHCS), which are also defined by nested cells, though these do not come with any parameters which affect the result since for the nonsymmetric upper bound we are only concerned about the order-of-growth.

1.3.2 Main concentration bounds

We first state a trivial, non-probabilistic bound which follows from Assumption 4:

Proposition 1. For some constant $C > 0$, for all $x_1, x_2, \ldots, x_n \in \mathcal{X}_f$,

$$\mathrm{TSP}_\Pi(x_1, \ldots, x_n) \leqslant Cn\,. \tag{1.20}$$

Proof. For each x_i, we select (arbitrarily) some $\boldsymbol{q}_i$ such that $[\boldsymbol{q}_i]_\mathcal{X} = x_i$; then by Assumption 4, there is some C such that $d_\Pi(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \leqslant C$ for all $i = 1, \ldots, n-1$, and hence we have a tour which goes through $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ by the order of indices with length $\leqslant C(n-1)$. □

We now state the main theorem, which gives a very high probability lower bound to the DSTSP, and a matching (up to a constant in n and f) upper bound when Π is symmetric. The lower bound uses the constant β, which in turn is based on the branching factor b (and the small-time constraint factor γ); the upper bound, being derived from an algorithm reliant on hierarchical cell structures, uses s and α. Both upper and lower bounds feature the growth rate $n^{1-\frac{1}{\gamma}}$ and a term which demonstrates how density f and agility g interact to affect the tour length.

Theorem 1. If the assumptions in Section 2.1 hold, then for any $\delta > 0$, the following holds for sufficiently large n:

$$\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \quad \textit{wvhp} \tag{1.21}$$

where β is the lower constant (see Definition 8).

Furthermore, if Π is symmetric and $\mathcal{X}_f$ can be covered with Symmetric Hierarchical Cell Structures with scaling parameter $s \geqslant 2$ and efficiency parameter $\alpha \leqslant 1$ (see Definition 11 and Assumption 10), then for any $\delta > 0$ the following holds for sufficiently large n:

$$\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big) n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \quad \textit{wvhp} \tag{1.22}$$

Finally, if Π is nonsymmetric and can be covered with Nonsymmetric Hierarchical Cell Structures (see Definition 14), then there is some constant $C > 0$

(which can depend on the dynamics and on the target point density f) such that

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant Cn^{1-\frac{1}{\gamma}} \quad wvhp\,. \tag{1.23}$$

These results extend to the expected value (see Appendix A.1.1 and Proposition 21 for the proof) due to Proposition 1:

Corollary 1. If the assumptions in Chapter 2 hold, then for any $\delta > 0$,

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[\mathrm{TSP}_\Pi(\{X_i\})] \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{1.24}$$

for all sufficiently large n. Furthermore, if Π is symmetric and $\mathcal{X}_f$ can be covered with Symmetric Hierarchical Cell Structures with scaling parameter $s \geqslant 2$ and efficiency parameter α (see Definition 11), then for any $\delta > 0$,

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[\mathrm{TSP}_\Pi(\{X_i\})] \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big) n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{1.25}$$

for all sufficiently large n.

Finally, if Π is nonsymmetric, then there is some constant $C > 0$ (which can depend on the dynamics and on the target point density f) such that

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[\mathrm{TSP}_\Pi(X_1, \ldots, X_n)] \leqslant Cn^{1-\frac{1}{\gamma}}\,. \tag{1.26}$$

Taken together, these results yield the general rule that for symmetric Π,

$$\mathrm{TSP}_\Pi(\{X_i\}) = \Theta\Big(n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\Big) \tag{1.27}$$

and the multiplicative gap between the upper and lower bounds is (up to the approximation factor)

$$c_{\mathrm{gap}} := c_{\mathrm{gap}}(\Pi) = 12s\alpha^{-\frac{1}{\gamma}}\beta \tag{1.28}$$

where $\beta \propto \max(\log(b), (3/2)^\gamma)$ (since Π is symmetric).

Note that this gap is a constant with regard to the density function f, since none of the parameters depend on f. Therefore, (1.28) can be viewed as a statement on how the TSP tour length varies as f is changed (provided it still satisfies the assumptions given in Section 2.1). While c_{gap} has no direct dependence on g, both g and c_{gap} depend on Π; however, the integral $\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx$ is still a good indicator of how fast a vehicle governed by Π can visit many targets distributed iid by f.

Remark 3. *Our techniques also allow us to achieve similar results to Theorem 1 for the case of the Euclidean TSP and targets distributed on a set of fractal dimension; see Appendix C.*

1.3.3 DSTSP lower bound

We now state more precise versions of the bounds from Theorem 1, in particular giving the convergence rates.

Proposition 2. Let the assumptions in Section 2.1 hold, and let

$$v_{f,g} = \operatorname{Var}_{f\sim X}\big[f(X)^{-\frac{1}{\gamma}} g(X)^{-\frac{1}{\gamma}}\big] \tag{1.29}$$

$$= \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}} g(x)^{-\frac{2}{\gamma}}\, dx - \left(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\right)^2 \tag{1.30}$$

Then for any $\delta > 0$, there is some $\lambda > 0$ such that for sufficiently large n,

$$\mathbb{P}\left[\operatorname{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\right] \tag{1.31}$$

$$\geqslant 1 - \exp\left(-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}\right) \tag{1.32}$$

$$- \exp\left(-\frac{n(1-\delta/2)\big(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\big)^2 \delta^2/32}{(1+\delta/2)v_{f,g} + (1-\delta/2)\big(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\big)^2 \delta/12}\right) \tag{1.33}$$

Proposition 3. If Π is symmetric with SHCS's with scaling parameter $s \geqslant 2$ and efficiency parameter α, for any $\delta > 0$: for $\gamma \geqslant 3$,

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{1.34}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1-2/\gamma}} \tag{1.35}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1/3}} \tag{1.36}$$

for all sufficiently large n. For a more precise bound, let $\boldsymbol{p} = (p_1, \ldots, p_m)$ be the probability masses of the m cells in the SHCS. Then

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{1.37}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1-2/\gamma}(\sum_{j=1}^m p_j^{1-1/\gamma})^2} \tag{1.38}$$

for all sufficiently large n. Note that $1 \leqslant \sum_{j=1}^m p_j^{1-1/\gamma} \leqslant m^{\frac{1}{\gamma}}$, and is maximized when all m cells in the SHCS have mass $= 1/m$.

If $\gamma = 2$, and (wlog) p_1 is the smallest nonzero value of $\boldsymbol{p} = (p_1, \ldots, p_m)$, then for all sufficiently large n,

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{1.39}$$

$$\geqslant 1 - e^{-\frac{(2/9)p_1 n(\sum_{j=1}^m p_j^{1/2})^2}{127-\log(1/p_1)+\log(n)}} \tag{1.40}$$

Remark 4. *Note that while the probability bound for $\gamma = 2$ converges to 1 faster in the limit, in the sense that it is $1 - e^{-\Theta(n/\log(n))}$ as opposed to $1 - e^{-\Theta(n^{1-2/\gamma})}$, the $\gamma \geqslant 3$ bound avoids the p_1 term, which can be very small. However, for $\gamma \geqslant 3$, we can get even faster convergence of $1 - e^{-\Theta(n)}$; but the probability of failure includes some terms which potentially stay relatively large until n is sufficiently big.*

1.3.4 DTSP with adversarial targets

Theorem 1 naturally leads to the following questions:

- Given a control system Π with agility function g and some bounded $\mathcal{X}^* \subseteq \mathcal{X}$, what probability density function f with support (contained in) $\mathcal{X}^*$ makes the TSP trajectory on $X_i \overset{iid}{\sim} f$ the longest as $n \to \infty$?

- How does this 'worst-case density function' compare to adversarial target points (i.e. chosen to maximize the length of the optimal TSP trajectory)?

We will show that the difference between these (both in expectation and with very high probability) is at most a constant factor which depends on certain characteristics of the dynamic constraints but not on the agility function g, resulting in the bound:

Theorem 2. Let Π be dynamics satisfying the assumptions in Section 2.1 and $\mathcal{X}^* \subseteq \mathcal{X}$ be bounded. Then for any $\delta > 0$, for any sufficiently large n,

$$\sup_{X_1,\dots,X_n \in \mathcal{X}^*} \big(\mathrm{TSP}_\Pi(\{X_i\})\big) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(x)^{-1}\, dx \Big)^{\frac{1}{\gamma}} \tag{1.41}$$

If Π is symmetric with a SHCS with scaling parameter $s \geqslant 2$ and efficiency parameter α, then for any $\delta > 0$, for any sufficiently large n,

$$\sup_{X_1,\dots,X_n \in \mathcal{X}^*} \big(\mathrm{TSP}_\Pi(\{X_i\})\big) \leqslant (1+\delta)\big(6s\alpha^{-\frac{1}{\gamma}}\big) n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(x)^{-1}\, dx \Big)^{\frac{1}{\gamma}}. \tag{1.42}$$

However, there exist examples where nonsymmetric Π satisfies the assumptions in Section 2.1 hold but

$$\sup_{X_1,\dots,X_n \in \mathcal{X}^*} \big(\mathrm{TSP}_\Pi(\{X_i\})\big) = \Theta(n) \tag{1.43}$$

so the matching adversarial upper bound is only achievable in general when Π is symmetric.

This means that for symmetric Π and any $\delta > 0$,

$$(1-\delta)\beta^{-1} \leqslant \frac{\sup_{X_1,\dots,X_n \in \mathcal{X}^*} \mathrm{TSP}_\Pi(X_1,\dots,X_n)}{n^{1-\frac{1}{\gamma}} (\int_{\mathcal{X}^*} g(x)^{-1}\, dx)^{\frac{1}{\gamma}}} \leqslant (1+\delta) 6s\alpha^{-\frac{1}{\gamma}} \tag{1.44}$$

for any sufficiently large n.

Note that these are not probabilistic bounds since $X_1, \ldots, X_n$ are not random anymore. The lower bound is directly derived from the 'worst-case target distribution' $f_g(x) \propto g(x)^{-1}$ or, to normalize,

$$f_g(x) = \frac{g(x)^{-1}}{\int_{\mathcal{X}^*} g(y)^{-1}\, dy}\,. \tag{1.45}$$

The fact that the given lower bound holds at all (let alone with very high probability) when $X_1, \ldots, X_n \overset{iid}{\sim} f_g$ then implies that a non-random adversary can choose a set of target points which makes the shortest tour at least this large. The upper bound is derived from our algorithm which covers the space using Symmetric Hierarchical Cell Structures.

Remark 5. *The DSTSP bound's dependence on f, g takes the form of an integral over $\mathcal{X}_f$ of a function of f, g; this is because in some sense (when $n \to \infty$) the length of the TSP depends only on local conditions. Even if two regions are close, with enough target density the efficiency of visiting targets in each becomes roughly independent. However, the Adversarial DTSP has an exponent outside the integral. This is because varying g locally will change where the targets go, and hence change the target density over the whole space.*

Remark 6. *One odd feature of Theorem 2 is that the adversarial upper bound (equation* (1.42)*) has a better absolute constant* 6 *than the constant* 12 *in thee equivalent stochastic bound (equation* (1.25)*). This is because in the adversarial case we get a hard (deterministic) bound, but in the random case when f does not maximize the integral, an 'unlucky' random placement of targets might cause the tour to be larger than it was expected to be. To get our probabilistic bounds we multiply by a buffer constant (we use* 2*) which was not needed in the adversarial case.*

Chapter 2

Assumptions and Geometry

In this chapter we discuss the assumptions required for our theorems and algorithms and the underlying geometry of control-affine systems. While the geometry of certain vehicles imply many of these assumptions as lemmas we present them as assumptions since it's possible that in practice one wants to show them directly; where they can be shown we give lemmas.

The chapter will be organized as follows:

- In Section 2.1 we state the assumptions needed for our results. We also state and discuss the *meta-assumptions*, which are more fundamental properties of the dynamics and the relationship between the workspace $\mathcal{X}$ and the configuration space $\mathcal{Q}$, and from which some of the assumptions can be derived as lemmas.

- In Section 2.2 we show that many of the assumptions from Section 2.1 follow directly if the dynamics take a certain (common and general) form, particularly from the Uniform Ball-Box Theorem (when the vehicle is symmetric) or the Nonsymmetric Ball-Box Conjecture (when the vehicle is nonsymmetric); the Uniform Ball-Box Theorem is a well-known theorem from sub-Riemannian geometry which is often used in motion planning, and the Nonsymmetric Uniform Ball-Box Conjecture[1] is a variant for nonsymmetric vehicles that we formulate.

[1]For a thorough analysis of the Nonsymmetric Uniform Ball-Box Conjecture, including the proofs of several partial results which may possibly be used to prove it in the future, see Appendix B.

This means that to use our results it is often sufficient to know that they system being considered takes a certain geometric form (rather than needing to show all the assumptions hold) when the vehicle is symmetric; when the vehicle is nonsymmetric, one can alternatively prove that the Nonsymmetric Uniform Ball-Box Conjecture holds for the dynamics in question.

- In Sections 2.3 and 2.4, we discuss the geometric structures on which our algorithms are based (and in turn give our upper bounds to the DSTSP). In both cases, we refer to these as *Hierarchical Cell Structures*, since they involve a system of recursively nesting cells (inscribed within reachable sets).

- In Section 2.5 we address the fact that in certain important cases some assumptions may not hold everywhere - for instance, Lipschitz continuity assumptions on the target density function f. In this case we formulate a notion of *approximately everywhere* which suffices to cover the main instances we wish to address, e.g. where the density function is piecewise constant (and thus has sharp discontinuities). We then show in this section that our main results still hold even if certain assumptions only hold approximately everywhere.

2.1 Assumptions

Our main results follow from a few assumptions. These generally deal either with the properties of the target point distribution f or of the properties of the dynamic constraints Π at small scales.

In this section we specify our main assumptions, which suffice to show our very high probability lower bound (see Chapter 3). Additional definitions and assumptions needed for the upper bounds are given in Section 2.3 (which deals with symmetric vehicles) and Section 2.4 (which deals with nonsymmetric vehicles).

2.1.1 Meta-assumptions

If one wants to apply our results for a given vehicle and target point distribution, one can directly prove all the assumptions given here (and Section 2.3 and Section 2.4) hold for the vehicle in question; however, if the vehicle is control-affine many of the assumptions follow from more basic *meta-assumptions*, which may be easier to show instead. We state the meta-assumptions here and give some intuition as to their meaning; they will be discussed in much greater depth in Section 2.2, where we show how they imply certain assumptions.

The first meta-assumption deals with the dynamics Π:

Meta-Assumption 1. The dynamics Π are *control-affine* and *equiregular* ([19], Definition 2.10) over $[\mathcal{X}_f]^{\mathcal{Q}}$, which is contained in some compact $\mathcal{Q}^* \subseteq \mathcal{Q}$.

This concerns the vector fields $h_0^{\text{cont}}, \ldots, h_m^{\text{cont}}$ which define a control-affine system (see Definition 2). They define the directions of possible motion from a given $\boldsymbol{q} \in \mathcal{Q}$; by combining controls we can directly move in the subspace of $T_{\boldsymbol{q}}\mathcal{Q}$ (the tangent space of $\mathcal{Q}$ at $\boldsymbol{q}$) spanned by $h_0^{\text{cont}}(\boldsymbol{q}), \ldots, h_m^{\text{cont}}(\boldsymbol{q})$. However, by using *commutators* (for instance, going forward along h_1^{cont}, then forward along h_2^{cont}, then backward along h_1^{cont} and finally backward along h_2^{cont}) it may be possible to simulate motions outside the span of $h_0^{\text{cont}}(\boldsymbol{q}), \ldots, h_m^{\text{cont}}(\boldsymbol{q})$, though at the cost of making the motion considerably slower (e.g. to move ε^2 in the new direction may take ε time). These commutators are called *Lie brackets*, and if they can be iterated until they span the tangent space (though the more iterations are produced, the slower the motion in the given direction) we call the dynamics *nonholonomic*. Meta-Assumption 1 condition says that how these Lie brackets (which are collectively called the *Lie algebra*) span the tangent spaces of $\mathcal{Q}$ is the same, broadly speaking, through $\mathcal{Q}$. For symmetric control-affine dynamics, this yields important tools such as the Chow-Rashevskii theorem and especially the Uniform Ball-Box Theorem ([19], Thm 2.4), which show that reachable sets around any $\boldsymbol{q} \in \mathcal{Q}$ always contain a rectilinear neighborhood of a certain size around $\boldsymbol{q}$. For nonsymmetric dynamics, the presence of the drift vector field (which cannot be moved backwards in) means we must define new, special variant of

the Uniform Ball-Box Theorem, which we call the *Nonsymmetric Uniform Ball-Box Conjecture* (since we are unable so far to prove it holds for all Π, see Section 2.2 for formal statement and Appendix B for more discussion). If this holds, then Meta-Assumption 1 also implies the same assumptions for nonsymmetric dynamics as for symmetric dynamics.

The second meta-assumption deals with the projection $[\cdot]_{\mathcal{X}}$:

Meta-Assumption 2. There is some $\mathcal{Q}^*$ such that $[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \mathcal{Q}^*$ which is closed, connected and projection-regular.

See Definition 10 for the formal definition of projection regularity, and Section 2.2.1 for additional discussion. In the same way that the results depend on how the Lie algebra of $h_0^{\text{cont}}, \ldots, h_m^{\text{cont}}$ spans the tangent spaces of $\mathcal{Q}$, they depend on how it projects onto $\mathcal{X}$. In particular, if we look at the projections of the Lie algebra vectors onto the workspace $\mathcal{X}$, we want them to *also* span the tangent spaces of the workspace in a consistent way everywhere.

We will note which assumptions follow from the meta-assumptions and which have to be made on their own (in particular, any assumption concerning the distribution f of the target points has nothing to do with the meta-assumption); the proofs are given in Section 2.2.

Remark 7. *Our results follow if the assumptions in this chapter are met, even if the meta-assumptions themselves do not hold.*

2.1.2 ζ-regularized approximations

To show our results, we will often want to assume that certain functions are Lipschitz continuous or bounded away from 0 or ∞, or both. However, in many important cases this may not hold. Thus, we want to use approximations of these functions that do have these properties, which we call ζ-*regularized approximations*; $\zeta > 0$ governs the degree of approximation and the Lipschitz constant and bound away from 0 or ∞. Depending on whether we want our approximation to be an upper or lower bound, we define the *upper* and *lower* ζ-regularized approximations.

Definition 9. For any function h and $\zeta > 0$ on domain $\mathcal{Y}$ with metric $d_{\mathcal{Y}}$, let

$$\widehat{\mathcal{F}}(h,\zeta) := \{h^* : h^*(y) \geqslant \max(h(y),\zeta), h^* \text{ is } (1/\zeta)\text{-Lipschitz}\} \tag{2.1}$$

$$\text{and} \quad \check{\mathcal{F}}(h,\zeta) := \{h^* : h^*(y) \leqslant \min(h(y),1/\zeta), h^* \text{ is } (1/\zeta)\text{-Lipschitz}\}\,. \tag{2.2}$$

We define the *ζ-regularized upper and lower approximations* $\hat{h}^{(\zeta)}$ and $\check{h}^{(\zeta)}$ of h respectively as

$$\hat{h}^{(\zeta)}(y) := \inf_{h^* \in \widehat{\mathcal{F}}(h,\zeta)} (h^*(y)) \tag{2.3}$$

$$\text{and} \quad \check{h}^{(\zeta)}(y) := \sup_{h^* \in \check{\mathcal{F}}(h,\zeta)} (h^*(y))\,. \tag{2.4}$$

We show that $\hat{h}^{(\zeta)}$ and $\check{h}^{(\zeta)}$ have desirable properties while not being far from h:

Lemma 1. *For any nonnegative h which is continuous almost everywhere and bounded above by some finite $h_{\max} = \sup_y h(y)$ on domain $\mathcal{Y}$, the following hold for the upper ζ-regularized function $\hat{h}^{(\zeta)}$:*

i. $\hat{h}^{(\zeta)}(y) \geqslant h(y)$ *and* $\hat{h}^{(\zeta)}(y) \geqslant \zeta$ *for all* $y \in \mathcal{Y}$.

ii. $\lim_{\zeta \to 0} \hat{h}^{(\zeta)}(y) = h(y)$ *almost everywhere.*

iii. $\hat{h}^{(\zeta)}(y)$ *is decreasing for any fixed* y *as a function of* ζ.

iv. $\hat{h}^{(\zeta)}(y)$ *is* $(1/\zeta)$*-Lipschitz continuous.*

v. If h has finite integral on a bounded set $\mathcal{A}$,

$$\lim_{\zeta \to 0} \int_{\mathcal{A}} \hat{h}^{(\zeta)}(x)\,dx = \int_{\mathcal{A}} h(x)\,dx \tag{2.5}$$

vi. For any $\zeta \leqslant h_{\max}$, we have $\max_y \hat{h}^{(\zeta)}(y) \leqslant h_{\max}$. (Or, in other words, $\sup_y \hat{h}^{(\zeta)}(y) = \max(\zeta, h_{\max})$).

Analogous results hold for the lower ζ-regularized function $\check{h}^{(\zeta)}$.

Proof. We show these as follows:

i: This follows from the definition of $\hat{h}^{(\zeta)}(y)$.

ii: This follows since it holds at all y at which h is continuous, which is almost everywhere.

iii: This follows since $\mathcal{F}(h,\zeta') \subseteq \mathcal{F}(h,\zeta)$ if $\zeta' > \zeta$, and therefore an infimum over values in $\mathcal{F}(h,\zeta)$ will be smaller than the same infimum over values in $\mathcal{F}(h,\zeta')$.

iv. This follows because the infimum of a set of c-Lipschitz continuous functions is also c-Lipschitz continuous. This is because of the following argument by contradiction. Suppose $\hat{h}^{(\zeta)}$ is not $(1/\zeta)$-Lipschitz continuous; then there is $y_1, y_2 \in \mathcal{Y}$ such that

$$|\hat{h}^{(\zeta)}(y_2) - \hat{h}^{(\zeta)}(y_1)| > (1/\zeta) d_{\mathcal{X}}(y_1, y_2)\,. \tag{2.6}$$

WLOG let $\hat{h}^{(\zeta)}(y_1) < \hat{h}^{(\zeta)}(y_2)$. Then there is some $h^* \in \mathcal{F}(h,\zeta)$ such that

$$h^*(y_1) - \hat{h}^{(\zeta)}(y_1) < |\hat{h}^{(\zeta)}(y_1) - \hat{h}^{(\zeta)}(y_2)| - (1/\zeta) d_{\mathcal{Y}}(y_1, y_2) \tag{2.7}$$

$$\implies h^*(y_2) - h^*(y_1) \geqslant \hat{h}^{(\zeta)}(y_2) - h^*(y_1) \tag{2.8}$$

$$> (\hat{h}^{(\zeta)}(y_2) - \hat{h}^{(\zeta)}(y_1)) - \big(|\hat{h}^{(\zeta)}(y_1) - \hat{h}^{(\zeta)}(y_2)| - (1/\zeta) d_{\mathcal{Y}}(y_1, y_2)\big) \tag{2.9}$$

$$= (1/\zeta) d_{\mathcal{Y}}(y_1, y_2) \tag{2.10}$$

so h^* is also not $(1/\zeta)$-Lipschitz continuous, which is a contradiction since $h^* \in \mathcal{F}(h,\zeta)$.

v. This follows from (ii) and the Dominated Convergence Theorem (where the dominating function is $f_{\max}$ over $\mathcal{X}_f$, which is integrable since by the conditions on f, $\mathcal{X}_f$ is bounded).

vi. This follows because the constant function with value $\max(\zeta, h_{\max})$ is an upper bound to h, at least as large as ζ and Lipschitz continuous with any parameter (since it is constant), and hence is in $\widehat{\mathcal{F}}(h,\zeta)$. Then by definition $\hat{h}^{(\zeta)}(y)$ must be at most $\max(\zeta, h_{\max})$ (and since $\hat{h}^{(\zeta)}(y) \geqslant \max(\zeta, h(y))$ for all y, this means $\sup_y \hat{h}^{(\zeta)}(y) =$

$\max(\zeta, h_{\max})$).

Proofs for the lower ζ-regularized function are analogous. □

The upper ζ-regularized approximation will be important to showing the TSP lower bound (in general we can show a lower bound by assuming that the vehicle is more agile than it really is, i.e. using an upper bound of g in place of g itself keeps the DSTSP lower bound valid) while the lower ζ-regularized approximation will be important to showing the TSP upper bound.

2.1.3 Assumptions about the dynamics

We now consider some assumptions we need to make about the dynamics Π. Let $d_{\mathcal{X}}(\cdot,\cdot)$ and $d_{\mathcal{Q}}(\cdot,\cdot)$ denote metrics on the manifolds $\mathcal{X}$ and $\mathcal{Q}$. These induce norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Q}}$ on the tangent spaces at any $x \in \mathcal{X}$ and $\boldsymbol{q} \in \mathcal{Q}$.

Note that the lengths of the tours are *not* necessarily measured in the metrics $d_{\mathcal{X}}$ or $d_{\mathcal{Q}}$ but rather by the time it takes to execute the tour, i.e. by d_Π – these are not necessarily equivalent because different control inputs to the vehicle might cause motions of different speeds relative to these metrics. We introduce these to aid in the analysis and in particular to make the following assumption well-defined:

Assumption 2. The manifold $\mathcal{X}$ has dimension $\geqslant 2$.

This is to avoid degenerate cases. Typically, when $\mathcal{X}$ is 1-dimensional, the TSP tour length is constant since one can simply sweep the entire space.

Assumption 3. There is some 'speed limit' c_Π of Π such that for any $\boldsymbol{q} \in \mathcal{Q}$ and control $\boldsymbol{u} \in \mathcal{U}$, we have $\|h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u})\|_{\mathcal{X}} \leqslant c_\Pi$.

Note that Assumption 2 and Assumption 3 together imply the following:

Lemma 2. *The small-time constraint factor γ is at least 2.*

Proof. Let $\mathcal{B}_\varepsilon(x)$ denote the radius-ε ball around $x \in \mathcal{X}$ under metric $d_{\mathcal{X}}$. Then

$$\bar{R}_\varepsilon(\boldsymbol{q}) \subseteq \mathcal{B}_{c_\Pi \varepsilon}([\boldsymbol{q}]_{\mathcal{X}}) \tag{2.11}$$

$$\implies \text{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})) \leqslant \text{Vol}_{\mathcal{X}}(\mathcal{B}_{c_\Pi \varepsilon}([\boldsymbol{q}]_{\mathcal{X}})) \tag{2.12}$$

and we know that $\mathrm{Vol}_{\mathcal{X}}(\mathcal{B}_{c_\Pi \varepsilon}([\boldsymbol{q}]_{\mathcal{X}})) \propto \varepsilon^{\dim(\mathcal{X})}$ for small ε. Thus, since $\log(\varepsilon) < 0$,

$$\gamma = \lim_{\varepsilon \to 0} \frac{\log(\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})))}{\log(\varepsilon)} \geqslant \frac{\log(\mathrm{Vol}_{\mathcal{X}}(\mathcal{B}_{c_\Pi \varepsilon}([\boldsymbol{q}]_{\mathcal{X}})))}{\log(\varepsilon)} = \dim(\mathcal{X}) \geqslant 2\,. \tag{2.13}$$

□

We also assume that any two configurations corresponding to points in the support $\mathcal{X}_f$ of the target point distribution are connected by a finite path $\pi \in \Pi$:

Assumption 4. For any $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$, we have $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') < \infty$. Furthermore, there is some constant C such that if $[\boldsymbol{q}]_{\mathcal{X}}, [\boldsymbol{q}']_{\mathcal{X}} \in \mathcal{X}_f$ (the support of target probability distribution f) then $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant C$.

For symmetric equiregular control-affine dynamics satisfying Chow's condition [19], the first part of this assumption follows from the Chow-Rashevskii theorem. However, for nonsymmetric control-affine dynamics it must be explicitly assumed since there are simple examples in which it doesn't hold (e.g. a boat swept downstream by a current that can't return upstream).[2]

2.1.4 Assumptions about the target distribution

We consider the support $\mathcal{X}_f$ of density f and how f behaves on it.

Assumption 5. The function f is a probability density function represents a full-dimensional continuous random variable, i.e. $f : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ is integrable, continuous almost everywhere, and satisfies $\int_{\mathcal{X}} f(x)\,dx = 1$ and

$$\mathbb{P}_{X \sim f}[X \in \mathcal{A}] = \int_{\mathcal{A}} f(x)\,dx \text{ for any subset } \mathcal{A} \subseteq \mathcal{X}. \tag{2.14}$$

Furthermore, f has a finite maximum value $f_{\max} := \sup_{x \in \mathcal{X}} f(x) < \infty$.

The target region $\mathcal{X}_f = \{x : f(x) > 0\}$ is also compact.

[2] To apply Theorem 1 for nonsymmetric vehicles, Assumption 4 must be shown to hold explicitly.

Except for very contrived examples, assuming that $\mathcal{X}_f$ is compact is really just to ensure it is bounded. If $\mathcal{X}_f$ is allowed to be unbounded, it is easy to produce target point distributions f for which the expected distance between $X_1, X_2 \overset{iid}{\sim} f$ is infinite by using the St. Petersburg paradox: pick an origin point $\mathbf{0} \in \mathcal{X}$ and set f to have support on a sequence of small regions $\mathcal{X}_1, \mathcal{X}_2, \cdots \subset \mathcal{X}$, where $\mathbb{P}_{X\sim f}[X \in \mathcal{X}_i] = 2^{-i}$ but $d_{\mathcal{X}}(\mathbf{0}, X) \approx 2^i$ for all $X \in \mathcal{X}_i$; then a random $X \sim f$ will have infinite expected distance from any point in $\mathcal{X}$, so X_2 will always have expected infinite distance from X_1. The length of the TSP tour will then be infinite in expectation even for 2 target points, let alone $n > 2$.

2.1.5 Assumptions about the reachable set

A key condition for our results to hold is that the agility function be well defined and that the limit it represents converges uniformly over the space:

Assumption 6. The control system Π has an *agility function* $g : \mathcal{Q} \to \mathbb{R}_{\geqslant 0}$ and a *small-time reachability factor* γ such that for any $\rho > 0$, there is some $\varepsilon^*_\rho > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^*_\rho$ and all $\boldsymbol{q} \in \mathcal{Q}$,

$$(1-\rho)g(\boldsymbol{q}) \leqslant \frac{\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leqslant (1+\rho)g(\boldsymbol{q}) \tag{2.15}$$

Note that this implies a similar bound on the workspace agility function:

Lemma 3. *For any $\rho > 0$, there is some $\varepsilon^*_\rho > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^*_\rho$ and all $x \in \mathcal{X}$,*

$$(1-\rho)g(x) \leqslant \sup_{\boldsymbol{q}:[\boldsymbol{q}]_{\mathcal{X}}=x} \frac{\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leqslant (1+\rho)g(x) \tag{2.16}$$

Proof. We let $\varepsilon^*_\rho > 0$ be the same value as in Assumption 6. Then, for any $\boldsymbol{q}$ such that $[\boldsymbol{q}]_{\mathcal{X}} = x$, we have

$$\frac{\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leqslant (1+\rho)g(\boldsymbol{q}) \leqslant (1+\rho)g(x) \tag{2.17}$$

Additionally, we have

$$\sup_{\boldsymbol{q}:[\boldsymbol{q}]_{\mathcal{X}}=x} \frac{\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \geqslant \sup_{\boldsymbol{q}:[\boldsymbol{q}]_{\mathcal{X}}=x} (1-\rho)g(\boldsymbol{q}) = (1-\rho)g(x) \tag{2.18}$$

and we have shown both sides of the inequality. □

We also assume bounds on g over the workspace:

Assumption 7. Letting $g_{\min} := \inf_{x\in\mathcal{X}} g(x)$ and $g_{\max} := \sup_{x\in\mathcal{X}} g(x)$,

$$0 < g_{\min} \leqslant g_{\max} < \infty\,. \tag{2.19}$$

2.1.6 The coverage assumptions

The conditions we need for our technique concern the ability to cover any 2ε-radius reachable set with a constant number of $\approx \varepsilon$-radius reachable sets, and to cover the starting set $\mathcal{Q}_0$ with a polynomially or even exponentially-increasing number of $\approx \varepsilon$-radius reachable sets. We note that this section deals primarily with *configuration* reachable sets, unlike Section 2.1.5.

We begin with the *b-coverability* condition. The *branching factor* $b > 1$ is an integer roughly corresponding to the number of ε-reachable sets needed to cover a 2ε-reachable set in $\mathcal{Q}$:

Assumption 8. There is some integer $b > 1$, called the *branching factor*, such that for any $\rho > 0$, there is some $\varepsilon^*_\rho > 0$ such that for any $\boldsymbol{q} \in \mathcal{Q}$ and $0 < \varepsilon \leqslant \varepsilon^*_\rho$, there is a set $Q_{\rho,\varepsilon}(\boldsymbol{q})$ of configurations such that

$$R_{2\varepsilon}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}'\in Q_{\rho,\varepsilon}(\boldsymbol{q})} R_{(1+\rho)\varepsilon}(\boldsymbol{q}') \text{ and } |Q_{\rho,\varepsilon}(\boldsymbol{q})| \leqslant b \tag{2.20}$$

Lemma 4. *If* Π *satisfies Meta-Assumption 1, then Assumption 8 holds if either* Π *is symmetric or if the Nonsymmetric Uniform Ball-Box Theorem holds for* Π*.*

We also have the *starting coverability* condition, which says that the set $[\mathcal{X}_f]^{\mathcal{Q}}$ (configurations on the support of f) can be covered with a polynomial number of ε-reachable sets:

Assumption 9. There exists a polynomial P such that there is some $\varepsilon^* > 0$ such that for any $\varepsilon < \varepsilon^*$ there is a set Q_ε^0 of configurations such that

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{\boldsymbol{q} \in Q_\varepsilon^0} R_\varepsilon(\boldsymbol{q}) \text{ and } |Q_\varepsilon^0| \leqslant P(1/\varepsilon) \tag{2.21}$$

Note that this uses configuration reachable sets rather than workspace reachable sets.

The real possibility that this guards against (given that $\mathcal{X}_f$ is bounded by Assumption 5) is that the set of configurations one can be in at any $x \in \mathcal{X}_f$ is unbounded. For symmetric control-affine systems, reachable sets contain rectilinear sets with volumes polynomial in ε as $\varepsilon \to 0$ (though the exponent is not γ as we are discussing configuration reachable sets), so if $[\mathcal{X}_f]^{\mathcal{Q}}$ is bounded, the above holds.

One quick way to check Assumption 9 holds is that if Π is symmetric, control-affine, and equiregular (satisfying Meta-Assumption 1), then Assumption 9 is equivalent to the closure of $[\mathcal{X}_f]^{\mathcal{Q}}$ being compact (which typically just means bounded since it's already closed):

Lemma 5. *If* Π *satisfies Meta-Assumption 1, then Assumption 9 holds if either* Π *is symmetric or if the Nonsymmetric Uniform Ball-Box Theorem holds for* Π*.*

This holds because, by the Uniform Ball-Box Theorem, for sufficiently small ε, each configuration ε-reachable set anchored at $\boldsymbol{q}$ contains a box centered at $\boldsymbol{q}$ whose with side lengths polynomial in ε, and therefore with volume polynomial in ε (with the same polynomial for each $\boldsymbol{q}$). Then, the compact $[\mathcal{X}_f]^{\mathcal{Q}}$ can be tiled by polynomially many of these boxes (allowing for some overlap) since $\mathcal{Q}$ is a manifold and is therefore approximately Euclidean on small scales. See Section 2.2 for the complete proof.

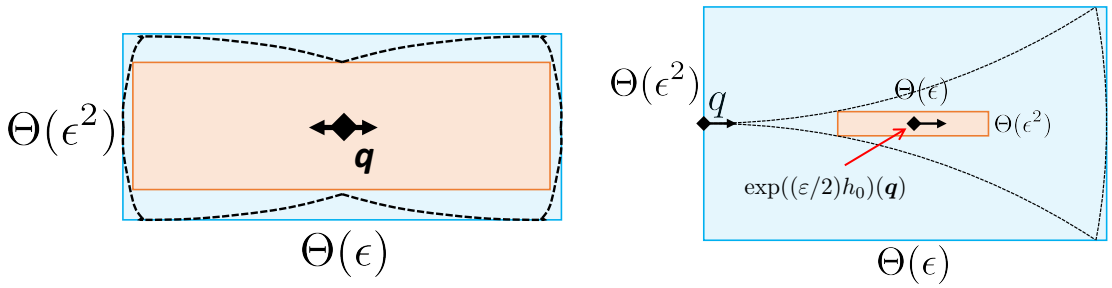

Figure 2-1: Visualization of the Uniform Ball-Box Theorem (left) and the Nonsymmetric Uniform Ball-Box Conjecture (right) with inscribed and circumscribed boxes. Note that this visualization is incomplete because it represents the workspace reachable sets while the Ball-Box Theorem concerns the configuration reachable set (which is harder to depict because in this case it is 3-dimensional).

2.2 Assumptions from Geometry

In this section we show how some of the assumptions in Section 2.1 follow automatically from Meta-Assumption 1, i.e. whenever the control system takes a certain form. We make use of the following notions:

- We have control vector fields $h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$ (see Section 1.2.3) over d-dimensional manifold $\mathcal{Q}$ which define our control-affine system. For a vector field h (not necessarily one of the control vector fields) we denote by $\exp(zh)$ the function which proceeds along h for z time, i.e. it takes $\boldsymbol{q} \in \mathcal{Q}$ and returns a new point $\exp(zh)(\boldsymbol{q})$ which is the point reached from $\boldsymbol{q}$ by following h for z time units.

- The *Lie algebra* of $h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$, which denote the set of directions which can be generated via linear combination and Lie brackets, and which we assume span the tangent space at each $\boldsymbol{q} \in \mathcal{Q}$ (Chow's Condition, [19] Definition 1.6). A Lie bracket I from the *free Lie algebra* on $[m]$ denotes a sequence of Lie bracket operations over m unspecified elements; it can then be applied to our set of control vector fields. We denote by h_I the vector in the Lie algebra over $h_1^{\text{cont}}, \ldots, h_m^{\text{cont}}$ which corresponds to I, and we denote the length of Lie bracket I (i.e. how many bracket operations needed to generate it) by $|I|$.

- *Adapted frames* ([19], page 36), which denote a set of vector fields which are used to generate a coordinate map around some $\boldsymbol{q} \in \mathcal{Q}$ (and which span $T_{\boldsymbol{q}}(\mathcal{Q})$),

in a way compatible with the Lie algebra; the boxes of the Uniform Ball-Box Theorem are defined via this coordinate map.

- *Regular points* ([19], Definition 2.4), which are configurations $\boldsymbol{q} \in \mathcal{Q}$ in which the Lie algebra behaves smoothly in a neighborhood around $\boldsymbol{q}$.

These allow us to use our main tool:

Uniform Ball-Box Theorem. [[19], Theorem 2.4] If $\mathcal{Q}^* \subseteq \mathcal{Q}$ is compact, there exist constants $0 < c < C$ and $\varepsilon_0 > 0$ such that, for $\boldsymbol{q} \in \mathcal{Q}^*$ and $0 < \varepsilon \leqslant \varepsilon_0$ and any adapted frame $\boldsymbol{H} = (h_{I_1}, \ldots, h_{I_d})$ (a set of elements of the Lie algebra of the control vector fields) at $(\boldsymbol{q}, \varepsilon)$,

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \tag{2.22}$$

where $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon) = \{\exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}) : |z_i| \leqslant \varepsilon^{|I_i|} \ \ \forall i\}$.

This states that at small scales ε the configuration ε-reachable set $R_\varepsilon(\boldsymbol{q})$ contains a box of particular dimensions and orientation, and is in turn contained by a box of the same dimensions and orientation scaled up by a constant. The coordinate system in which these boxes are defined is given by $\boldsymbol{H}$: a point at coordinates $(z_1, \ldots, z_d)$ is the one reached from $\boldsymbol{q}$ by traveling for z_d along h_{I_d}, then z_{d-1} along $h_{I_{d-1}}$, and so forth. The limits on $|z_i|$ give the dimensions of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon)$. This then yields, intuitively, a way to cover $R_\varepsilon(\boldsymbol{q})$ with a number of configuration $(\varepsilon/2)$-reachable sets, by taking the circumscribed box of $R_\varepsilon(\boldsymbol{q})$ and tiling it with the inscribed boxes of $R_{\varepsilon/2}(\boldsymbol{q}')$ for a set of well-chosen $\boldsymbol{q}'$.

We also give the nonsymmetric (or drifting) version:

Nonsymmetric Uniform Ball-Box Conjecture. If $\mathcal{Q}^* \subseteq \mathcal{Q}$ is compact, there exist constants $0 < c < C$ and $\varepsilon_0 > 0$ such that, for $\boldsymbol{q} \in \mathcal{Q}^*$ and $0 < \varepsilon \leqslant \varepsilon_0$ and any adapted frame $\boldsymbol{H} = (h_{I_1}, \ldots, h_{I_d})$ (a set of elements of the Lie algebra of the

control vector fields) at $(\boldsymbol{q}, \varepsilon)$,

$$\mathrm{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), C\varepsilon)\,. \tag{2.23}$$

The difference here is that the same boxes as in the Uniform Ball-Box Theorem are being pushed $\varepsilon/2$ down the drift vector field. The reason why it's being pushed $\varepsilon/2$ rather than ε is that, since the vehicle is unable to directly control its movement down drift vector field $h_0 = h_0^{\mathrm{cont}}$, it controls it by taking advantage of the fact that the reachable set is the set of points reachable in time *at most* ε: the default is to only move for $\varepsilon/2$ time, and to move less than $\varepsilon/2$ down h_0 the vehicle can end its motion early; while to move more than $\varepsilon/2$ down h_0 the vehicle can extend its motion for any amount of time between $\varepsilon/2$ to ε.

As discussed, this is a conjecture. Nevertheless, formulating it brings two distinct advantages:

- It prepares the way for potentially proving it future; if it's proved, the main results then immediately follow.
- It offers a shortcut for using our main DSTSP results. For a system of dynamic that one is interested in, instead of proving all the assumptions in Section 2.1, one can prove the Nonsymmetric Ball-Box Theorem holds for the system in question instead.

2.2.1 Projection regularity

We note that merely discussing reachable sets in the configuration space is not sufficient because we often want to deal with workspace reachable sets, in particular when considering their volume. This means that we want to know that the projection function $[\cdot]_{\mathcal{X}}$ is *projection regular* (as stated in Meta-Assumption 2); while this was discussed in Section 2.2.1, we define it formally here. This concept mirrors the concept of regularity as discussed in [19].

First, note that a vector field h on $\mathcal{Q}$ produces for any $\boldsymbol{q} \in \mathcal{Q}$ a tangent vector $h(\boldsymbol{q}) \in T_{\boldsymbol{q}}\mathcal{Q}$; we let d denote the number of dimensions of $\mathcal{Q}$. We also (as

in Section 2.3.1) make the following definitions. Taking the projection to $\mathcal{X}$ (and letting $x(\boldsymbol{q}) = [\boldsymbol{q}]_{\mathcal{X}}$), we get a tangent vector on $[h(\boldsymbol{q})]_{\mathcal{X}} \in T_x\mathcal{X}$ (moving along $h(\boldsymbol{q})$ from $\boldsymbol{q} \in \mathcal{Q}$ projects to a movement along $[h(\boldsymbol{q})]_{\mathcal{X}}$ from $x(\boldsymbol{q})$). We note that $\boldsymbol{H}(\boldsymbol{q}) = \{h_{I_1}(\boldsymbol{q}), \ldots, h_{I_d}(\boldsymbol{q})\}$ is by definition a basis of $T_{\boldsymbol{q}}\mathcal{Q}$; therefore

$$[\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{[h_{I_1}(\boldsymbol{q})]_{\mathcal{X}}, \ldots, [h_{I_d}(\boldsymbol{q})]_{\mathcal{X}}\} \tag{2.24}$$

spans $T_x\mathcal{X}$. Let us denote this by

$$\bar{\boldsymbol{H}}(\boldsymbol{q}) := [\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{\bar{h}_{I_1}(\boldsymbol{q}), \ldots, \bar{h}_{I_d}(\boldsymbol{q})\}\,. \tag{2.25}$$

We now consider at each $\boldsymbol{q}$ the following: let $\triangle_{\boldsymbol{q}}^{(j)}$ denote the subspace of $T_{x(\boldsymbol{q})}\mathcal{X}$ spanned by $\{\bar{h}_{I_i}(\boldsymbol{q}) : |I_i| \leqslant j\}$, and let $\xi^{(j)} = \dim(\triangle_{\boldsymbol{q}}^{(j)})$. While this theoretically can depend on $\boldsymbol{q}$, we would like it to be consistent over the whole space. We thus define:

Definition 10. We say that $\boldsymbol{q}$ is *projection-regular* under $[\cdot]_{\mathcal{X}}$ if there is some neighborhood $\mathcal{Q}'$ of $\boldsymbol{q}$ such that $\xi^{(j)}$ is constant over $\mathcal{Q}'$. We say that $\mathcal{Q}^*$ is projection-regular if all $\boldsymbol{q} \in \mathcal{Q}^*$ are projection-regular.

Remark 8. *We note that γ is intimately related to the $\xi^{(j)}$'s. In particular, consider $\xi^{(j)} - \xi^{(j-1)}$ (where by definition $\xi^{(0)} = 0$); this denotes the increase in dimension of the spanned subspace when adding in $|I_i| = j$ from the adapted frame. The volume of the resulting boxes at scale ε will then be $\varepsilon^{\sum_j j(\xi^{(j)} - \xi^{(j-1)})}$ (note that once j is high enough that $T_x\mathcal{X}$ is spanned, the $\xi^{(j)}$ become constant and the sum ends). Let $d^{(\mathcal{X})}$ be the dimension of $\mathcal{X}$ and j^* be the minimum j for which $\xi^{(j)} = d^{(\mathcal{X})}$; we then get*

$$\gamma = \sum_{j=1}^{j^*} j(\xi^{(j)} - \xi^{(j-1)}) = j^* d^{(\mathcal{X})} - \sum_{j=1}^{j^*-1} \xi^{(j)} \tag{2.26}$$

This yields the following:

Lemma 6. *If $\mathcal{Q}^*$ is connected and projection-regular, then $\xi^{(j)}$ is constant over $\mathcal{Q}^*$ for all j.*

Proof. Let $\mathcal{Q}'(\boldsymbol{q})$ be the (open) neighborhoods of $\boldsymbol{q}$ on which the projection-regularity condition is known to hold. Then trivially,

$$\mathcal{Q}^* \subseteq \bigcup_{\boldsymbol{q} \in \mathcal{Q}^*} \mathcal{Q}'(\boldsymbol{q}) \,. \tag{2.27}$$

We know that all $\mathcal{Q}'(\boldsymbol{q})$ have constant $\xi^{(j)}$'s. Let $\xi^{(j)}(\boldsymbol{q})$ denote the value of $\xi^{(j)}$ at $\boldsymbol{q}$, and suppose (by way of contradiction) that we have some $\boldsymbol{q}, \boldsymbol{q}'$ such that $\xi^{(j)}(\boldsymbol{q}) \neq \xi^{(j)}(\boldsymbol{q}')$. Let

$$\mathcal{Q}'_1 := \bigcup_{\boldsymbol{q}'' : \xi^{(j)}(\boldsymbol{q}'') = \xi^{(j)}(\boldsymbol{q})} \mathcal{Q}'(\boldsymbol{q}'') \text{ and } \mathcal{Q}'_2 := \bigcup_{\boldsymbol{q}'' : \xi^{(j)}(\boldsymbol{q}'') \neq \xi^{(j)}(\boldsymbol{q})} \mathcal{Q}'(\boldsymbol{q}'') \,. \tag{2.28}$$

Both are open since they are both unions of open sets; and both are nonempty because $\boldsymbol{q} \in \mathcal{Q}'_1$ and $\boldsymbol{q}' \in \mathcal{Q}'_2$. Finally, they are (by definition) disjoint and cover $\mathcal{Q}^*$. But this contradicts the fact that $\mathcal{Q}^*$ is connected and we have our desired contradiction. □

Thus, if Meta-Assumption 2 holds (i.e. projection regularity across some connected and closed set $\mathcal{Q}^*$ containing $[\mathcal{X}_f]^{\mathcal{Q}}$), by Lemma 6 we can consider $\xi^{(j)}$ as a constant.

This is a meta-assumption because, like Meta-Assumption 1, we don't use it directly in our results; rather, we use it to prove that under certain geometric conditions, our normal assumptions hold automatically. Note that the $\mathcal{Q}^*$ in Meta-Assumption 2 can be intersected with the $\mathcal{Q}^*$ in Meta-Assumption 1 to create a compact $\mathcal{Q}^*$ containing $[\mathcal{X}_f]^{\mathcal{Q}}$ on which the two meta-assumptions both hold.

This will come into play in Section 2.3.1 and Section 2.4.3 as well.

2.2.2 Coverage assumptions

We now consider the coverage assumptions Assumption 8 and Assumption 9, which both assert that certain sets can be covered by reachable sets in a particular way. Note that since these assumptions deal with *configuration* reachable sets, there is no need for projection regularity.

We begin with Lemma 4, which asserts that if Meta-Assumption 1 holds, then Assumption 8 follows.

Proof of Lemma 4. We will prove that a slightly stronger version of Assumption 8 holds, in which there is some $b > 1$ such that there is some $\varepsilon^* > 0$ such that for any $\boldsymbol{q} \in \mathcal{Q}$ and $0 < \varepsilon \leqslant \varepsilon^*$ there is a set $Q_\varepsilon(\boldsymbol{q})$ of configurations such that

$$R_{2\varepsilon}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_\varepsilon(\boldsymbol{q})} R_\varepsilon(\boldsymbol{q}') \text{ and } |Q_\varepsilon(\boldsymbol{q})| \leqslant b\,. \tag{2.29}$$

This is stronger than Assumption 8 in that it drops the $\rho > 0$ approximation factor, and therefore if this holds then it Assumption 8 trivially holds as well (we can simply let $Q_{\rho,\varepsilon}(\boldsymbol{q}) = Q_\varepsilon(\boldsymbol{q})$ for all $\rho > 0$). As a remark, Assumption 8 is phrased in the more general way because there are instances where it yields a lower value of b, which improves our lower bound in Theorem 1.

We choose some regular $\boldsymbol{q} \in [\mathcal{X}_f]^{\mathcal{Q}}$, which is contained inside some compact subset of $\mathcal{Q}$. Thus, we can find an adapted frame $\boldsymbol{H}$ at $\boldsymbol{q}$ which is an adapted frame for all $\boldsymbol{q}'$ in a neighborhood of $\boldsymbol{q}$; let ε be sufficiently small so that all $\boldsymbol{q}' \in R_\varepsilon(\boldsymbol{q})$ are in this neighborhood, and that $\varepsilon < \varepsilon_0$ from the Uniform Ball-Box Theorem. Then, if $0 < c < C$ are as defined in the Uniform Ball-Box Theorem,

$$R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon)\,. \tag{2.30}$$

This is a box whose size along dimension $i \in [d]$ is $2(C\varepsilon)^{|I_i|}$ Furthermore, for any $\boldsymbol{q}' \in R_\varepsilon(\boldsymbol{q})$, we know

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon/2) \subset R_{\varepsilon/2}(\boldsymbol{q}')\,. \tag{2.31}$$

But these are boxes in the same coordinate system as $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon)$, whose size along dimension $i \in [d]$ is $2(c\varepsilon/2)^{|I_i|}$; this means it is a $(c/(2C))^{|I_i|}$ fraction of the size of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon)$ along dimension i, which is notably constant with regard to ε. Thus, we

can use

$$b = \prod_{i=1}^{d} \lceil (2C/c)^{|I_i|} \rceil \tag{2.32}$$

such boxes to cover $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon)$. Call their centers $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_b$. Then we have

$$R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \subseteq \bigcup_{j=1}^{b} \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_j, c\varepsilon/2) \subseteq \bigcup_{j=1}^{b} R_{\varepsilon/2}(\boldsymbol{q}_j) \tag{2.33}$$

and we are done.

To modify the above proof for the nonsymmetric case, we pull back the centers the appropriate amount: we find the b centers $\boldsymbol{q}_j$ whose boxes cover the box which contains $R_\varepsilon(\boldsymbol{q})$, and then pull them back, i.e.:

$$\mathrm{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), C\varepsilon) \subseteq \bigcup_{j=1}^{b} \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_j, c\varepsilon/2) \tag{2.34}$$

$$\subseteq \bigcup_{j=1}^{b} R_{\varepsilon/2}(\exp(-(\varepsilon/4)h_0)(\boldsymbol{q}_j)) \tag{2.35}$$

which then concludes the proof. □

We then consider Lemma 5, which asserts that if Meta-Assumption 1, then Assumption 9 follows.

Proof of Lemma 5. For simplicity we assume that $[\mathcal{X}_f]^{\mathcal{Q}}$ is itself compact (otherwise we take a compact superset, and if $P(1/\varepsilon)$ configuration ε-reachable sets can cover it then $[\mathcal{X}_f]^{\mathcal{Q}}$ is trivially covered by them too).

By the Uniform Ball-Box Theorem, there exists constants $0 < c < C$ and $0 < \varepsilon_0$ such that any $\boldsymbol{q}$ has an adapted frame $\boldsymbol{H} = (h_{I_1}, \ldots, h_{I_d})$ such that for any $0 < \varepsilon \leqslant \varepsilon_0$ and $\boldsymbol{q} \in [\mathcal{X}_f]^{\mathcal{Q}}$,

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon)\,. \tag{2.36}$$

We let $\varepsilon^*(\boldsymbol{q})$ denote any sufficiently small radius of the reachable set so that

$R_{\varepsilon^*(\boldsymbol{q})}(\boldsymbol{q})$ is contained in the neighborhood of $\boldsymbol{q}$ on which $\boldsymbol{H}$ is also an adapted frame at all other $\boldsymbol{q}'$. Without loss of generality we can assume $\varepsilon^*(\boldsymbol{q}) \leqslant \varepsilon_0$. We then consider the open version of $R_{\varepsilon^*(\boldsymbol{q})}(\boldsymbol{q})$, which is an open set containing $\boldsymbol{q}$ (since control-affine dynamics satisfying Chow's Condition induce reachable sets around $\boldsymbol{q}$ containing a neighborhood of $\boldsymbol{q}$). Thus (with the open reachable sets)

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{\boldsymbol{q} \in [\mathcal{X}_f]^{\mathcal{Q}}} R_{\varepsilon^*(\boldsymbol{q})}(\boldsymbol{q}) \tag{2.37}$$

and since $[\mathcal{X}_f]^{\mathcal{Q}}$ is compact, there is some finite subcover. Let S be the number of sets in this finite subcover, and let $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_S$ be the configurations at their centers, so that

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{i=1}^{S} R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i) \tag{2.38}$$

By the proof of Lemma 4 above, we know that for any $\boldsymbol{q}' \in R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i)$ and any $\varepsilon < \varepsilon^*(\boldsymbol{q}_i)$, the set $R_\varepsilon(\boldsymbol{q}')$ can be covered by b configuration $(\varepsilon/2)$-reachable sets, which can each in turn be covered by b configuration $(\varepsilon/4)$-reachable sets, and so forth. Thus, for any integer $k \geqslant 1$, $R_\varepsilon(\boldsymbol{q}')$ can be covered by b^k configuration $(2^{-k}\varepsilon)$-reachable sets. In particular, $R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i)$ can be covered by b^k configuration $(2^{-k}\varepsilon^*(\boldsymbol{q}_i))$-reachable sets.

Let $\varepsilon^*_{\min} := \min_i \varepsilon^*(\boldsymbol{q}_i)$ and $\varepsilon^*_{\max} := \max_i \varepsilon^*(\boldsymbol{q}_i)$. Then given $0 < \varepsilon \leqslant \varepsilon^*_{\min}$, how many configuration ε-reachable sets are needed to cover $[\mathcal{X}_f]^{\mathcal{Q}}$? We let

$$k_\varepsilon := \min(k : \varepsilon \leqslant 2^{-k}\varepsilon^*_{\max}) = \lceil \log_2(\varepsilon^*_{\max}/\varepsilon) \rceil \leqslant \log_2(\varepsilon^*_{\max}/\varepsilon) + 1 \tag{2.39}$$

This also implies that

$$\varepsilon \geqslant 2^{-k_\varepsilon}\varepsilon^*_{\max} \tag{2.40}$$

Then each $R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i)$ can be covered by b^{k_ε} configuration $2^{-k_\varepsilon}\varepsilon^*_{\max}$-reachable sets, which (using the same anchors) means each $R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i)$ can be covered by b^{k_ε} config-

uration ε-reachable sets. But we can bound

$$b^{k_\varepsilon} \leqslant b^{\log_2(\varepsilon^*_{\max}/\varepsilon)+1} = b(\varepsilon^*_{\max}/\varepsilon)^{\log_2 b} = c^*\varepsilon^{-\log_2 b} \tag{2.41}$$

where $c^* = b(\varepsilon^*_{\max})^{\log_2 b}$, which is constant with regards to ε. Noting that we have to do this with all S boxes $R_{\varepsilon^*(\boldsymbol{q}_i)}(\boldsymbol{q}_i)$ (where S also does not depend on ε), we then get that $[\mathcal{X}_f]^{\mathcal{Q}}$ can be covered with $\leqslant Sc^*\varepsilon^{-\log_2 b}$ configuration ε-reachable sets, which is a polynomial number in $1/\varepsilon$, so we are done.

To modify the above proof for the Nonsymmetric Uniform Ball-Box Conjecture, we simply take the same boxes with centers $\boldsymbol{q}_i$ and do the same pull-back, using $\boldsymbol{q}'_i = \exp(-(\varepsilon^*(\boldsymbol{q}_i)/2)h_0)(\boldsymbol{q}_i)$ as our centers for covering $[\mathcal{X}_f]^{\mathcal{Q}}$. □

2.3 Symmetric Hierarchical Cell Structures

In this section we describe a key notion we need for our algorithm for the Symmetric DSTSP, which we will explore in Chapter 4. We call this structure a *Symmetric Hierarchical Cell Structure* (Symmetric HCS or SHCS); our algorithm depends on its existence everywhere in $\mathcal{X}_f$ (or 'approximately everywhere', see Section 2.5), a condition which holds automatically if the system is control-affine and regular (see Section 2.2, Proposition 4).

A Symmetric HCS is a structure of nesting cells which are inscribed in reachable sets of increasingly small radii. It generally reflects the intuition that a ε-workspace reachable set should be divisible into 2^γ $(\varepsilon/2)$-workspace reachable sets, which should each be divisible into 2^γ $(\varepsilon/4)$-workspace reachable sets, and so on, forming a sort of hierarchy of reachable sets of exponentially decreasing radius. Our algorithm assumes that any $x \in \mathcal{X}_f$ is contained in some SHCS, which, as we will show, then allows us to cover the space with (finitely many) SHCS's with minimal overlap; furthermore, if the dynamics (which are symmetric in this chapter) satisfy Meta-Assumption 1, we can show that every $x \in \mathcal{X}_f$ does in fact fall inside some SHCS. We define:

Definition 11. A *Symmetric Hierarchical Cell Structure* $\mathrm{HCS}(\boldsymbol{q}_0, \varepsilon_0, \alpha, s, \zeta)$ rooted at $\boldsymbol{q}_0 \in \mathcal{Q}$ with radius $\varepsilon_0 > 0$, efficiency parameter $\alpha \leqslant 1$, scaling parameter s, and is ζ-regular, is a structure consisting of the following elements:

- A Jordan-measurable *cell* $S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \mathcal{X}$ such that

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ and } \mathrm{Vol}_{\mathcal{X}}(S(\boldsymbol{q}_0, \varepsilon_0)) \geqslant \alpha \check{g}^{(\zeta)}([\boldsymbol{q}_0]_{\mathcal{X}})\varepsilon_0^{\gamma} \tag{2.42}$$

  where $\check{g}^{(\zeta)}$ is the lower ζ-regularized approximation of g.

- A set of s^{γ} disjoint ζ-regular hierarchical cell structures with radius ε_0/s, efficiency parameter α, scaling parameter s, and rooted at $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{s^{\gamma}}$ such that

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bigcup_{j=1}^{s^{\gamma}} S(\boldsymbol{q}_j, \varepsilon_0/s) \quad \text{and} \quad \boldsymbol{q}_j \in \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ for all } j \tag{2.43}$$

  We call these the *sub-SHCS*'s of the original SHCS.

We will sometimes refer to a set of SHCS's as $S_1, \ldots, S_m$; in this case, we say $x \in S_j$ if x is contained in the top-level cell. Note that this is a recursive definition, i.e. that to give $\mathrm{HCS}(\boldsymbol{q}_0, \varepsilon_0, \alpha, s, \zeta)$ we also need to give its components $\mathrm{HCS}(\boldsymbol{q}_j, \varepsilon_0/s, \alpha, s, \zeta)$ for j from 1 to s^{γ}, and then in turn *their* components, and so forth. Thus, to know the hierarchical cell structure at $\boldsymbol{q}_0, \varepsilon_0$ requires knowing a full hierarchy of cells which can be represented as an infinite tree with s^{γ} branches at each node (as we will do when we define the *hierarchical collection problem*). This is generally possible if we have some kind of regular structure to exploit, for instance the rectilinear sets inscribed within reachable sets implied by the Uniform Ball-Box Theorem (see Proposition 4).

The assumption of Jordan-measurability of the cells is meant to exclude contrived cell structures. One important property of Jordan-measurability is that the volume of a Jordan-measurable set is the same as the volume of its interior, a fact which is used in the proof of Lemma 19.

Assumption 10. There is some α and s such that for any $\zeta > 0$, there exists some ε_0 such that at any $x \in \mathcal{X}_f$, there is a ζ-regular SHCS with radius ε_0, efficiency α, and scaling parameter s containing x in its interior.

2.3.1 Existence of Symmetric HCS's

We now show how, under Meta-Assumption 1 (and a few additional assumptions concerning the properties of the function $[\cdot]_{\mathcal{X}}$), we can guarantee the existence of a Symmetric Hierarchical Cell Structure about any given $x \in \mathcal{X}_f$ with $s = 2$ by taking advantage of the Uniform Ball-Box Theorem (see Section 2.2). Since we are dealing with symmetric dynamics in this section, we don't need the nonsymmetric version.

Then we have:

Proposition 4. If Π is symmetric and Meta-Assumption 1 and Meta-Assumption 2 hold then Assumption 10 holds with $s = 2$.

For convenience, we define the *Local Hierarchical Cell Structure*(LHCS), which is a modified version of a SHCS:

Definition 12. A ζ-regular *Local Hierarchical Cell Structure* LHCS$(\boldsymbol{q}_0, \varepsilon_0, \alpha, s, \zeta)$ rooted at $\boldsymbol{q}_0 \in \mathcal{Q}$ with radius $\varepsilon_0 > 0$, efficiency parameter $0 < \alpha \leqslant 1$, and scaling parameter s is a structure consisting of the following elements:

- A Jordan-measurable *cell* $S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \mathcal{X}$ such that

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ and } \mathrm{Vol}_{\mathcal{X}}(S(\boldsymbol{q}_0, \varepsilon_0)) \geqslant \alpha \check{g}^{(\zeta)}(\boldsymbol{q}_0)\varepsilon_0^{\gamma} \tag{2.44}$$

  where $\check{g}^{(\zeta)}$ is the lower ζ-regularized approximation of g.

- A set of s^{γ} disjoint ζ-regular LHCS's with radius ε_0/s, efficiency parameter

α, scaling parameter s, and rooted at $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{s^\gamma}$ such that

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bigcup_{j=1}^{s^\gamma} S(\boldsymbol{q}_j, \varepsilon_0/s) \quad \text{and} \quad \boldsymbol{q}_j \in \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ for all } j \tag{2.45}$$

The difference between Definition 12 and Definition 11 is that the volume of the cell is only expected to be a certain size relative to $\check{g}^{(\zeta)}(\boldsymbol{q}_0)\varepsilon_0^\gamma$ (which bounds the volume of the workspace reachable set at $\boldsymbol{q}_0$ for small ε_0) as opposed to $\check{g}^{(\zeta)}([\boldsymbol{q}_0]_\mathcal{X})\varepsilon_0^\gamma$; that is, the cell is only expected to be big relative to the workspace reachable set from the anchor $\boldsymbol{q}_0$ rather than relative to the largest such cell around $x_0 = [\boldsymbol{q}_0]_\mathcal{X}$.

Note that the condition of having a LHCS around every $\boldsymbol{q}_0 \in \mathcal{Q}$ is actually *stronger* than having a SHCS around every $x_0 \in \mathcal{X}_f$; this is because by the definition $g(x) = \sup_{\boldsymbol{q}:[\boldsymbol{q}]_\mathcal{X}=x} g(\boldsymbol{q})$ we know that we can find $\boldsymbol{q}$ such that $[\boldsymbol{q}]_\mathcal{X} = x_0$ such that $g(\boldsymbol{q})$ is arbitrarily close to $g(x_0)$, and then just use that as our $\boldsymbol{q}_0$ (taking the approximation $\check{}^{(\zeta)}$ doesn't affect this as $\zeta \to 0$ except at a closed set of measure 0, see Section 2.5).

We consider first a stylized case with $\mathcal{Q} = \mathcal{X}$ (for instance, where the Dubins car needs to visit target points where each target point has a random direction in which it needs to be visited), in which case the projection function is the identity. Let us fix some $\boldsymbol{q}_0 \in \mathcal{Q}$ and $\boldsymbol{H} = (h_{I_1}, \ldots, h_{I_d})$ be an adapted frame at $\boldsymbol{q}$ and scale ε where I_1 are elements of the Lie algebra; we let $r_i = |I_i|$ and $r = \sum_{i=1}^d r_i$; since $\mathcal{Q} = \mathcal{X}$, in this case $\gamma = r$ (since the volume of this box is proportional to ε^r as $\varepsilon \to 0$). Then $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, \varepsilon)$ is a box in the coordinate map produced by the adapted frame with sides parallel to the axes and length $2\varepsilon^{r_i}$ along dimension i in the coordinate map. Then there is a neighborhood $\mathcal{R}$ of $\boldsymbol{q}_0$ and some $\varepsilon_0^* > 0$ (we add the $\cdot^*$ to distinguish it from the ε_0 in the SHCS and LHCS definitions) and constants $0 < c < C$ such that for any $\boldsymbol{q}$ in a neighborhood of $\boldsymbol{q}_0$ and any $\varepsilon < \varepsilon_0^*$

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \tag{2.46}$$

(see the Uniform Ball-Box Theorem).

Given $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$, we note that we can cover it with 2^r instances of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c(\varepsilon/2))$, since these are smaller boxes with length $2(c\varepsilon/2)^{r_i} = 2^{-r_i} \cdot 2(c\varepsilon)^{r_i}$, hence it has length

exactly 2^{-r_i} of the length of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ on dimension i. Since $r_i = |I_i|$ is an integer (the length of the Lie bracket I_i), this means 2^{r_i} is also an integer and hence we can divide the length of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ into 2^{r_i} pieces of exactly length 2^{-r_i}. Repeating this for every $i = 1, \ldots, d$ yields a division of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ into exactly $\prod_{i=1}^{d} 2^{r_i} = 2^r = 2^\gamma$ boxes of the form $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c(\varepsilon/2))$. Thus, using $S(\boldsymbol{q}, \varepsilon) = \text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ as our cells, we have our Local Hierarchical Cell Structure in this case; because of the circumscribing box $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon)$, we know that there is a sufficiently small $\varepsilon^* > 0$ such that if the SHCS has scale $\leqslant \varepsilon^*$, the efficiency parameter α (the ratio of the cell volume to the volume of the reachable set it's inscribed in) satisfies $\alpha \geqslant \frac{1}{2}(c/C)^r$.

Remark 9. *The volume of the 'boxes' on the manifold is not the same as the volumes of their representations on the coordinate map; for one thing, the vectors in $\boldsymbol{H}$ may have different magnitudes, and the curvature of the manifold $\mathcal{Q}$ will also have an effect. However, the* ratio *of the volume of* $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon)$ *to the volume of* $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', C\varepsilon)$*, as $\varepsilon \to 0$, approaches $(c/C)^r$. Thus, there is a $\varepsilon^* > 0$ such that for all $\varepsilon \leqslant \varepsilon^*$ and $\boldsymbol{q}'$ in the neighborhood of $\boldsymbol{q}$*

$$\frac{\text{Vol}_{\mathcal{Q}}(\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon))}{\text{Vol}_{\mathcal{Q}}(\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon))} \geqslant \frac{1}{2}(c/C)^r \tag{2.47}$$

However, we also want to extend this to cases where $\mathcal{X} \neq \mathcal{Q}$; although the same cell structure will achieve covering (just project all cells down to the workspace), the issue is that when $\mathcal{X} \neq \mathcal{Q}$ we get $\gamma < r$ and hence covering with 2^r sub-cells is too many. Instead we will choose a size-2^γ subset of the configuration space sub-cells whose projections onto $\mathcal{X}$ will cover the projection of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ (up to an approximation factor which decreases as $\varepsilon \to 0$).

As discussed in Section 1.2, we assume that $[\cdot]_{\mathcal{X}}$ locally behaves as a projection from a (d-dimensional) Euclidean space to a (d'-dimensional) subspace. Let us fix some $\boldsymbol{q} \in \mathcal{Q}$, and let $x = [\boldsymbol{q}]_{\mathcal{X}}$. We can now treat a sufficiently small neighborhood $\mathcal{V}$ of $\boldsymbol{q}$ and its projection $[\mathcal{V}]_{\mathcal{X}}$ onto $\mathcal{X}$ as both linear (using $\boldsymbol{H}$ as our basis in $\mathcal{Q}$). We note for any $\boldsymbol{q}' \in \mathcal{V}$ and any sufficiently small ε, by the Uniform Ball-Box Theorem we can inscribe $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon)$ in $R_\varepsilon(\boldsymbol{q}')$; as before, we can divide this into 2^r sub-boxes of

the form $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}'', c(\varepsilon/2))$. We now consider the projection of these boxes onto $\mathcal{X}$: we will show that an appropriately-chosen set of 2^γ of the 2^r sub-boxes (with a tiny bit of scaling up, which will diminish as $\varepsilon \to 0$) will cover the projection $[\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon)]_{\mathcal{X}}$.

A vector field h on $\mathcal{Q}$ produces for any $\boldsymbol{q} \in \mathcal{Q}$ a tangent vector $h(\boldsymbol{q}) \in T_{\boldsymbol{q}}\mathcal{Q}$. Taking the projection to $\mathcal{X}$ (and letting $x = [\boldsymbol{q}]_{\mathcal{X}}$), we get a tangent vector on $[h(\boldsymbol{q})]_{\mathcal{X}} \in T_x\mathcal{X}$ (moving along $h(\boldsymbol{q})$ from $\boldsymbol{q} \in \mathcal{Q}$ projects to a movement along $[h(\boldsymbol{q})]_{\mathcal{X}}$ from $x = [\boldsymbol{q}]_{\mathcal{X}} \in \mathcal{X}$). We note that $\boldsymbol{H}(\boldsymbol{q}) = \{h_{I_1}(\boldsymbol{q}), \ldots, h_{I_d}(\boldsymbol{q})\}$ is a basis of $T_{\boldsymbol{q}}\mathcal{Q}$; therefore

$$[\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{[h_{I_1}(\boldsymbol{q})]_{\mathcal{X}}, \ldots, [h_{I_d}(\boldsymbol{q})]_{\mathcal{X}}\} \tag{2.48}$$

spans $T_x\mathcal{X}$. Let us denote $\bar{h}_i(\boldsymbol{q}) := [h_{I_i}(\boldsymbol{q})]_{\mathcal{X}}$ and denote

$$\bar{\boldsymbol{H}}(\boldsymbol{q}) := [\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{\bar{h}_1(\boldsymbol{q}), \ldots, \bar{h}_d(\boldsymbol{q})\}\,. \tag{2.49}$$

For any $A \subseteq [d]$, we let $\boldsymbol{H}_A := \{h_{I_i} : i \in A\}$ (and $\boldsymbol{H}_A(\boldsymbol{q}) := \{h_{I_i}(\boldsymbol{q}) : i \in A\}$) and likewise $\bar{\boldsymbol{H}}_A := \{\bar{h}_i(\boldsymbol{q}) : i \in A\}$.

Remark 10. *It's important to remember that, despite the notation, $\bar{h}_i(\boldsymbol{q})$ is a tangent vector of $T_x\mathcal{X}$ at $x = [\boldsymbol{q}]_{\mathcal{X}}$; the $\boldsymbol{q}$ is there to tell us which tangent vector. Likewise, $\bar{\boldsymbol{H}}(\boldsymbol{q})$ is a collection of d tangent vectors in $T_x\mathcal{X}$.*

We then define for any $A \subset [d]$ such that $|A| = d'$ the following for any $\boldsymbol{q}' \in \mathcal{V}$ and $x' = [\boldsymbol{q}']_{\mathcal{X}}$:

$$\mathrm{Box}_{\boldsymbol{H}_A}(\boldsymbol{q}', \varepsilon) = \{\exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}') : |z_i| \leqslant \varepsilon^{|I_i|} \ \ \forall i \in A \text{ and } z_i = 0 \ \ \forall i \notin A\} \tag{2.50}$$

This is like $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ but we may only use vector fields in $\boldsymbol{H}_A$; note that $\mathrm{Box}_{\boldsymbol{H}_A}(\boldsymbol{q}', \varepsilon) \subseteq \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$. Let us also denote $r_A := \sum_{i \in A} r_i$. Then, when ε is sufficiently small, if $\bar{\boldsymbol{H}}_A$ is linearly independent,

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}_A}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) = \Theta(\varepsilon^{r_A})\,. \tag{2.51}$$

We then consider another stylized case where for any size-d' subset $A \subseteq [d]$, $\bar{\boldsymbol{H}}_A(\boldsymbol{q})$ is linearly independent. Since our vector fields and projection function are smooth, this property will hold in some neighborhood around $\boldsymbol{q}$; we will assume that it holds in the neighborhood $\mathcal{V}$ discussed above (otherwise, just take the intersection of $\mathcal{V}$ with the neighborhood for which this holds).

Remark 11. *Note that this stylized case doesn't always apply: for example, the Reeds-Shepp car on $\mathbb{R}^2$ does not satisfy it, since the vector field $h_2 = [0; 0; 1]$ (note that $h_1 = [\cos(\theta), \sin(\theta), 0]$) which is controlled to steer the vehicle projects to nothing on $\mathcal{X} = \mathbb{R}^2$ (i.e. it is not linearly independent of anything else). Traveling along this vector field rotates the vehicle, which does not affect its location in $\mathcal{X} = \mathbb{R}^2$.*

In this case, without loss of generality, let $r_1 \leqslant \cdots \leqslant r_d$. Then we let $A^* = [d']$, which clearly satisfies $A^* = \arg\min_A r_A$. We first claim that $r_{A^*} = \gamma$.

For any $A \subseteq [d]$ such that $|A| = d'$, consider a sequence $\boldsymbol{a} = (a_i : i \in [d] \backslash A)$ where $a_i \in \{-1, 1\}$ for all $i \in [d] \backslash A$; that is, we assign -1 or 1 for every index *not* in A. Then we define:

$$\mathrm{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon) = \tag{2.52}$$

$$\{\exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}') : |z_i| \leqslant \varepsilon^{|I_i|} \ \ \forall i \in A, z_i = a_i \varepsilon^{|I_i|} \ \ \forall i \notin A\} \tag{2.53}$$

This denotes a d'-dimensional facet of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ whose dimensions run along h_{I_i} for $i \in A$; the sequence $\boldsymbol{a}$ denotes which facet (i.e. for the remaining dimensions, which side of the box is it on). Note that there are a fixed number $2^{d-d'}$ of these facets for each A, and a fixed number $\binom{d}{d'}$ of size-d' sets A, hence a total of $2^{d-d'}\binom{d}{d'}$ such facets. We now note that

$$[\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}} = \bigcup_{A \subseteq [d] : |A| = d'} \bigcup_{\boldsymbol{a} \in \{-1,1\}^{[d] \backslash A}} [\mathrm{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}} \tag{2.54}$$

which is just saying that the projection of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ onto $\mathcal{X}$ is the same as the

union of the projections of all its dimension-d' facets onto $\mathcal{X}$. This means

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \leqslant \sum_{A \subseteq [d]: |A| = d'} \sum_{\boldsymbol{a} \in \{-1,1\}^{[d] \setminus A}} \mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \tag{2.55}$$

$$= \Theta(\varepsilon^{r_{A^*}}) \tag{2.56}$$

since it is a sum of (a bounded number of) terms of order ε^{r_A} for various A, and $r_{A^*} = \min_A r_A$. Furthermore, since $\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}', \varepsilon) \subseteq \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$, we have

$$\Theta(\varepsilon^{r_{A^*}}) = \mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \leqslant \mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \tag{2.57}$$

and hence we have

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) = \Theta(\varepsilon^{r_{A^*}}) \tag{2.58}$$

But since the small-time constraint factor was defined as the value γ satisfying

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) = \Theta(\varepsilon^{\gamma}) \tag{2.59}$$

so $\gamma = r_{A^*}$. This is not surprising as both reflect the volume of $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ projected on $\mathcal{X}$.

We now take a cell in our Local Hierarchical Cell Structure about $\boldsymbol{q}'$ at scale ε to be $S(\boldsymbol{q}', \varepsilon) = [\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}$. We check that it is indeed a LHCS (with scaling factor 2). We can subdivide $\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}', \varepsilon)$ into $2^{r_{A^*}} = 2^{\gamma}$ sub-boxes $\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}'', \varepsilon/2)$ in the same way as the Uniform Ball-Box Theorem does with $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$; the projections of these sub-boxes then become the sub-cells. One thing is that it may not have a good efficiency parameter α and it may be possible to do much better; but we are only proving the existence of a LHCS here.

What happens when $\bar{\boldsymbol{H}}_A(\boldsymbol{q})$ is not linearly independent for all A? It is still possible

that an A' such that $\bar{\boldsymbol{H}}_{A'}(\boldsymbol{q})$ is not independent can still have

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}_{A'}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) > 0 \tag{2.60}$$

though the volume must be $o(\varepsilon^{r_{A'}})$. However, if it has volume, it is only through a Lie bracket interaction; thus, it has volume $\Theta(\varepsilon^{r'_{A'}})$ where $r'_{A'} > r_{A'}$ is an integer, and

$$\min(r_A : A \subseteq [d] \text{ s.t. } |A| = d' \text{ and } \bar{\boldsymbol{H}}_A(\boldsymbol{q}') \text{ is linearly independent}) \leqslant r'_{A'}\,. \tag{2.61}$$

Therefore, we may choose

$$A^* := \arg\min(r_A : A \subseteq [d] \text{ s.t. } |A| = d' \text{ and } \bar{\boldsymbol{H}}_A(\boldsymbol{q}') \text{ is linearly independent}) \tag{2.62}$$

Then the arguments from above (particularly equations (2.54) through (2.59)) still hold, showing that $r_{A^*} = \gamma$. We can then use $S(\boldsymbol{q}', \varepsilon) = [\mathrm{Box}_{\boldsymbol{H}_{A^*}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}$ as our cell, as before.

Then, almost everywhere in $\mathcal{X}_f$, for any $\delta > 0$ we can choose $\boldsymbol{q}$ such that $[\boldsymbol{q}]_{\mathcal{X}} = x$ and $\check{g}^{(\zeta)}(\boldsymbol{q}) \geqslant (1-\delta)\check{g}^{(\zeta)}(x) \geqslant (1-\delta)g_{\min}$; the Local HCS around $\boldsymbol{q}$ is then a SHCS around x. There exists a efficiency parameter $\alpha > 0$ which, for sufficiently small ε, holds for all the sub-SHCS's because $\check{g}^{(\zeta)}(\boldsymbol{q})$ is bounded away from 0 and is Lipschitz continuous, i.e. on small scales it can be treated as constant.

2.4 Nonsymmetric Hierarchical Cell Structures

In the same vein as the Symmetric HCS, we define a Nonsymmetric Hierarchical Cell Structure on which we can build a DSTSP algorithm for nonsymmetric vehicles. In this section, for simplicity, we denote the drift vector field by $h_0 := h_0^{\mathrm{cont}}$ and the control vector fields by $h_j := h_j^{\mathrm{cont}}$ for $j = 1, \ldots, m$.

Recall as well that if we have a vector field h on $\mathcal{Q}$, we denote the *flow* along h by time t as $\exp(t\,h)(\boldsymbol{q})$ (the point $\boldsymbol{q}'$ produced by moving along h from $\boldsymbol{q}$ for t time).

2.4.1 Deviation sets

While a tile inscribed within a reachable set is sufficient for the case of symmetric dynamics, it is difficult to use such cells with nonsymmetric dynamics because the trajectory of the vehicle as it leaves a cell may be constrained by the location of the target point that it visits; this can cause a chain of increasingly complicated effects (it can change what cell is visited next or what target point within that cell is visited next, which causes more changes, and so forth) which make the formulation and analysis of such an algorithm very difficult.[3]

To solve this, we define *deviation sets*: the basic idea is given $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$ where $\boldsymbol{q}' \in R_\varepsilon(\boldsymbol{q})$, the deviation set is the set of configurations $\boldsymbol{q}'$ for which $\boldsymbol{q} \to \boldsymbol{q}'' \to \boldsymbol{q}'$ is possible in $\leqslant 2\varepsilon$ time. Then, inscribing a cell into the deviation set will allow the vehicle to arrive at $\boldsymbol{q}$, visit at target point in the cell, and then 'reset' to $\boldsymbol{q}'$ (which will lead into another cell, and so forth); thus we will have to show that even though the deviation sets can be smaller than reachable sets[4], they still contain cells of volume $\propto \varepsilon^\gamma$. However, given $\boldsymbol{q}$, the configuration $\boldsymbol{q}'$ must be chosen carefully, as it's possible even if $\boldsymbol{q}' \in R_\varepsilon(\boldsymbol{q})$ for the deviation set to have zero volume.

Given control law $h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u})$, which returns the instantaneous velocity (in the configuration space) where $\boldsymbol{u}$ is the control input and $\boldsymbol{q}$ is the current configuration, we define $h_{\boldsymbol{u}}$ to be the vector field resulting from putting in a (constant) $\boldsymbol{u} \in \mathcal{U}$. Note that when $h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{u})$ is control-affine, we have $h^{\text{cont}}(\boldsymbol{q}, \boldsymbol{0}) = h_0(\boldsymbol{q})$ and hence $h_{\boldsymbol{0}} = h_0$ (putting in a zero input yields the drift vector field).

Then we formally define:

Definition 13. Given $\boldsymbol{q} \in \mathcal{Q}$ and $\varepsilon > 0$, the *deviation set* anchored at $\boldsymbol{q}$ with scale ε and control $\boldsymbol{u}$ is

$$D_\varepsilon(\boldsymbol{q}, \boldsymbol{u}) := \{\boldsymbol{q}' : d_\Pi(\boldsymbol{q}, \boldsymbol{q}') + d_\Pi(\boldsymbol{q}', \exp(\varepsilon h_{\boldsymbol{u}})(\boldsymbol{q})) \leqslant 2\varepsilon\}. \tag{2.63}$$

[3] The symmetric dynamics algorithm, by contrast, always 'resets' the vehicle back to the anchor of the cell to make its behavior predictable.

[4] Note that reachable sets are constrained at one end but deviation sets are constrained at both ends

As with reachable sets, we denote the projection to the workspace by $\bar{\cdot}$:

$$\bar{D}_\varepsilon(\boldsymbol{q}, \boldsymbol{u}) := [D_\varepsilon(\boldsymbol{q}, \boldsymbol{u})]_{\mathcal{X}} . \tag{2.64}$$

We refer to this as the *workspace deviation set*.

When Π is control-affine we denote

$$D_\varepsilon(\boldsymbol{q}) := D_\varepsilon(\boldsymbol{q}, \mathbf{0}) = \{\boldsymbol{q}' : d_\Pi(\boldsymbol{q}, \boldsymbol{q}') + d_\Pi(\boldsymbol{q}', \exp(\varepsilon h_0)(\boldsymbol{q})) \leqslant 2\varepsilon\} \tag{2.65}$$

$$\text{and } \bar{D}_\varepsilon(\boldsymbol{q}) := \bar{D}_\varepsilon(\boldsymbol{q}, \mathbf{0}) \tag{2.66}$$

i.e. for control-affine vehicles if the control is not specified then the deviation set takes the input $\boldsymbol{u} = \mathbf{0}$ by default.

As with the reachable sets, we are interested in the volume of the workspace deviation sets. However, it is possible at any sufficiently small scale ε to pick $\boldsymbol{q}, \boldsymbol{u}$ such that $D_\varepsilon(\boldsymbol{q}, \boldsymbol{u})$ has zero volume; for instance, if the vehicle is the Dubins car with $\mathcal{U} = [-1, 1]$ and $\boldsymbol{u} = -1$ i.e. maximum turning to the left (see Example 1), then it is impossible to reach $\exp(\varepsilon h_{\boldsymbol{u}})(\boldsymbol{q})$ from $\boldsymbol{q}$ in $\leqslant 2\varepsilon$ time without just following $\boldsymbol{u} = -1$ and hence the deviation set (and workspace deviation set) has volume 0. However, as mentioned, following control $\boldsymbol{u} = 0$ (as in (2.65)) yields a workspace deviation set of the correct volume.

As we will see in Section 2.4.3, when Meta-Assumption 1 and the Nonsymmetric Uniform Ball-Box Conjecture hold for Π the workspace deviation sets contain (and are contained by) boxes whose sizes scale like that of workspace reachable sets.

2.4.2 The Nonsymmetric Hierarchical Cell Structure

As noted in Section 1.3, our upper bound results for nonsymmetric vehicles are focused entirely on showing the correct order-of-growth of the DSTSP shortest tour; hence we can be looser with the constant factors and in particular will discard any reference to the agility function g (which remains important for the lower bound in the nonsymmetric case). We also will remove the scaling factor s and only define it for

$s = 2$; this suffices for dynamics satisfying Meta-Assumption 1 (if the Nonsymmetric Uniform Ball-Box Conjecture holds) thanks to Lemma 8.

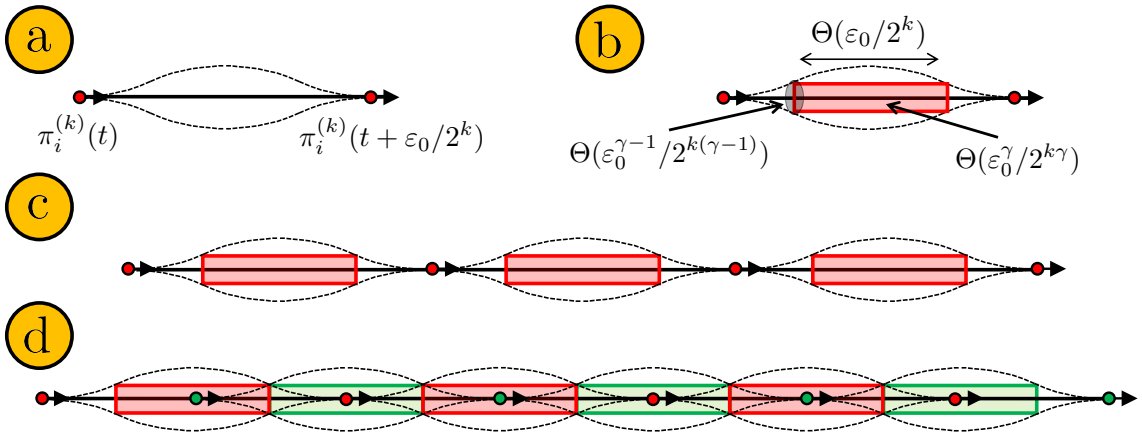

Figure 2-2: Deviation sets and cells along a track $\pi_i^{(k)}$. (a) A single deviation set built around track $\pi_i^{(k)}$; (b) the cell inscribed in that set; (c) a string of cells which can be visited in sequence; (d) multiple strings of cells which together form an unbroken row around $\pi_i^{(k)}$. Note that multiple runs along the track are needed to visit every cell (one for the red and one for the green). These tracks can then be stacked to cover a Nonsymmetric Hierarchical Cell Structure.

Definition 14. A ε_0-scale *Nonsymmetric Hierarchical Cell Structure* (Nonsymmetric HCS or NHCS) with structural constants $\alpha_1, \alpha_2 > 0$ is a structure consisting of an infinite sequence of *levels* $k = 0, 1, \ldots$. Each level k then consists of the following elements:

- $2^{k\gamma}$ *cells* which are subsets of the workspace

$$S_1^{(k)}, \ldots, S_{2^{k\gamma}}^{(k)} \subseteq \mathcal{X}\,. \tag{2.67}$$

  We denote the set of level-k cells as $\mathcal{S}^{(k)}$.

- $2^{k(\gamma-1)}$ *tracks*, which are trajectories

$$\pi_1^{(k)}, \ldots, \pi_{2^{k(\gamma-1)}}^{(k)} : [0, \alpha_1] \to \mathcal{Q} \tag{2.68}$$

where each $\pi_i^{(k)}$ is associated with some control $\boldsymbol{u}_i^{(k)} \in \mathcal{U}$ and some starting configuration $\boldsymbol{q}_i^{(k)}$ and satisfies

$$\pi_i^{(k)}(t) := \exp(t\, h_{\boldsymbol{u}_i^{(k)}})(\boldsymbol{q}_i^{(k)}) \text{ for all } k \text{ and } i \in 1, 2, \ldots, 2^{k(\gamma-1)}\,. \tag{2.69}$$

This means that $\pi_i^{(k)}$ is the trajectory produced when applying control $\boldsymbol{u}_i^{(k)}$ starting from $\boldsymbol{q}_i^{(k)}$ for α_1 time.

These then need to satisfy the following relationships with each other:

- For each k, the level-k cells are all disjoint.
- For each k, every level-k cell $S_i^{(k)}$ is partitioned into 2^γ level-$(k+1)$ cells which we call its *sub-cells*.
- For each k, every level-k track $\pi_i^{(k)}$ is associated with 2^k different level k cells; WLOG we can assign the indices of the cells associated with $\pi_i^{(k)}$ as $S_{j;i}^{(k)}$ where $j = 1, 2, \ldots, 2^k$. We denote the set of all cells associated with track $\pi_i^{(k)}$ as

$$\mathcal{S}_i^{(k)} := \{S_{j;i}^{(k)} : j \in [2^k]\}\,. \tag{2.70}$$

- For each $S \in \mathcal{S}_i^{(k)}$, there is some $t \in [0, \alpha_1]$ such that

$$S \subseteq \bar{D}_{\varepsilon_0/2^k}(\pi_i^{(k)}(t); \boldsymbol{u}_i^{(k)})\,. \tag{2.71}$$

- Each $S \in \mathcal{S}^{(k)}$ at level k has volume

$$\mathrm{Vol}_{\mathcal{X}}(S) \geqslant \alpha_2 \varepsilon_0^\gamma / 2^{k\gamma}\,. \tag{2.72}$$

and the single level-0 cell has nonzero volume.

We remark that ε_0 is the scale, as in the Symmetric HCS (the top level cell is contained in a deviation set of radius ε_0, and the volume of level-k cells is generally

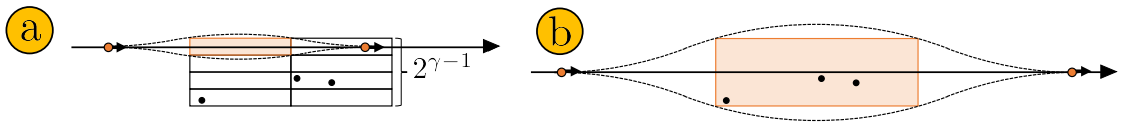

Figure 2-3: (a) 2^γ cells at level k combine to form (b) a cell at level $k-1$; black dots represent (unvisited) targets. Note that most of the cells at level k are empty, but the level-$(k-1)$ cell is not.

proportional to $\varepsilon_0/2^{k\gamma}$); the structural constant α_1 is a bound on the lengths of the tracks; and the structural constant α_2 is a bound on the constant factor of the volume of the cells at level k when attached to $\varepsilon_0/2^k$. Cells and deviation sets along a track (for the example of a Dubins car) are illustrated in Section 2.4.2, and how 2^γ cells at level k combine to form a cell at level $k-1$ is illustrated in Section 2.4.2.

Finally, we make an assumption analogous to Assumption 10:

Assumption 11. For any $x \in \mathcal{X}_f$, there is some Nonsymmetric HCS containing x in its interior.

The idea of the Nonsymmetric HCS is to permit a *pass* at any level k: a trajectory through the NHCS which takes $\propto 2^{k(\gamma-1)}$ time which can visit a target point in every (nonempty) cell. Then the vehicle can make a sequence of passes (not necessarily all at the same level) until all the targets within the NHCS are visited, and move on to the next NHCS. The main tradeoff to consider in this structure is the level k at which to make passes: since we visit $2^{k\gamma}$ cells (and thus visit potentially $2^{k\gamma}$ targets) in time only $2^{k(\gamma-1)}$, increasing k can make the trajectory more efficient, visiting more targets per time spent; however, when k is too large, most cells will be empty (for instance, if $2^{k\gamma} > n$ then by the Pigeonhole Principle at least one cell will be empty, and if the targets are random likely many more cells will be empty). Thus, picking the correct k for each pass is critical to finding an efficient tour. We call this problem the *Nonsymmetric Hierarchical Collection Problem* (NHCP), and discuss it in Chapter 5.

2.4.3 Nonsymmetric HCS from Geometry

In this section we assume Meta-Assumption 1; we will also exclusively work with deviation sets under the $\mathbf{0}$ input, i.e. defined by (2.65); thus our deviation sets depend significantly on the drift. We also assume that the Nonsymmetric Uniform Ball-Box Conjecture holds, at least for our dynamic constraints Π. Finally, we make an additional assumption that the drift moves the vehicle by some minimum speed in the workspace:

Assumption 12. There is some constant $c_{\min} > 0$ such that for all $x \in \mathcal{X}_f$,

$$\sup_{\boldsymbol{q} \in [x]^{\mathcal{Q}}} \|[h_0(\boldsymbol{q})]_{\mathcal{X}}\|_{\mathcal{X}} \geqslant c_{\min}\,. \tag{2.73}$$

That is, for any $x \in \mathcal{X}_f$, there is some $\boldsymbol{q}$ mapping to x such that h_0 moves the vehicle at a rate $\geqslant c_{\min}$ in metric $d_{\mathcal{X}}$. Combined with the vector fields being smooth (Assumption 1), this means that for any $x \in \mathcal{X}_f$, we can find some $\boldsymbol{q} \in [x]^{\mathcal{Q}}$ (i.e. a configuration corresponding to the location of x in the workspace) and some neighborhood $\mathcal{V}$ of $\boldsymbol{q}$ in $\mathcal{Q}$ such that for $\boldsymbol{q}' \in \mathcal{V}$,

$$\|[h_0(\boldsymbol{q}')]_{\mathcal{X}}\|_{\mathcal{X}} \geqslant c_{\min}/2\,. \tag{2.74}$$

This allows us to construct the deviation sets below with some minimum speed, so in our algorithm (which strings deviation sets together) we are guaranteed to cross $\mathcal{X}_f$ a bounded amount of time.

Remark 12. *If Assumption 12 doesn't hold but there is some $\boldsymbol{u}^*$ in the interior of $\mathcal{U} \subseteq \mathbb{R}^m$ for which $\sup_{\boldsymbol{q} \in [x]^{\mathcal{Q}}} \|[h_{\boldsymbol{u}^*}(\boldsymbol{q})]_{\mathcal{X}}\|_{\mathcal{X}} \geqslant c_{\min}$, we can do the following: we restrict our controls to some neighborhood $\mathcal{U}^* \subseteq \mathcal{U}$ of $\mathcal{U}$ which is radially symmetric about $\boldsymbol{u}^*$ and then subtract $\boldsymbol{u}^*$ from $\mathcal{U}^*$ to get $\mathcal{U}' := \{\boldsymbol{u} : \boldsymbol{u} + \boldsymbol{u}^* \in \mathcal{U}^*\}$. Then the dynamics Π are adjusted so that control $\boldsymbol{u} \in \mathcal{U}'$ yield the same results as $\boldsymbol{u} + \boldsymbol{u}^* \in \mathcal{U}^*$, i.e.*

$$\dot{\boldsymbol{q}} = h_0(\boldsymbol{q}) + \sum_{j=1}^{m} h_j(\boldsymbol{q})(u_j + u_j^*) = \left(h_0(\boldsymbol{q}) + \sum_{j=1}^{m} h_j(\boldsymbol{q}) u_j^*\right) + \sum_{j=1}^{m} h_j(\boldsymbol{q}) u_j \tag{2.75}$$

i.e. we replace the drift vector field h_0 with $h_0 + \sum_{j=1}^{m} h_j u_j^*$. By the properties of the Lie bracket the resulting Lie algebra is unchanged.

To analyze the properties of the deviation set, we define for any control-affine dynamics (satisfying Definition 2) the *reverse reachable set*, i.e. the set of configurations from which $\boldsymbol{q}$ is reachable:

$$\overleftarrow{R}_\varepsilon(\boldsymbol{q}) := \{\boldsymbol{q}' \in \mathcal{Q} : d_\Pi(\boldsymbol{q}', \boldsymbol{q}) \leqslant \varepsilon\}, \tag{2.76}$$

or, equivalently, $\overleftarrow{R}_\varepsilon(\boldsymbol{q}) = \{\boldsymbol{q}' : \boldsymbol{q} \in R_\varepsilon(\boldsymbol{q}')\}$. Note that if Π is symmetric we have $R_\varepsilon(\boldsymbol{q}) = \overleftarrow{R}_\varepsilon(\boldsymbol{q})$; however, the distinction is important for nonsymmetric dynamics. One very notable property of the reverse reachable set $\overleftarrow{R}_\varepsilon(\boldsymbol{q})$ is, under the Nonsymmetric Uniform Ball-Box Conjecture, the boxes containing and contained within it are the same general dimensions (i.e. the dimensions scale the same way relative to ε) as those in the reachable set $R_\varepsilon(\boldsymbol{q})$; this is because it can be viewed as the reachable set of the dynamics with vector fields $-h_0, h_1, \ldots, h_m$, i.e. Π with the drift vector field reversed, and since the Lie algebra contains the same vectors either way (with some extra minus signs, but this doesn't affect any of the spanned spaces) the boxes have the same scaling laws.

While the deviation set is more constrained than a reachable set (note that a reachable set is constrained at the beginning, but a deviation set is constrained both at the beginning and at the end), it satisfies the following relationships to the reachable set and reverse reachable set:

Lemma 7. *If $\mathcal{Q}^* \subseteq \mathcal{Q}$ and $\varepsilon_0 > 0$ are defined as in the Nonsymmetric Uniform Ball-Box Conjecture, for any $\boldsymbol{q} \in \mathcal{Q}^*$ and $0 < \varepsilon \leqslant \varepsilon_0$ the following holds:*

$$R_\varepsilon(\boldsymbol{q}) \cap \overleftarrow{R}_\varepsilon(\exp(\varepsilon h_0)(\boldsymbol{q})) \subseteq D_\varepsilon(\boldsymbol{q}) \subseteq R_{2\varepsilon}(\boldsymbol{q}). \tag{2.77}$$

Proof. The first inclusion holds because $\boldsymbol{q}' \in R_\varepsilon(\boldsymbol{q}) \cap \overleftarrow{R}_\varepsilon(\exp(\varepsilon h_0)(\boldsymbol{q}))$ implies that $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon$ (since it's in the reachable set) and $d_\Pi(\boldsymbol{q}', \exp(\varepsilon h_0)(\boldsymbol{q})) \leqslant \varepsilon$ (since it's

in the reverse reachable set of $\exp(\varepsilon h_0)(\boldsymbol{q})$); hence by definition

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') + d_\Pi(\boldsymbol{q}', \exp(\varepsilon h_0)(\boldsymbol{q})) \leqslant 2\varepsilon \tag{2.78}$$

and hence $\boldsymbol{q}' \in D_\varepsilon(\boldsymbol{q})$ by definition.

The second inclusion holds because if $\boldsymbol{q}' \in D_\varepsilon(\boldsymbol{q})$, then

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant d_\Pi(\boldsymbol{q}, \boldsymbol{q}') + d_\Pi(\boldsymbol{q}', \exp(\varepsilon h_0)(\boldsymbol{q})) \leqslant 2\varepsilon \tag{2.79}$$

and hence by definition $\boldsymbol{q}' \in R_{2\varepsilon}(\boldsymbol{q})$. □

Lemma 7 then implies, if the Nonsymmetric Uniform Ball-Box Conjecture holds, that the deviation sets contain the relevant boxes:

Lemma 8. *If $\mathcal{Q}^* \subseteq \mathcal{Q}$ is compact and the Nonsymmetric Uniform Ball-Box Conjecture holds, then exist constants $c > 0$ and $\varepsilon_0 > 0$ such that, for $\boldsymbol{q} \in \mathcal{Q}^*$ and $0 < \varepsilon \leqslant \varepsilon_0$ and any adapted frame $\boldsymbol{H} = (h_{I_1}, \ldots, h_{I_d})$ (a set of elements of the Lie algebra of the control vector fields) at $(\boldsymbol{q}, \varepsilon)$,*

$$\mathrm{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), c\varepsilon) \subseteq D_\varepsilon(\boldsymbol{q})\,. \tag{2.80}$$

Proof. This follows from Lemma 7: from the first inclusion (noting that the adapted frames are the same because they are based on the same Lie algebras) we know that since

$$\exp((\varepsilon/2)h_0)(\boldsymbol{q}) = \exp((\varepsilon/2)(-h_0))(\exp(\varepsilon h_0)\boldsymbol{q}) \tag{2.81}$$

(going forward by ε along h_0 and then forward by $\varepsilon/2$ along $-h_0$ is the same as going forward by $\varepsilon/2$ by h_0), and by the Nonsymmetric Uniform Ball-Box Conjecture there

are $c_1, c_2 > 0$ such that

$$\text{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), c_1\varepsilon) \subseteq R_\varepsilon(\boldsymbol{q}) \tag{2.82}$$

$$\text{and } \text{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), c_2\varepsilon) \subseteq \overleftarrow{R}_\varepsilon(\exp(\varepsilon h_0)(\boldsymbol{q}))\,. \tag{2.83}$$

Thus, letting $c = \min(c_1, c_2) > 0$ yields

$$\text{Box}_{\boldsymbol{H}}(\exp((\varepsilon/2)h_0)(\boldsymbol{q}), c\varepsilon) \subseteq R_\varepsilon(\boldsymbol{q}) \cap \overleftarrow{R}_\varepsilon(\exp(\varepsilon h_0)(\boldsymbol{q})) \subseteq D_\varepsilon(\boldsymbol{q})\,. \tag{2.84}$$

□

We now show that (given the Nonsymmetric Uniform Ball-Box Conjecture) the geometry of control-affine dynamic constraints implies Assumption 11:

Proposition 5. If the dynamic constraints Π satisfy Meta-Assumption 1 and the Nonsymmetric Uniform Ball-Box Conjecture holds for Π, then Assumption 11 holds.

Proof. Let $x_0 \in \mathcal{X}_f$, and let $\varepsilon_0 > 0$ and $0 < c < C$ be as defined in the Nonsymmetric Uniform Ball-Box Conjecture, $\boldsymbol{H}$ be an adapted frame, and $\boldsymbol{q}_0 \in [x_0]^{\mathcal{Q}}$.

Without loss of generality, let $I_1 = [0]$, i.e. h_{I_1} (the Lie bracket associated with I_1) is just the original drift vector h_0 (any maximal linearly-independent subset of the vectors $h_0(\boldsymbol{q}), h_1(\boldsymbol{q}), \ldots, h_m(\boldsymbol{q})$ can be in the adapted frame, since they all come with $|[i]| = 1$ and so without loss of generality we can include h_0). Note that in $\text{Box}_{\boldsymbol{H}}$ the last direction traversed is I_1, i.e. along h_0. Let us denote

$$\boldsymbol{H}' := \boldsymbol{H} \backslash \{h_{I_1}\} = \boldsymbol{H} \backslash \{h_0\} \tag{2.85}$$

i.e. the adapted frame with $h_{I_1} = h_0$ removed. Now it contains $d-1$ vectors and hence $\boldsymbol{H}'(\boldsymbol{q})$ cannot span the d-dimensional tangent space $T_{\boldsymbol{q}}\mathcal{Q}$. We now consider $\text{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)$ in relation to $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$; we note that by definition for any $\boldsymbol{q} \in \text{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$ has a corresponding $\boldsymbol{q}' \in \text{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)$ such that there is some $t \in$

$[-c\varepsilon_0, c\varepsilon_0]$ for which

$$\boldsymbol{q} = \exp(t\, h_0)(\boldsymbol{q}') \tag{2.86}$$

We then pull back $\mathrm{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)$ by $\varepsilon_0 + c\varepsilon_0$; let

$$\mathcal{P} := \{\boldsymbol{q}'' : \boldsymbol{q}'' = \exp(-(\varepsilon_0/2 + c\varepsilon_0/2)\, h_0)(\boldsymbol{q}') \text{ for some } \boldsymbol{q}' \in \mathrm{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)\} \tag{2.87}$$

We now consider the collection of potential 'tracks' which are trajectories starting at some $\boldsymbol{q}'' \in \mathcal{P}$ and then following h_0 for $\varepsilon_0 + c\varepsilon_0$ (we will select a subset of these to be our actual tracks as required in Definition 14). We denote $\alpha_1 = \varepsilon_0 + c\varepsilon_0 = (1+c)\varepsilon_0$ (note that ε_0 is fixed); for any $\boldsymbol{q}'' \in \mathcal{P}$ we let $\pi_{\boldsymbol{q}''} : [0, \alpha_1] \to \mathcal{Q}$ be the trajectory where

$$\pi_{\boldsymbol{q}''}(t) = \exp(t\, h_0)(\boldsymbol{q}'')\,. \tag{2.88}$$

We now consider any $0 < \varepsilon \leqslant \varepsilon_0$ and any $\boldsymbol{q} \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$. Note that by Lemma 8 we have

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subseteq D_\varepsilon(\exp(-(\varepsilon/2)h_0)(\boldsymbol{q}))\,. \tag{2.89}$$

We note that this means that for any $0 < \varepsilon \leqslant \varepsilon_0$ and any $\boldsymbol{q} \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$ there is some $\boldsymbol{q}''$ and $t \in [0, \alpha_1]$ such that

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subseteq D_\varepsilon(\exp(-(\varepsilon/2)h_0(\boldsymbol{q}))) = D_\varepsilon(h_0(\pi_{\boldsymbol{q}''}(t))) \tag{2.90}$$

because we can find $\boldsymbol{q}'', t$ such that $\boldsymbol{q} = \exp((t + \varepsilon/2)h_0)(\boldsymbol{q}'')$. This means that all $c\varepsilon$ boxes with centers $\boldsymbol{q} \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$ are contained in a ε-deviation set of a point on one of the potential tracks starting in $\mathcal{P}$.

We now consider how to make our cells at each level. Note that, as in the Symmetric HCS, our cells are in the workspace and hence we must project these boxes

down into $\mathcal{X}$. For any $\boldsymbol{q} \in \mathcal{Q}$ and $\varepsilon > 0$, we let

$$\overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon) := [\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon)]_{\mathcal{X}} \subseteq \mathcal{X} \tag{2.91}$$

(the projection of the box into the workspace). Note that $\overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}, \varepsilon)$ is not necessarily a box itself. As in Section 2.3.1, at sufficiently small scales we can treat the projection function $[\cdot]_{\mathcal{X}}$ as a linear projection from a higher dimensional space (with the same number of dimensions as $\mathcal{Q}$) to a lower dimensional space (with the same number of dimensions as $\mathcal{X}$). We assume that ε_0 is sufficiently small (WLOG since ε_0 can be made arbitrarily small and still satisfy its definition in the Nonsymmetric Uniform Ball-Box Conjecture).

We then note that for any $\boldsymbol{q}$ and $\varepsilon \leqslant \varepsilon_0$, we can 'pull back' $\boldsymbol{q}$ along the drift vector field h_0 by $\varepsilon/2$ to get

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subseteq D_{\varepsilon}(\exp(-(\varepsilon/2)h_0)(\boldsymbol{q})) \subseteq \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \tag{2.92}$$

$$\text{and hence } \overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subseteq \bar{D}_{\varepsilon}(\exp(-(\varepsilon/2)h_0)(\boldsymbol{q})) \subseteq \overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \tag{2.93}$$

Note that $\overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$ is not necessarily a box thanks to the projection.

We will use $S^{(0)}_{1;1} \subseteq \overline{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$ (the top-level cell of our NHCS, and since it's the only cell on the only track at level 0 it gets $1;1$ in the subscript), and our top-level track is $\pi^{(0)}_1 = \exp(t\, h_0)(\boldsymbol{q}^{(0)}_1)$ where $\boldsymbol{q}^{(0)}_1 = \exp(-(\varepsilon_0/2 + c\varepsilon_0/2)\, h_0)(\boldsymbol{q}') \in \mathcal{P}$ (since $\boldsymbol{q}_0 \in \mathrm{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)$). By (2.93) we have the inclusion of the cell within the deviation set (using $t^{(0)}_{1;1} = 0$) that we need.

We now take a similar strategy as in the proof of Proposition 4. We start with a stylized case in which $\mathcal{X} = \mathcal{Q}$ (i.e. the projection is just the identity function). In this case, we use the following system of 'coordinates': let $r_i = |I_i|$, and fix k; we then let $\mathcal{A}^{(k)} := \prod_{i=2}^{d} [2^{kr_i}]$ and $\boldsymbol{a} = (a_2, a_3, \ldots, a_d) \in \mathcal{A}^{(k)}$ (i.e. where $a_i \in [2^{kr_i}]$), and $z_i = (\frac{2a_i - 1}{2^{kr_i}} - 1)(c\varepsilon_0)^{r_i}$ (generalizing from the $k = 1$ case, this cuts the range $[-(c\varepsilon_0)^{r_i}, (c\varepsilon_0)^{r_i}]$ into 2^{kr_i} equal intervals and puts z_i in the center of each interval)

and again let

$$\boldsymbol{q}_{\boldsymbol{a}}^{(k)} := \exp(-(\varepsilon_0/2 + c\varepsilon_0/2)\, h_0) \circ \exp(z_2 h_{I_2}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}_0) \in \mathcal{P}\,. \tag{2.94}$$

It is in $\mathcal{P}$ because we know that

$$\exp(z_2 h_{I_2}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}_0) \in \mathrm{Box}_{\boldsymbol{H}'}(\boldsymbol{q}_0, c\varepsilon_0)\,. \tag{2.95}$$

Note that there are $2^{\sum_{i=2}^d kr_i} = 2^{\sum_{i=1}^d kr_i - k} = 2^{k(\gamma-1)}$ such $\boldsymbol{a}$ (the property that $\gamma = \sum_{i=1}^d r_i$ is unique to the $\mathcal{Q} = \mathcal{X}$ case); thus we can associate each of our $2^{k(\gamma-1)}$ tracks at level k with some a sequence $\boldsymbol{a} \in \mathcal{A}^{(k)}$, and hence we replace the track index (which we called i in Definition 14, not to be confused with the index i in r_i) with $\boldsymbol{a} \in \mathcal{A}^{(k)}$.

We then define our cells iteratively starting with level $k = 0$. We first define 'upper bounds' (supersets) of our cells as follows. We note that if we index the $2^{k(\gamma-1)}$ tracks at level k by $\boldsymbol{a} \in \mathcal{A}^{(k)}$, we can index the $2^{k\gamma}$ cells by pairs $(a_1; \boldsymbol{a})$ where $a_1 \in [2^k]$ (note that by definition $r_1 = |I_1| = 1$ since I_1 denotes the 'Lie bracket' of just h_0) and $\boldsymbol{a} \in \mathcal{A}^{(k)}$ (putting back the first index corresponding to $h_{I_1} = h_0$).

Formally, to get the cell center $\boldsymbol{q}_{a_1;\boldsymbol{a}}^{(k)}$ corresponding to $(a_1; \boldsymbol{a})$ we let $z_1 = (\frac{2a_1-1}{2^k} - 1)(c\varepsilon_0)$ (the same formula as the other a_i, noting that $r_1 = 1$). Then we have

$$\boldsymbol{q}_{a_1;\boldsymbol{a}}^{(k)} := \exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}_0) \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)\,. \tag{2.96}$$

We note that for all $\boldsymbol{q}_{a_1;\boldsymbol{a}}^{(k)}$, we have the following:

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_{a_1;\boldsymbol{a}}^{(k)}, c\varepsilon_0/2^k) \subseteq D_{\varepsilon_0/2^k}(\pi_{\boldsymbol{a}}^{(k)}(t)) \tag{2.97}$$

for some $t \in [0, \alpha_1]$, where $\pi_{\boldsymbol{a}}^{(k)}(t)$ is the track starting at $\boldsymbol{q}_{\boldsymbol{a}}^{(k)}(t)$.

We first let $\widehat{S}_{1;\boldsymbol{a}}^{(0)} = \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$, where $\boldsymbol{a} = (1, 1, \ldots, 1)$ is the only sequence where each a_i is in $[2^{0 \cdot r_i}] = [1]$ (since it is level 0). We let $\mathrm{Ch}(a_1; \boldsymbol{a})$ denote the set of sub-cells (or children) from $(a_1; \boldsymbol{a})$, i.e. the 2^γ level-$(k+1)$ cells which are contained in the level-k cell represented by $(a_1; \boldsymbol{a})$. This happens by representing them by $(a_1'; \boldsymbol{a}')$

where $a_i' \in \{2^{r_i}(a_i - 1) + j : j \in [2^{r_i}]\}$ for all i (including $i = 1$).

Then, starting with $k = 0$ and increasing it, we let δ_k be the smallest constant such that for all $a_1 \in [2^k]$ and $\boldsymbol{a} \in \mathcal{A}^{(k)}$,

$$\widehat{S}^{(k)}_{a_1;\boldsymbol{a}} \subseteq \bigcup_{(a_1';\boldsymbol{a}') \in \mathrm{Ch}(a_1;\boldsymbol{a})} \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}^{(k+1)}_{a_1';\boldsymbol{a}'}, (1 + \delta^{(k)}) c\varepsilon_0 / 2^{k+1}) \tag{2.98}$$

i.e. we take the boxes you would expect but need to expand them a little to ensure we cover the cells from the previous layer. We then let

$$\widehat{S}^{(k)}_{a_1';\boldsymbol{a}'} := \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}^{(k+1)}_{a_1';\boldsymbol{a}'}, (1 + \delta^{(k)}) c\varepsilon_0 / 2^{k+1}) \tag{2.99}$$

for all these boxes. They are not our cells because they overlap somewhat, but overlaps can be removed (while keeping the area of each cell within a constant multiple) to get our cells $S^{(k)}_{a_1;\boldsymbol{a}}$.

However, we note that at each layer k, the multiplicative factor we need to expand by is proportional to $1 + O(2^{-k})$. This is because if we have

$$\boldsymbol{q} = \exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}_0) \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0) \tag{2.100}$$

and then we move within a box centered at $\boldsymbol{q}$, i.e. go to

$$\boldsymbol{q}' = \exp(z_1' h_{I_1}) \circ \cdots \circ \exp(z_d' h_{I_d})(\boldsymbol{q}) \in \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon_0 / 2^k) \tag{2.101}$$

this only differs from

$$\boldsymbol{q}'' = \exp((z_1' + z_1) h_{I_1}) \circ \cdots \circ \exp((z_d' + z_d) h_{I_d})(\boldsymbol{q}_0) \tag{2.102}$$

because of the order in which we execute the flow operations. The error is thus composed of additional Lie brackets of the Lie algebra, which scale down with the box scale. Thus, $1 + \delta^{(k)}$ can be bounded above by an infinite product of terms of the form $(1 + O(2^{-k}))$, which must converge, and hence we can bound all $(1 + \delta^{(k)})$ by a

constant. Thus, there is a constant $c' \geqslant (1+\delta^{(k)})c$ for all k such that

$$S^{(k)}_{a_1;\boldsymbol{a}} \subseteq \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}^{(k)}_{a_1;\boldsymbol{a}}, c'\varepsilon_0/2^{k+1}) \subseteq D_{\alpha_2\varepsilon/2^k}(\pi^{(k)}_{\boldsymbol{a}}(t)) \tag{2.103}$$

for some $t \in [0, \alpha_2]$, where $\alpha_2 = c'/c$, and we are done with the stylized $\mathcal{X} = \mathcal{Q}$ case.

To extend this to the case where $\mathcal{X} \neq \mathcal{Q}$, we have to use the projection function $[\cdot]_{\mathcal{X}}$. We use the following definitions from Section 2.2.1. We let d denote the number of dimensions of $\mathcal{Q}$, and d' denote the number of dimensions of $\mathcal{X}$. Taking the projection to $\mathcal{X}$ (and letting $x(\boldsymbol{q}) = [\boldsymbol{q}]_{\mathcal{X}}$), we get a tangent vector on $[h(\boldsymbol{q})]_{\mathcal{X}} \in T_{x(\boldsymbol{q})}\mathcal{X}$ (moving along $h(\boldsymbol{q})$ from $\boldsymbol{q} \in \mathcal{Q}$ projects to a movement along $[h(\boldsymbol{q})]_{\mathcal{X}}$ from $x(\boldsymbol{q})$). We note that $\boldsymbol{H}(\boldsymbol{q}) = \{h_{I_1}(\boldsymbol{q}), \ldots, h_{I_d}(\boldsymbol{q})\}$ is by definition a basis of $T_{\boldsymbol{q}}\mathcal{Q}$; therefore

$$[\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{[h_{I_1}(\boldsymbol{q})]_{\mathcal{X}}, \ldots, [h_{I_d}(\boldsymbol{q})]_{\mathcal{X}}\} \tag{2.104}$$

spans $T_{x(\boldsymbol{q})}\mathcal{X}$. Let us denote this by

$$\bar{\boldsymbol{H}}(\boldsymbol{q}) := [\boldsymbol{H}(\boldsymbol{q})]_{\mathcal{X}} = \{\bar{h}_{I_1}(\boldsymbol{q}), \ldots, \bar{h}_{I_d}(\boldsymbol{q})\}\,. \tag{2.105}$$

We also let $\triangle^{(j)}_{\boldsymbol{q}}$ denote the subspace of $T_{x(\boldsymbol{q})}\mathcal{X}$ spanned by $\{\bar{h}_{I_i} : |I_i| \leqslant j\}$, and let $\xi^{(j)} = \dim(\triangle^{(j)}_{\boldsymbol{q}})$, which by Meta-Assumption 2 is fixed.

We will be using projections of the same tracks and cells from the $\mathcal{Q} = \mathcal{X}$ case; however, we will only need a subset of them ($2^{k(\gamma-1)}$ tracks and 2^k cells per track) to cover the projections. We do this by the following method:

We consider $k = 0, 1, \ldots$ inductively. For $k = 0$, we have one $\boldsymbol{q}^{(0)}_{\boldsymbol{a}}$ and since a deviation set coming from $\pi^{(0)}_{\boldsymbol{a}}$ contains cell $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$, the corresponding workspace deviation set contains $\bar{\mathrm{Box}}_{\boldsymbol{H}}(\boldsymbol{q}_0, c\varepsilon_0)$, which we use as our cell $S^{(0)}_{a_1;\boldsymbol{a}}$ (recall that there is only one setting of $(a_1; \boldsymbol{a})$ when $k = 0$).

Then, given a track $\pi^{(k)}_{a_1;\boldsymbol{a}}$ at level k, we consider how to find $2^{\gamma-1}$ 'child' tracks $\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}$ at level $k+1$. The requirement is that for any

$$x \in \overline{\mathrm{Box}}_{\boldsymbol{H}}(\pi^{(k)}_{a_1;\boldsymbol{a}}(t), c\varepsilon_0/2^k) \text{ where } t \in [0, \alpha_1] \tag{2.106}$$

there should be some child $\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}$ and $t' \in [0, \alpha_1]$ such that

$$x \in \overline{\mathrm{Box}}_{\boldsymbol{H}}(\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}(t'), c\varepsilon_0/2^{k+1}) \tag{2.107}$$

We note that since we are allowed to adjust t', we can move $\overline{\mathrm{Box}}_{\boldsymbol{H}}(\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}(t'), c\varepsilon_0/2^{k+1})$ along h_0. Hence, we can focus instead on finding for any

$$x \in \overline{\mathrm{Box}}_{\boldsymbol{H}'}(\pi^{(k)}_{a_1;\boldsymbol{a}}(t), c\varepsilon_0/2^{k}) \text{ where } t \in [0, \alpha_1] \tag{2.108}$$

a child $\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}$ and $t' \in [0, \alpha_1]$ such that

$$x \in \overline{\mathrm{Box}}_{\boldsymbol{H}'}(\pi^{(k+1)}_{a'_1;\boldsymbol{a}'}(t'), c\varepsilon_0/2^{k+1})\,. \tag{2.109}$$

This then reduces the problem to how $\mathrm{Box}_{\boldsymbol{H}'}$ behaves under projection, analogous to how $\mathrm{Box}_{\boldsymbol{H}}$ behaves under projection in Proposition 4. Repeating the same analysis (examining the subsets of $\boldsymbol{H}$ whose projections onto $\mathcal{X}$ are linearly independent), we can select $2^{\gamma-1}$ children ($2^{\gamma-1}$ because eliminating h_0 reduces the dimension of the box by 1 since by Assumption 12 the drift vector field's projection onto $\mathcal{X}$ has magnitude bounded away from 0).

Since for any $x_0 \in \mathcal{X}_f$ we can select an appropriate $\boldsymbol{q}_0 \in [x_0]^{\mathcal{Q}}$ and ε_0 and get an NHCS, we are done. □

Remark 13. *Unlike the proof of Proposition 4, we did not need to deal with an analogue of Local Hierarchical Cell Structures in the above proof because we are not concerned with maximizing the agility in any given region (since for nonsymmetric vehicles our upper bound only concerns the order-of-growth).*

2.5 Approximately everywhere

There are certain important cases where the assumptions don't strictly hold over the entire space – for instance, if the parameters of the dynamics have a discontinuity (say, a Dubins car whose turning radius sharply changes when it enters a given re-

gion), Assumption 6 may not hold because $\boldsymbol{q}$ near the boundary may require ε to be arbitrarily small, so no fixed ε^*_ζ will work. In order to deal with this instance, we define the notion of an assumption holding *approximately everywhere* on some subset $\mathcal{X}^* \subseteq \mathcal{X}_f$ or $\mathcal{Q}^* \subseteq \mathcal{Q}$.

Definition 15. An assumption holds *approximately everywhere* on a set A in a space $\mathcal{A}$ with metric $d_\mathcal{A}$ and volume $\mathrm{Vol}_\mathcal{A}$ if, for every $\eta > 0$, there is some A_η such that:

- The assumption holds on A_η.
- $\mathrm{Vol}_\mathcal{A}(A \backslash A'_\eta) \leqslant \eta$ where $A'_\eta := \{a \in \mathcal{A} : \inf_{a' \notin A_\eta} d_\mathcal{A}(a, a') \geqslant \eta\}$ (the η-interior of A_η).

We call A_η the *η-approximator* of A.

In short, the assumption should hold on a set A_η whose η-interior A'_η (the set of points at least η away from points outside the set) takes up most of A, i.e. $A \backslash A'_\eta$ has volume at most η. While we use the same $\eta > 0$ as the depth of the interior and the cap on the volume, they could be separated into $\eta_1, \eta_2 > 0$ and the definition would be equivalent (by using $\eta = \min(\eta_1, \eta_2)$).

Remark 14. *One thing to note is that as $\eta \to 0$, A_η and A'_η will generally expand to more completely fill A, and the condition $A_{\eta_1} \subseteq A_{\eta_2}$ if $\eta_1 \geqslant \eta_2$ can be enforced without affecting the definition. Let's define $A^* = \bigcup_{\eta \to 0} A_\eta$: noting that our assumption in question holds on all A_η for $\eta > 0$, can we say that it must hold on A^*? No – many of our assumptions take the form 'for all $\zeta > 0$, there exists some $\varepsilon^*_\zeta > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^*_\zeta$ a certain condition holds for all a'; however, a different ε^*_ζ may be needed for each given A_η (so really it should be $\varepsilon^*_{\zeta,\eta}$), and no $\varepsilon^*_\zeta > 0$ may work for all $\eta > 0$.*

Remark 14 shows why we need Definition 15 rather than the more common 'almost everywhere'. We then note that we only need our assumptions to hold approximately everywhere for our main results to hold:

Proposition 6. If Proposition 2 and Proposition 3 are true when all the assumptions hold everywhere, they are also true when all the assumptions hold approximately everywhere on the support $\mathcal{X}_f$.

Proof. First, we note that we have a finite number k of assumptions, which we can label as $1, \ldots, k$. Then for $\eta > 0$, let $\mathcal{X}_\eta^{(i)}$ denote the η-approximator of $\mathcal{X}_f$; we then denote $\mathcal{X}_\eta := \bigcap_{i=1}^k \mathcal{X}_\eta^{(i)}$ and $\mathcal{X}_\eta'$ be its η-interior. We claim that:

i. All the assumptions hold over $\mathcal{X}_\eta$.

ii. $\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f \backslash \mathcal{X}_\eta') \leqslant k\eta$.

To show (i), we note that all our assumptions take the form discussed in Remark 14, i.e. 'for all $\zeta > 0$, there is a $\varepsilon_\zeta^* > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon_\zeta^*$, a certain condition holds across the set'. We then fix $\zeta > 0$ and we let $\varepsilon_{\zeta,\eta}^{(i)} > 0$ be the value which corresponds to assumption i over $\mathcal{X}_\eta^{(i)}$; then we let $\varepsilon_{\zeta,\eta}^* = \min_i \varepsilon_{\zeta,\eta}^{(i)}$. In this case, $\varepsilon < \varepsilon_{\zeta,\eta}^* \implies \varepsilon < \varepsilon_{\zeta,\eta}^{(i)}$ and hence for such ε the conditions all hold over their respective $\mathcal{X}_\eta^{(i)}$, and hence hold over $\mathcal{X}_\eta$.

To show (ii), we first show that $\mathcal{X}_\eta' = \bigcap_{i=1}^k \mathcal{X}_\eta^{(i)'}$ where $\mathcal{X}_\eta^{(i)'}$ is the η-interior of $\mathcal{X}_\eta^{(i)}$. This is because

$$x \in \mathcal{X}_\eta' \iff \inf_{x' \notin \mathcal{X}_\eta} d_{\mathcal{X}}(x, x') \geqslant \eta \tag{2.110}$$

$$\iff \inf_{i \in [k]} \inf_{x' \notin \mathcal{X}_\eta^{(i)}} d_{\mathcal{X}}(x, x') \geqslant \eta \tag{2.111}$$

$$\iff x \in \mathcal{X}_\eta^{(i)} \text{ for all } i \in [k] \tag{2.112}$$

$$\iff x \in \bigcap_{i=1}^k \mathcal{X}_\eta^{(i)'} \tag{2.113}$$

Then we can compute:

$$\begin{aligned}
\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f \backslash \mathcal{X}'_\eta) &= \mathrm{Vol}_{\mathcal{X}}\Big(\mathcal{X}_f \backslash \bigcap_{i=1}^{k} \mathcal{X}_\eta^{(i)'}\Big) && (2.114)\\
&= \mathrm{Vol}_{\mathcal{X}}\Big(\bigcup_{i=1}^{k} (\mathcal{X}_f \backslash \mathcal{X}_\eta^{(i)'})\Big) && (2.115)\\
&\leqslant \sum_{i=1}^{k} \mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f \backslash \mathcal{X}_\eta^{(i)'}) && (2.116)\\
&\leqslant k\eta && (2.117)
\end{aligned}$$

Thus, in particular we know that any point $x \in \mathcal{X}'_\eta$ has a η-radius ball around it in which all the assumptions hold and that as $\eta \to 0$, the volume of $\mathcal{X}_f$ that's outside the set $\mathcal{X}'_\eta$ goes to 0, which in turn shows that

$$\lim_{\eta \to 0} \mathbb{P}_{X \sim f}[X \in \mathcal{X}'_\eta] = 1 \text{ and } \lim_{\eta \to 0} \int_{\mathcal{X}'_\eta} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx = \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{2.118}$$

since f represents a continuous probability distribution and g is bounded below.

We now consider our main theorems, and specifically how they are proven.

First, we consider Proposition 2 (the very high probability lower bound to the DSTSP). We then fix $\eta > 0$ and note that for any $X_1, \ldots, X_n$,

$$\mathrm{TSP}_{\Pi}(\{X_1, \ldots, X_n\} \cap \mathcal{X}'_\eta) \leqslant \mathrm{TSP}_{\Pi}(X_1, \ldots, X_n) \tag{2.119}$$

Thus, any very high probability lower bound for the DSTSP where the trajectory only needs to visit target points in $\mathcal{X}'_\eta$ also applies to the original problem. Furthermore, as $\lim_{\eta \to 0} \mathbb{P}_{X \sim f}[X \in \mathcal{X}'_\eta] = 1$, for any $\delta_1 > 0$ we can set η sufficiently small so that

$$\mathbb{P}_{X \sim f}[X \notin \mathcal{X}'_\eta] \leqslant \delta_1/2 \tag{2.120}$$

Then $\{X_1, \ldots, X_n\} \cap \mathcal{X}'_\eta$ will with very high probability have $\geqslant (1 - \delta_1)n$ targets by the Chernoff bound.

Thus, we can modify the problem to distribute $(1-\delta_1)n$ targets over $\mathcal{X}'_\eta$ with density proportional to f; this yields a TSP tour length at most as long (with very high probability) as distributing n targets according to f and keeping only whose which fall in $\mathcal{X}'_\eta$, which in turn is at most as long as having to visit all n targets.

We note that the scaled version of f on $\mathcal{X}'_\eta$ is

$$f|_{\mathcal{X}'_\eta}(x) = \begin{cases} f(x)/\mathbb{P}_{X\sim f}[X \in \mathcal{X}'_\eta] & \text{if } x \in \mathcal{X}'_\eta \\ 0 & \text{if } x \notin \mathcal{X}'_\eta \end{cases} \tag{2.121}$$

and that this value for all $x \in \mathcal{X}'_\eta$ can be bounded as

$$f(x) \leqslant f|_{\mathcal{X}'_\eta}(x) \leqslant (1-\delta_1/2)^{-1} f(x) \tag{2.122}$$

which in turn means that

$$\int_{\mathcal{X}'_\eta} f|_{\mathcal{X}'_\eta}(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \leqslant (1-\delta_1/2)^{-(1-\frac{1}{\gamma})} \int_{\mathcal{X}'_\eta} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{2.123}$$

$$\leqslant (1-\delta_1/2)^{-(1-\frac{1}{\gamma})} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{2.124}$$

We can then turn to the Orienteering problem, in particular the Cost-Balanced Orienteering problem Definition 18 and the very high probability upper bound Proposition 8 we derive for it. We can consider this problem on $(1-\delta_1)n$ random target points for approximation factor δ_2. We fix cost lower bound ζ as in the proof of Proposition 8 and set

$$\lambda < \frac{\eta\zeta}{c_\Pi} \tag{2.125}$$

where c_Π is the speed limit of Π. Then any cost-λ trajectory must have length $\leqslant \lambda/\zeta < \eta/c_\Pi$, which can cover a distance of at most η in the metric on $\mathcal{X}$. But since all targets X_i in this problem are in $\mathcal{X}'_\eta$, which is the η-interior of $\mathcal{X}_\eta$, any trajectory visiting any target point (which are the only ones we care about) must remain entirely in $\mathcal{X}_\eta$, which is exactly where we know our assumptions hold. Therefore, the proof of

Proposition 8 holds and we obtain the bound

$$\mathbb{P}[\mathrm{CBO}_\Pi(X_1, X_2, \dots, X_n; \lambda, f) \leqslant (1+\delta_2)\beta\lambda\,(1-\delta_1)^{\frac{1}{\gamma}} n^{\frac{1}{\gamma}}] \geqslant 1 - e^{-\frac{4}{5}(1+\delta_2)\log(b)\lambda(1\delta_1)^{\frac{1}{\gamma}} n^{\frac{1}{\gamma}}} \tag{2.126}$$

Finally, we apply this to the TSP lower bound in the same way. In this case, instead of visiting n target points, we need to visit $(1-\delta_1)n$, but the argument is the same. Then, by setting δ_1, δ_2 small enough we can achieve an approximation factor δ for the TSP for any $\delta > 0$ with very high probability. □

Chapter 3

DSTSP Lower Bound

In this chapter we show the DSTSP lower bound from Theorem 1, which not only includes the order of growth $\Theta(n^{1-\frac{1}{\gamma}})$ with regard to the number n of target points, but which describes the relationship between the density f of target points, the agility function g of the vehicle, and the small-time constraint factor γ. We do this by showing a very high probability Orienteering upper bound and translating it into a corresponding DSTSP upper bound. One wrinkle, however, is that while the usual Orienteering problem suffices to show the order-of-growth of the DSTSP lower bound, it is not precise enough to give the constant factor demonstrating the effect of density f and agility g; this obstacle has to do with the fact that a TSP solution must visit all the targets, which may be spread throughout $\mathcal{X}_f$, while an Orienteering path can choose to restrict itself to only a very small region within $\mathcal{X}_f$ – presumably one with a high density of target points and/or where the vehicle has better agility. Thus, the DSTSP depends on f and g over all of $\mathcal{X}_f$ while the usual Orienteering essentially depends only on most advantageous or *lucrative* area, in which target points can be visited most rapidly, while f and g elsewhere (with high probability) do not affect the solution at all. This is discussed in Section 3.1.

We thus obtain our DSTSP lower bound with the following steps:

1. Define a variant of the Orienteering problem, which we call *Cost-Balanced Orienteering* (CBO) and which (approximately) balances out the lucrativity over

the whole space (Section 3.2).

2. Show a very high probability upper bound to the CBO using the 'implicit discretization' we developed in [20],[1] followed by the probabilistic technique developed by [11] (Section 3.3).

3. Use the very high probability CBO upper bound to show a corresponding very high probability lower bound for the DSTSP.

3.1 Lucrativity and Cost-Balancing

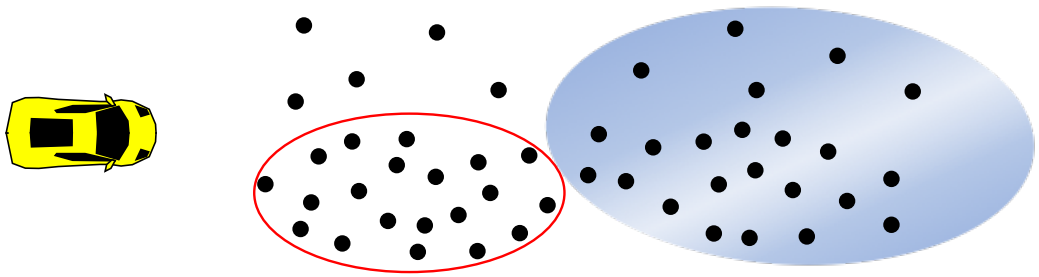

Figure 3-1: Illustration of lucrativity for a Dubins car with a patch of ice on the ground (which reduces turning rate and hence agility); targets are represented by black dots. The circled area is the most lucrative, as it has both a large density of targets and high vehicle agility.

In order to measure exactly how f, g contribute to the 'lucrativity' of a region, we consider the following: for small $\varepsilon > 0$, we have $\mathrm{Vol}_{\mathcal{X}}(\bar{R}_\varepsilon(\boldsymbol{q})) \approx g(\boldsymbol{q})\varepsilon^\gamma$. Thus, we can estimate the probability that any given target falls into $\bar{R}_\varepsilon(\boldsymbol{q})$:

$$\mathbb{P}_{X\sim f}[X \in \bar{R}_\varepsilon(\boldsymbol{q})] \approx f([\boldsymbol{q}]_{\mathcal{X}})g(\boldsymbol{q})\varepsilon^\gamma \tag{3.1}$$

$$\leqslant f([\boldsymbol{q}]_{\mathcal{X}})g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma \tag{3.2}$$

Thus, the expected number of targets that fall within a given radius-ε reachable set

[1] In [20] our focus was on the order-of-growth and hence we didn't need to use cost-balancing. However, the technique can be extended to the CBO, as we will do here.

anchored at $\boldsymbol{q} \in [x]^{\mathcal{Q}}$ can be bounded by

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[|\{X_1, \ldots, X_n\} \cap \bar{R}_\varepsilon(\boldsymbol{q})|] \leqslant n f(x) g(x) \varepsilon^\gamma \,. \tag{3.3}$$

This means that if $\varepsilon = n^{-\frac{1}{\gamma}} (f(x)g(x))^{-\frac{1}{\gamma}}$, the expected number of targets in a ε-radius reachable set is at most 1; we can then take ε to be a rough measure of the average time to reach the nearest target from x, and hence $1/\varepsilon = n^{\frac{1}{\gamma}} (f(x)g(x))^{\frac{1}{\gamma}}$ is roughly the *rate* at which we visit target points in the vicinity of x. Thus, we may define the *lucrativity function* over $\mathcal{X}$ as

$$\text{¢}^*(x) := \text{¢}^*_{f,g}(x) := \big(f(x)g(x)\big)^{\frac{1}{\gamma}} \tag{3.4}$$

which is proportional to the rate at which the vehicle can expect to encounter target points near x (when n is large).

Thus, we want to modify the Orienteering problem to balance out the lucrativity over the whole space and not have it favor any region over any other. Intuitively, this can be done by using lucrativity function as a cost function: the vehicle is 'charged' $\text{¢}^*(x)$ cost per unit length for movement at x. This then means that the rate it can expect to encounter target points is roughly 1 per unit cost, everywhere; we refer to this as the problem being 'balanced'.

Remark 15. *Unfortunately, while this intuition generally holds across the space under a very broad set of conditions, there are many cases of interest in which this will not hold everywhere. In particular,* (3.2) *may fail wherever* $f([\boldsymbol{q}]_{\mathcal{X}}) = 0$ *or* $g([\boldsymbol{q}]_{\mathcal{X}}) = 0$*, or wherever* f *or* g *has a discontinuity. This leads in particular to a problem where movement outside of* $\mathcal{X}_f$ *(i.e. where* $f(x) = 0$*) is 'free' (no cost). This makes the Orienteering problem with cost function* ¢^* *too powerful to provide a useful lower bound to the TSP.*

To deal with the issue raised in Remark 15, we modify the lucrativity function to obtain a cost function for our dynamics:

Definition 16. The *adjusted cost-balancing function* (which we will generally refer to as the *cost function*) of density f and agility g with regularization factor $\zeta > 0$ is $\mathfrak{c} : \mathcal{X} \to \mathbb{R}$ is

$$\mathfrak{c}_\zeta(x) := (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \tag{3.5}$$

i.e. $\mathfrak{c}_\zeta(x)$ is the product of the upper ζ-regularizations of f and g (the minimal upper bounds of f, g which are $\geqslant \zeta$ and $(1/\zeta)$-Lipschitz continuous, see Definition 9).

Then, for a trajectory $\pi \in \Pi$, its cost is

$$\ell^{\mathfrak{c}}(\pi) := \ell^{\mathfrak{c}}_\zeta(\pi) := \int_0^{\ell(\pi)} \mathfrak{c}\big(\pi(t)\big)dt. \tag{3.6}$$

For convenience, we want to be able to reparameterize π by cost. For any $t \in [0, \ell(\pi)]$, we define

$$t^{\mathfrak{c}} := t^{\mathfrak{c}}_\pi := \int_0^t \mathfrak{c}_\zeta(\pi(t'))\, dt' \tag{3.7}$$

and define $\pi^{\mathfrak{c}} : [0, \ell^{\mathfrak{c}}(\pi)] \to \mathcal{Q}$ as the trajectory satisfying

$$\pi^{\mathfrak{c}}(t^{\mathfrak{c}}) = \pi(t)\,. \tag{3.8}$$

We will use the '$\mathfrak{c}$' symbol in general to denote cost-denominated versions of definitions from the previous section, e.g. $\bar{R}^{\mathfrak{c}}_\varepsilon(\boldsymbol{q})$ for the region reachable in $\leqslant \varepsilon$ cost from $\boldsymbol{q}$. In general, the value of ζ will be fixed and ζ will be left out of the notation.

We need to show that the cost function satisfies certain important properties, specifically being bounded above and below (away from 0) and being Lipschitz continuous:

Lemma 9. *The cost function* $\mathfrak{c}_\zeta(x) = (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}}$ *satisfies the following:*

i. $\mathfrak{c}_\zeta(x) \geqslant \mathfrak{c}^*(x)$ *(where* $\mathfrak{c}^*(x)$ *is the lucrativity function* $(f(x)g(x))^{\frac{1}{\gamma}}$*) for all* x*.*

ii. $\mathfrak{c}_\zeta(x)$ *is uniformly bounded away from* 0 *and is Lipschitz continuous. In particular, if* $\gamma \geqslant 1$*, then it is* α*-Lipschitz continuous where*

$$\alpha = \frac{1}{\gamma}\zeta^{\frac{2}{\gamma}-3}(f_{\max} + g_{\max}) \tag{3.9}$$

and $\mathfrak{c}_\zeta(x) \geqslant \zeta^{\frac{2}{\gamma}}$ *everywhere.*

iii. $\lim_{\zeta\to 0} \mathfrak{c}_\zeta(x) = \mathfrak{c}^*(x)$ *almost everywhere.*

Proof. We prove these in order.

i. This follows from the fact that $\hat{f}^{(\zeta)}$ and $\hat{g}^{(\zeta)}$ are upper bounds for f and g (which are nonnegative) by construction, hence

$$\mathfrak{c}_\zeta(x) = (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \geqslant (f(x)g(x))^{\frac{1}{\gamma}} = \mathfrak{c}^*(x)\,. \tag{3.10}$$

ii. This follows because by construction $\hat{f}^{(\zeta)}$ and $\hat{g}^{(\zeta)}$ are both $\geqslant \zeta$ everywhere, so

$$\mathfrak{c}_\zeta(x) = (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \geqslant \zeta^{\frac{2}{\gamma}}\,. \tag{3.11}$$

To show that they are Lipschitz continuous, we note that both f and g are bounded above (by $f_{\max}$ and $g_{\max}$ respectively) and that $\sup_x \hat{f}^{(\zeta)}(x) = f_{\max}$ and $\sup_x \hat{g}^{(\zeta)}(x) = g_{\max}$, and by construction $\hat{f}^{(\zeta)}$ and $\hat{g}^{(\zeta)}$ are $(1/\zeta)$-Lipschitz continuous.

However, the composition of two Lipschitz-continuous function is also Lipschitz-continuous (and the Lipschitz continuity factor of the composition is the product of the factors of the original two functions). Furthermore $(\cdot)^{\frac{1}{\gamma}}$ is Lipschitz-continuous if the domain is bounded away from 0 below and upper bounded away from ∞; and when the input is bounded below by ζ^2 and above by $f_{\max}g_{\max}$ (as in this case), it is α_1-Lipschitz continuous where

$$\alpha_1 = \begin{cases} \frac{1}{\gamma}\zeta^{2(\frac{1}{\gamma}-1)} & \text{if } \gamma \geqslant 1 \\ \frac{1}{\gamma}(f_{\max}g_{\max})^{\frac{1}{\gamma}-1} & \text{if } \gamma < 1 \end{cases} \tag{3.12}$$

We now analyze the Lipschitz-continuity of $\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x)$; the product of two bounded Lipschitz continuous functions is also a bounded Lipschitz continuous function. Specifically, if $h_1(x), h_2(x)$ are respectively ζ_1- and ζ_2-Lipschitz continuous nonnegative functions with respective fixed upper bounds $h_1^{\max}, h_2^{\max}$, then $h(x) = h_1(x)h_2(x)$ is also nonnegative, is bounded above by $h_1^{\max}h_2^{\max}$ and is $(\zeta_1 h_2^{\max} + \zeta_2 h_1^{\max})$-Lipschitz continuous. Thus, $\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x)$ is nonnegative, bounded above by $f_{\max}g_{\max}$ (and below by ζ^2), and is α_2-Lipschitz continuous where

$$\alpha_2 = (1/\zeta)(f_{\max} + g_{\max})\,. \tag{3.13}$$

Hence $\mathfrak{c}_\zeta(x)$ is α-Lipschitz continuous where $\alpha = \alpha_1\alpha_2$. In particular, when $\gamma \geqslant 2$ (the case we are most interested in) we have

$$\alpha = \frac{1}{\gamma}\zeta^{\frac{2}{\gamma}-3}(f_{\max} + g_{\max})\,. \tag{3.14}$$

iii. This follows as $\hat{f}^{(\zeta)} \to f$ and $\hat{g}^{(\zeta)} \to g$ almost everywhere as $\zeta \to 0$; hence *both* occur almost everywhere (the union of where they don't converge is measure 0). □

We also define cost-denominated versions of distance and reachable sets:

Definition 17. Given $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$, the cost-distance between them is

$$d_\Pi^{\mathfrak{c}}(\boldsymbol{q}, \boldsymbol{q}') = \inf(\ell^{\mathfrak{c}}(\pi) : \boldsymbol{q} \xrightarrow{\pi} \boldsymbol{q}', \pi \in \Pi) \tag{3.15}$$

Given $\varepsilon > 0$ and $\boldsymbol{q} \in \mathcal{Q}$, the *$\varepsilon$-cost reachable set* (in both $\mathcal{Q}$ and $\mathcal{X}$) are

$$R_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) := \{\boldsymbol{q}' \in \mathcal{Q} : d_\Pi^{\mathfrak{c}}(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon\} \quad \text{and} \quad \bar{R}_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) = [R_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})]_{\mathcal{X}} \tag{3.16}$$

Although intuitively on small scales the cost-reachable sets will resemble our original length-reachable sets (except scaled by the cost function $\mathfrak{c}$), they are not exactly the same shape. Thus, we need to show that this change does not alter our main assumptions, namely Assumption 8 and Assumption 9; we also need to show that (with sufficiently small $\zeta > 0$ regularization) it achieves the balancing effect we wanted.

Lemma 10. *Let b be the branching factor from Assumption 8. Then there are functions $\rho_i : \mathbb{R}_{>0} \to \mathbb{R}_{\geqslant 0}$ for $i = 1, 2, 3, 4$ satisfying $\lim_{\zeta \to 0} \rho_i(\zeta) = 0$ and a polynomial $P^{\mathfrak{c}}$ (which depends through the cost function $\mathfrak{c} := \mathfrak{c}_\zeta$ on ζ) such that for any $\zeta > 0, \boldsymbol{q} \in \mathcal{Q}$ and $0 < \varepsilon \leqslant \rho_1(\zeta)$ the following hold:*

i. There is a set $Q_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})$ such that $|Q_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})| \leqslant b$ and

$$R_{2\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})} R_{(1+\rho_2(\zeta))\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}')\,. \tag{3.17}$$

ii. There is a set $Q_\varepsilon^{\mathfrak{c},0}$ such that $|Q_\varepsilon^{\mathfrak{c},0}| \leqslant P^{\mathfrak{c}}(1/\varepsilon)$ and

$$\mathcal{Q}_f \subseteq \bigcup_{\boldsymbol{q} \in Q_\varepsilon^{\mathfrak{c},0}} R_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) \tag{3.18}$$

iii. $\mathbb{P}_{X \sim f}[X \in \bar{R}_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})] \leqslant (1 + \rho_3(\zeta))\varepsilon^\gamma$.

iv. $\mathbb{E}_{X \sim f}[\,\mathfrak{c}_\zeta(X)^{-1}] \geqslant (1 - \rho_4(\zeta))\mathbb{E}_{X \sim f}[(f(X)g(X))^{-\frac{1}{\gamma}}]$.

Proof. We assume without loss of generality that $\zeta \leqslant 1$ (so $\alpha_1 = \zeta^{\frac{2}{\gamma}} \leqslant 1$) and ε is sufficiently small so that Assumption 8 and Assumption 9 hold.

The lemma then holds due to the properties of $\mathfrak{c}_\zeta(x)$, namely that it is bounded below by α_1 and α_2-Lipschitz continuous; these two conditions show that on small scales, it cannot change too much in a multiplicative sense. First, we define

$$\overleftrightarrow{R}_\varepsilon(\boldsymbol{q}) \text{ and } \overleftrightarrow{R}_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) \tag{3.19}$$

to be the sets reachable from $\boldsymbol{q}$ using (possibly alternating) backwards and forwards trajectories in Π of, respectively total length $\leqslant \varepsilon$ or total cost $\leqslant \varepsilon$ (note that if Π is symmetric then $R_\varepsilon(\boldsymbol{q}) = \overleftrightarrow{R}_\varepsilon(\boldsymbol{q})$ and $R_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) = \overleftrightarrow{R}_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})$).

Then, letting $c_\Pi < \infty$ be the speed limit of Π in $\mathcal{X}$ (i.e. $\|[\dot{\pi}(t)]_{\mathcal{X}}\| \leqslant c_\Pi$ for any

$\pi \in \Pi$), which of course also applies for backwards trajectories, we get

$$\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}) - (5c_\Pi(\alpha_2/\alpha_1))\varepsilon \leqslant \mathfrak{c}_\zeta([\boldsymbol{q}']_\mathcal{X}) \leqslant \mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}) + (5c_\Pi(\alpha_2/\alpha_1))\varepsilon \tag{3.20}$$

$$\text{for any } \boldsymbol{q} \in \mathcal{Q} \text{ and } \boldsymbol{q}' \in \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{3.21}$$

since $\mathfrak{c}$ is α_2-Lipschitz continuous and $\boldsymbol{q}'$ is by definition reachable from $\boldsymbol{q}$ using Π (and backwards Π) in at most $(5/\alpha_1)\varepsilon$ time, which translates to at most $(5c_\Pi/\alpha_1)\varepsilon$ distance in the metric on $\mathcal{X}$, which finally translates to a change of at most $(5c_\Pi(\alpha_2/\alpha_1))\varepsilon$ in the value of $\mathfrak{c}$ by Lipschitz continuity. Then since $\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}) \geqslant \alpha_1$, we can turn this into multiplicative bounds:

$$(1 - (5c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}) \leqslant \mathfrak{c}_\zeta([\boldsymbol{q}']_\mathcal{X}) \leqslant (1 + (5c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}) \tag{3.22}$$

$$\text{for any } \boldsymbol{q} \in \mathcal{Q} \text{ and } \boldsymbol{q}' \in \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{3.23}$$

We then let $\rho^*_\zeta(\varepsilon) := (5c_\Pi(\alpha_2/\alpha_1^2))\varepsilon$ (noting that α_1, α_2 depend on ζ, as well as γ and f and g through $f_{\max}, g_{\max}$); for any fixed ζ, we have $\lim_{\varepsilon\to 0} \rho^*_\zeta(\varepsilon) = 0$.

Furthermore, since $\mathfrak{c}_\zeta(x) \geqslant \alpha_1$ for all x, we know that

$$\overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q}) \subseteq \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{3.24}$$

since any trajectory (allowing backwards movement) of cost $\leqslant 5\varepsilon$ must have length $\leqslant (5/\alpha_1)\varepsilon$. Thus, our bounds (3.22) hold for all $\boldsymbol{q}' \in \overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q})$ as well.

We now fix $\boldsymbol{q} \in \mathcal{Q}$ and consider $R^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}')$; if $R^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}') \subseteq \overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q})$, then we can conclude

$$R_{\varepsilon'/((1+\rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq R^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}') \subseteq R_{\varepsilon'/((1-\rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{3.25}$$

since at any $\boldsymbol{q}'' \in R^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}') \subseteq \overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q})$, we know that the cost is between $(1-\rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X})$ and $(1 + \rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X})$. Similarly, if $\overleftrightarrow{R}^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}') \subseteq \overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q})$, then

$$\overleftrightarrow{R}_{\varepsilon'/((1+\rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq \overleftrightarrow{R}^{\mathfrak{c}}_{\varepsilon'}(\boldsymbol{q}') \subseteq \overleftrightarrow{R}_{\varepsilon'/((1-\rho^*_\zeta(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{3.26}$$

(Note: the distinction between ε' and ε is very important in the above).

We now prove part (i). Equation (3.25) gives us the following:

$$R^{\complement}_{2\varepsilon}(\boldsymbol{q}) \subseteq R_{2\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \tag{3.27}$$

We can then apply Assumption 8 (since the right-hand side is the normal reachable set) to get a cardinality-b set $Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})$ such that

$$R_{2\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})} R_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}') \tag{3.28}$$

Note that without loss of generality, for any ε' and any $\boldsymbol{q}'$ (note: this $\boldsymbol{q}'$ is not related to $Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})$, it is any $\boldsymbol{q}' \in \mathcal{Q}$), we can assume

$$Q_{\varepsilon'}(\boldsymbol{q}') \subset \overleftrightarrow{R}_{3\varepsilon'}(\boldsymbol{q}') \tag{3.29}$$

since WLOG we can assume that any $\boldsymbol{q}'' \in Q_{\varepsilon'}(\boldsymbol{q}')$ has the property

$$R_{\varepsilon'}(\boldsymbol{q}'') \cap R_{2\varepsilon'}(\boldsymbol{q}') \neq \varnothing \tag{3.30}$$

as the points in $Q_{\varepsilon'}(\boldsymbol{q}')$ are being used to cover $R_{2\varepsilon'}(\boldsymbol{q}')$. This in turn yields

$$R_{\varepsilon'}(\boldsymbol{q}'') \subseteq \overleftrightarrow{R}_{4\varepsilon'}(\boldsymbol{q}') \tag{3.31}$$

for any $\boldsymbol{q}'' \in Q_{\varepsilon'}(\boldsymbol{q}')$. Thus we can apply this to

$$\varepsilon' = \varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}})) \text{ and } \boldsymbol{q}' = \boldsymbol{q} \tag{3.32}$$

which then yields for any $\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}$ that

$$R_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}') \subset \overleftrightarrow{R}_{4\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\complement_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \tag{3.33}$$

Then, for $\varepsilon \leqslant \alpha_1^2/(45\alpha_2 c_\Pi)$ we have

$$\rho_\zeta^*(\varepsilon) \leqslant 1/9 \tag{3.34}$$

which in turn means

$$4(1+\rho_\zeta^*(\varepsilon))/(1-\rho_\zeta^*(\varepsilon)) \leqslant 4(10/9)/(8/9) = 5 \tag{3.35}$$

and hence by (3.26) we have

$$\overleftrightarrow{R}_{4\varepsilon/((1-\rho_\zeta^*(\varepsilon))\, \mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \subseteq \overleftrightarrow{R}^{\mathfrak{c}}_{5\varepsilon}(\boldsymbol{q}) \tag{3.36}$$

This then finally yields by (3.25) that for all $\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\, \mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})$,

$$R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\, \mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}') \subseteq R^{\mathfrak{c}}_{((1+\rho_\zeta^*(\varepsilon))/(1-\rho_\zeta^*(\varepsilon)))\varepsilon}(\boldsymbol{q}') \tag{3.37}$$

when $\varepsilon \leqslant \alpha_1^2/(45\alpha_2 c_\Pi)$. We can then define

$$\rho_1(\zeta) = \zeta\alpha_1^2/(45\alpha_2 c_\Pi) \text{ and } \rho_2(\zeta) = \frac{1+\zeta/9}{1-\zeta/9} - 1 \tag{3.38}$$

in which case when $\varepsilon \leqslant \rho_1(\zeta)$ we get

$$\frac{1+\rho_\zeta^*(\varepsilon)}{1-\rho_\zeta^*(\varepsilon)} \leqslant \frac{1+\rho_\zeta^*(\rho_1(\zeta))}{1-\rho_\zeta^*(\rho_1(\zeta))} \leqslant \frac{1+\zeta/9}{1-\zeta/9} = 1+\rho_2(\zeta) \tag{3.39}$$

where $\lim_{\zeta\to 0} \rho_2(\zeta) = 0$. Thus, we finally have

$$R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\, \mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}') \subseteq R^{\mathfrak{c}}_{(1+\rho_2(\zeta))\varepsilon}(\boldsymbol{q}') \tag{3.40}$$

Putting this together with the above, we get

$$R^{\mathscr{C}}_{2\varepsilon}(\boldsymbol{q}) \subseteq R_{2\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \tag{3.41}$$

$$\subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})} R_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}') \tag{3.42}$$

$$\subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})} R^{\mathscr{C}}_{(1+\rho_2(\zeta))\varepsilon}(\boldsymbol{q}') \tag{3.43}$$

and $|Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})| \leqslant b$, and hence we may finally conclude that part (i) is true, with

$$Q^{\mathscr{C}}_{\varepsilon}(\boldsymbol{q}) = Q_{\varepsilon/((1-\rho^*_\zeta(\varepsilon))\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q})\,. \tag{3.44}$$

Part (ii) follows by considering $\mathscr{C}^*_{\max} := \sup_x (f(x)g(x))^{\frac{1}{\gamma}}$, which is a constant with regard to ζ and ε. Then, noting that $R_{\varepsilon/\mathscr{C}^*_{\max}}(\boldsymbol{q}') \subseteq R^{\mathscr{C}}_{\varepsilon}(\boldsymbol{q}')$ since $\mathscr{C}_\zeta(x) \leqslant \mathscr{C}^*_{\max}$, we can simply use $P^{\mathscr{C}}$ such that $P^{\mathscr{C}}(1/\varepsilon) = P(\mathscr{C}^*_{\max}/\varepsilon)$ and $Q^{\mathscr{C},0}_{\varepsilon} = Q_{\varepsilon/\mathscr{C}^*_{\max}}$ and we are done.

Part (iii) follows because when $\varepsilon \leqslant \rho_1(\zeta)$ and $\boldsymbol{q}' \in R^{\mathscr{C}}_{\varepsilon}(\boldsymbol{q})$,

$$(1 - (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \leqslant \mathscr{C}_\zeta([\boldsymbol{q}']_{\mathcal{X}}) \leqslant (1 + (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \tag{3.45}$$

(under the same logic as before but with radius ε rather than 5ε). Thus we know that

$$R^{\mathscr{C}}_{\varepsilon}(\boldsymbol{q}) \subseteq R_{\varepsilon/((1-(c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \tag{3.46}$$

Then we set $\rho_3(\zeta) = (1-\zeta/45)^{-\gamma} - 1$. Then, when $\varepsilon \leqslant \rho_1(\zeta) = \zeta\alpha_1^2/(45\alpha_2 c_\Pi)$, we have

$$1 - (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon \geqslant 1 - \zeta/45 \tag{3.47}$$

which thus means (taking the above and projecting to $\mathcal{X}$)

$$\bar{R}^{\mathscr{C}}_{\varepsilon}(\boldsymbol{q}) \subseteq \bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q}) \tag{3.48}$$

We then consider the volume of the set above. Since reachable set volumes approach $g(\boldsymbol{q})\varepsilon^\gamma$, for any $\rho_{3*}(\zeta)$ there is some $\rho_{1*}(\zeta)$ such that when $\varepsilon \leqslant \rho_{1*}(\zeta)$,

$$\mathrm{Vol}(\bar{R}_\varepsilon(\boldsymbol{q})) \leqslant (1+\rho_{3*}(\zeta))g(\boldsymbol{q})\varepsilon^\gamma \leqslant (1+\rho_{3*}(\zeta))g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma \tag{3.49}$$

Using $\varepsilon(1-\zeta/45)^{-1}/\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})$ instead of ε gives (for $\varepsilon \leqslant (1-\zeta/45)\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})\rho_{1*}(\zeta) \leqslant (1-\zeta/45)\zeta^{\frac{2}{\gamma}}\rho_{1*}(\zeta)$)

$$\begin{aligned}\mathrm{Vol}(\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})) &\leqslant (1+\rho_{3*}(\zeta))(1-\zeta/45)^{-\gamma}\,\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})^{-\gamma}g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma &(3.50)\\ &= (1+\rho_{3*}(\zeta))(1-\zeta/45)^{-\gamma}\hat{f}^{(\zeta)}([\boldsymbol{q}]_{\mathcal{X}})^{-1}\varepsilon^\gamma &(3.51)\end{aligned}$$

Finally we note that the entire reachable set is (by definition) within a distance of at most $\varepsilon(1-\zeta/45)^{-1}/\mathscr{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}})$ and hence the Lipschitz continuity (and boundedness away from 0) implies that for any $\rho_{3**}(\zeta)$, there is some $\rho_{1**}(\zeta)$ such that for all $\varepsilon < \rho_{1**}(\zeta)$,

$$\hat{f}^{(\zeta)}(\boldsymbol{q}') \leqslant (1+\rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q}) \text{ for all } \boldsymbol{q}' \text{ within distance } \varepsilon \text{ of } \boldsymbol{q} \tag{3.52}$$

Hence, setting $\varepsilon < \rho_{1**}(\zeta)(1-\zeta/45)\,\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \leqslant \rho_{1**}(\zeta)(1-\zeta/45)\zeta^{\frac{2}{\gamma}}$ then yields

$$\begin{aligned}
\mathbb{E}_{X\sim f}[X \in \bar{R}^{\mathscr{c}}_{\varepsilon}(\boldsymbol{q})] &\leqslant \mathbb{E}_{X\sim f}[X \in \bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})] && (3.53)\\
&= \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} f(x)\,dx && (3.54)\\
&\leqslant \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} \hat{f}^{(\zeta)}(x)\,dx && (3.55)\\
&\leqslant \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} (1+\rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})\,dx && (3.56)\\
&= (1+\rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})\mathrm{Vol}(\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathscr{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})) && (3.57)\\
&\leqslant (1+\rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})(1+\rho_{3*}(\zeta))(1-\zeta/45)^{-\gamma}\hat{f}^{(\zeta)}([\boldsymbol{q}]_{\mathcal{X}})^{-1}\varepsilon^{\gamma} && (3.58)\\
&= (1+\rho_{3**}(\zeta))(1+\rho_{3*}(\zeta))(1-\zeta/45)^{-\gamma}\varepsilon^{\gamma} && (3.59)
\end{aligned}$$

Thus, noting that $\rho_{3*}(\zeta)$ and $\rho_{3**}(\zeta)$ can be made arbitrarily small as $\zeta \to 0$ (and $(1-\zeta/45)^{-\gamma} \to 1$ as $\zeta \to 0$), we can let

$$\rho_3(\zeta) = (1+\rho_{3**}(\zeta))(1+\rho_{3*}(\zeta))(1-\zeta/45)^{-\gamma} - 1 \tag{3.60}$$

and the above will hold for all ε such that

$$\varepsilon \leqslant \rho_{1*}(\zeta)(1-\zeta/45)\zeta^{\frac{2}{\gamma}} \text{ and } \rho_{1**}(\zeta)(1-\zeta/45)\zeta^{\frac{2}{\gamma}} \tag{3.61}$$

We then set ρ_1 to be the minimum of the above values (of which there are only a fixed, finite number, so it remains positive).

Finally, part (iv) follows from the Monotone Convergence Theorem since

$$\mathbb{E}_{X\sim f}[(f(X)g(X))^{-\frac{1}{\gamma}}] = \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx < \infty \tag{3.62}$$

and $\lim_{\zeta\to 0}\mathscr{c}_\zeta(x)^{-1} = (f(x)g(x))^{-\frac{1}{\gamma}}$ wherever $f(x)g(x)$ is continuous, which is almost everywhere. □

Lemma 10(i) is the cost-reachable equivalent to the b-coverability assumption (Assumption 8) for our original (length) reachable sets; (ii) is similarly equivalent to the starting coverability assumption (Assumption 9); (iii) is a statement on the probability weights of the cost-reachable sets; and (iv) refers to the ability of $\mathfrak{c}_\zeta$ to approximate $\mathfrak{c}^*$ (while being bounded away from 0 and Lipschitz continuous).

3.2 Cost-Balanced Orienteering

Definition 18. The *cost-bounded trajectory set* of dynamic system Π is denoted

$$\Pi_\lambda^{\mathfrak{c}} := \{\pi \in \Pi : \ell^{\mathfrak{c}}(\pi) \leqslant \lambda\} \tag{3.63}$$

Then, the *Cost-Balanced Orienteering* problem is defined by

$$\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda) := \max_{\pi \in \Pi_\lambda^{\mathfrak{c}}} (|\bar{\pi} \cap \{X_i\}_{i=1}^n|) \tag{3.64}$$

i.e. the maximum number of targets which can be visited by a trajectory of cost at most λ.

Note that unlike normal Orienteering or TSP, even if we treat $X_1, \ldots, X_n$ as fixed, the problem depends on the density function f as that influences the cost function.

We now show a (with very high probability) lower bound on the CBO, which can then be turned into a corresponding upper bound on the Dynamic Stochastic TSP.

Proposition 7. [CBO upper bound] Let Π be a symmetric dynamic system and f be a probability density function satisfying the assumptions in Section 2.1, and let $\lambda > 0$ be fixed and sufficiently small. Then, there exists $0 < \beta$ (dependent on Π through branching factor b and small-time constraint factor γ but not directly on agility function g or density f) such that if $X_1, X_2, \ldots X_n \overset{iid}{\sim} f$, then

$$\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \leqslant \beta \lambda \, n^{\frac{1}{\gamma}} \tag{3.65}$$

with very high probability, where β does not depend on f, g, or λ.

Note that β depends on neither f nor g thanks to the cost balancing: changing f or g (by altering the dynamic system Π) also changes $\mathscr{C}_\zeta$ to preserve the balance.

We make the constants and probability bound of Proposition 7 more explicit:

Proposition 8. If the assumptions in Section 2.1 hold, then for any $\delta > 0$,

$$\mathbb{P}[\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \leqslant (1+\delta)\beta\lambda\, n^{\frac{1}{\gamma}}] \geqslant 1 - e^{-\frac{4}{5}(1+\delta)\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{3.66}$$

for sufficiently large n (where 'sufficiently large' can depend on δ).

Definition 19. Given a cost bound $\lambda > 0$, scale $\varepsilon > 0$, and approximation factor $\rho > 0$, we define a *λ-cost, ε-scale, ρ-approximate representation sequence* (to be referred to in general as a *representation sequence*) to be any sequence of configurations

$$\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_{\lceil \lambda/((1-\rho)\varepsilon)\rceil}) \quad \text{where} \quad \psi_k \in \mathcal{Q} \text{ for all } k \tag{3.67}$$

$$\text{such that} \quad \psi_0 \in Q^{\mathscr{C},0}_{(1-\rho)\varepsilon} \text{ and } \psi_k \in Q^{\mathscr{C}}_{(1-\rho/2)\varepsilon}(\psi_{k-1}) \text{ for all } k \geqslant 1 \tag{3.68}$$

Note that we scale back by $(1-\rho)$ for the initial covering set $Q^{\mathscr{C},0}_{(1-\rho)\varepsilon}$ but only by $(1-\rho/2)$ for the b-covering sets $Q^{\mathscr{C}}_{(1-\rho/2)\varepsilon}$.

We say $\boldsymbol{\psi}$ *represents* trajectory π, denoted $\pi \sqsupset \boldsymbol{\psi}$, if

$$\pi^{\mathscr{C}}(k(1-\rho)\varepsilon) \in R^{\mathscr{C}}_\varepsilon(\psi_k) \text{ for all } k \in \{0, 1, \ldots, n\} \tag{3.69}$$

We denote the set of all representation sequences (with parameters $\lambda, \varepsilon, \rho$) as $\Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho}$. Furthermore, we denote the cost-reachable set of $\boldsymbol{\psi}$ (with radius ε) as the union of the reachable sets of $\psi_0, \psi_1, \ldots, \psi_{\lceil \lambda/((1-\rho)\varepsilon)\rceil}$, which we denote

$$R^{\mathscr{C}}_\varepsilon(\boldsymbol{\psi}) := \bigcup_{k=0}^{\lceil \lambda/((1-\rho)\varepsilon)\rceil} R^{\mathscr{C}}_\varepsilon(\psi_k) \tag{3.70}$$

and we denote its workspace projection as

$$\bar{R}^{\mathfrak{C}}_{\varepsilon}(\boldsymbol{\psi}) := \bigcup_{k=0}^{\lceil \lambda/((1-\rho)\varepsilon)\rceil} \bar{R}^{\mathfrak{C}}_{\varepsilon}(\psi_k)\,. \tag{3.71}$$

Note that any representation sequence may represent many (even infinitely many) different valid trajectories, and any trajectory may be represented by multiple representation sequences. We now show a few of lemmas concerning representation sequences:

Lemma 11. *For any $\rho > 0$, there is some sufficiently small $\zeta > 0$ (which defines the cost function) and $\varepsilon^{(\rho)} > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^{(\rho)}$ such that any trajectory $\pi \in \Pi^{\mathfrak{C}}_{\lambda}$ such that $\pi(0) \in [\mathcal{X}_f]^{\mathcal{Q}}$ is represented by some $\boldsymbol{\psi} \in \Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho}$.*

Proof. First, we define $\zeta > 0$ sufficiently small so that $1 + \rho_2(\zeta) \leqslant 1/(1-\rho/2)$; we then use $\varepsilon^{(\rho)} = \rho_1(\zeta)$ (see Lemma 10 for the definitions of $\rho_2(\zeta)$ and $\rho_1(\zeta)$). This in particular means that for any $0 < \varepsilon \leqslant \varepsilon^{(\rho)}$ and any $\boldsymbol{q} \in \mathcal{Q}$, there is a set $Q^{\mathfrak{C}}_{(1-\rho)\varepsilon}(\boldsymbol{q})$ such that $|Q^{\mathfrak{C}}_{\varepsilon}(\boldsymbol{q})| \leqslant b$ and

$$R^{\mathfrak{C}}_{(2-\rho)\varepsilon}(\boldsymbol{q}) = R^{\mathfrak{C}}_{2(1-\rho/2)\varepsilon}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\boldsymbol{q})} R^{\mathfrak{C}}_{(1+\rho_2(\zeta))(1-\rho/2)\varepsilon}(\boldsymbol{q}') \subseteq \bigcup_{\boldsymbol{q}' \in Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\boldsymbol{q})} R^{\mathfrak{C}}_{\varepsilon}(\boldsymbol{q}')\,. \tag{3.72}$$

We then construct $\boldsymbol{\psi} \in \Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho}$ such that $\boldsymbol{\psi} \sqsupset \pi$ inductively.

Base case: We choose ψ_0 such that $\pi(0) = \pi^{\mathfrak{C}}(0) \in R^{\mathfrak{C}}_{\varepsilon}(\psi_0)$. This must exist because by definition the ε-radius reachable sets from points in $Q^{\mathfrak{C},0}_{(1-\rho)\varepsilon}$ cover $[\mathcal{X}_f]^{\mathcal{Q}}$.

Inductive step: For any $k > 0$, given ψ_{k-1} such that $\pi^{\mathfrak{C}}((k-1)(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{\varepsilon}(\psi_{k-1})$, we need to choose $\psi_k \in Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\psi_{k-1})$ such that $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{\varepsilon}(\psi_k)$. We do this by noting that by definition $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{\varepsilon}(\pi^{\mathfrak{C}}((k-1)(1-\rho)\varepsilon))$ (since they are only cost $(1-\rho)\varepsilon$ apart on valid trajectory π). Thus, $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{(2-\rho)\varepsilon}(\psi_{k-1})$ (since we can get from ψ_{k-1} to $\pi((k-1)(1-\rho)\varepsilon)$ with a ε-cost valid trajectory, and then from $\pi((k-1)(1-\rho)\varepsilon)$ to $\pi(k(1-\rho)\varepsilon)$ with an additional $(1-\rho)\varepsilon$-cost valid trajectory, thus in total a $(2-\rho)\varepsilon$-cost valid trajectory). But then by (3.72) we know that $R^{\mathfrak{C}}_{(2-\rho)\varepsilon}(\psi_{k-1})$

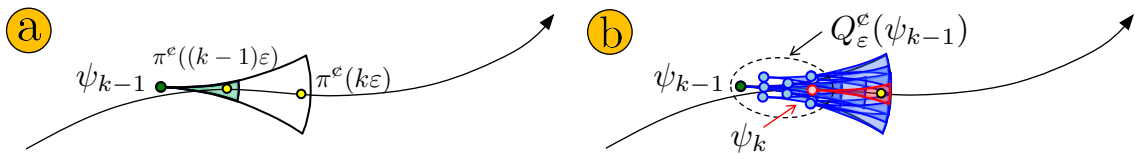

Figure 3-2: Implicit discretization: inductively building $\boldsymbol{\psi}$ representing π for a Dubins car; for simplicity the ρ error term is omitted from the illustration. (a) The situation at $\pi^{\mathfrak{C}}((k-1)\varepsilon)$, and extending the cost-reachable set from ψ_{k-1} to radius 2ε to encompass $\pi^{\mathfrak{C}}(k\varepsilon)$; (b) using Lemma 10 to cover $R^{\mathfrak{C}}_{2\varepsilon}(\psi_{k-1})$ with ε-cost reachable sets to select ψ_k. Note that these reachable sets are configuration reachable sets so the illustration is missing a dimension (the orientation).

is covered by the ε-radius cost reachable sets from points in $Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\psi_{k-1})$, and hence we can choose the one containing $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{(2-\rho)\varepsilon}(\psi_{k-1})$ to be ψ_k, thus maintaining $\psi_k \in Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\psi_{k-1})$ and $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon) \in R^{\mathfrak{C}}_{\varepsilon}(\psi_k)$.

Thus, we can build $\psi_0, \psi_1, \ldots$ like this until we have covered all of π. Since $\pi^{\mathfrak{C}}$ only takes in inputs from 0 to λ, we only need to do this up to $\pi^{\mathfrak{C}}(\lfloor \lambda/((1-\rho)\varepsilon)\rfloor\varepsilon)$ and we are done (we may even have to add another point to the representation sequence if $\lfloor \lambda/((1-\rho)\varepsilon)\rfloor = \lceil \lambda/((1-\rho)\varepsilon)\rceil - 1$, which we can do from $Q^{\mathfrak{C}}_{\varepsilon}(\psi_{\lfloor \lambda/((1-\rho)\varepsilon)\rfloor})$). □

Lemma 12. *For any $\boldsymbol{\psi} \sqsupset \pi \in \Pi$, we have $\pi \subset R^{\mathfrak{C}}_{2\varepsilon}(\boldsymbol{\psi})$; if Π is symmetric, then $\pi \subset R^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi})$. Furthermore, $\bar{\pi} \subset \bar{R}^{\mathfrak{C}}_{2\varepsilon}(\boldsymbol{\psi})$ holds for any Π and $\bar{\pi} \subset \bar{R}^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi})$ holds for symmetric Π.*

Proof. For any $\pi^{\mathfrak{C}}(t)$, we can write $t = k(1-\rho)\varepsilon + \tau$ where k is an integer (and is between 0 and $\lceil \lambda/\varepsilon \rceil$) and $\tau < (1-\rho)\varepsilon < \varepsilon$. Then $\pi^{\mathfrak{C}}(t) \in R^{\mathfrak{C}}_{2\varepsilon}(\psi_k)$ since $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon)$ is within ε cost from ψ_k and $\pi^{\mathfrak{C}}(t)$ is within $\tau < (1-\rho)\varepsilon < \varepsilon$ cost from $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon)$.

If Π is symmetric, then we divide into two cases: (i) $\tau \leqslant \varepsilon/2$; (ii) $\tau > \varepsilon/2$. In case (i), we have that $\pi^{\mathfrak{C}}(t)$ is at most $\tau \leqslant \varepsilon/2$ cost (moving forwards) from $\pi^{\mathfrak{C}}(k(1-\rho)\varepsilon)$, which is at most ε cost from ψ_k, hence

$$\pi^{\mathfrak{C}}(t) \in R^{\mathfrak{C}}_{(3/2)\varepsilon}(\psi_k) \subseteq R^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi}). \tag{3.73}$$

In case (ii) we have that $\pi^{\mathfrak{C}}(t)$ is at most $(1-\rho)\varepsilon - \tau < \varepsilon/2$ from $\pi^{\mathfrak{C}}((k+1)(1-\rho)\varepsilon)$ (moving backward along π, which is allowed by symmetric Π), and $\pi^{\mathfrak{C}}((k+1)(1-\rho)\varepsilon)$

is at most ε cost from ψ_{k+1}, hence

$$\pi^{\mathfrak{C}}(t) \in R^{\mathfrak{C}}_{(3/2)\varepsilon}(\psi_{k+1}) \subseteq R^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi}). \tag{3.74}$$

In both cases, we have $\pi^{\mathfrak{C}}(t) \in R^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi})$ for any $t \in [0, \lambda]$ and hence $\pi \subseteq R^{\mathfrak{C}}_{(3/2)\varepsilon}(\boldsymbol{\psi})$.

The statements for these sets projected into $\mathcal{X}$ then follow trivially. □

Lemma 13. *Let b be the branching factor and $P^{\mathfrak{C}}$ the polynomial from Lemma 10(ii),*

$$|\Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho}| \leqslant P^{\mathfrak{C}}(1/((1-\rho)\varepsilon))b^{\lceil \lambda/((1-\rho)\varepsilon)\rceil} \tag{3.75}$$

Additionally, for any fixed $\delta > 0$, there is some $\rho^{(\delta)} > 0$ and $\varepsilon^{(\delta)} > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^{(\delta)}$,

$$|\Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho^{(\delta)}}| \leqslant b^{(1+\delta)\lambda/\varepsilon}. \tag{3.76}$$

Proof. This follows from the fact that for any $\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_{\lceil \lambda/((1-\rho)\varepsilon)\rceil}) \in \Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho}$, we have $\psi_0 \in Q^{\mathfrak{C},0}_{\varepsilon}$ and $\psi_k \in Q^{\mathfrak{C}}_{\varepsilon}(\psi_{k-1})$ for all $k > 0$. Thus, we have $|Q^{\mathfrak{C},0}_{(1-\rho)\varepsilon}| \leqslant P^{\mathfrak{C}}(1/((1-\rho)\varepsilon))$ choices for ψ_0, and then for each $k > 0$ (iterating from 1 to $\lceil \lambda/((1-\rho)\varepsilon)\rceil$) we have $|Q^{\mathfrak{C}}_{(1-\rho/2)\varepsilon}(\psi_{k-1})| \leqslant b$ choices. Thus the total number of ways to construct $\boldsymbol{\psi} \in \Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho}$ is at most $P^{\mathfrak{C}}(1/((1-\rho)\varepsilon))b^{\lceil \lambda/((1-\rho)\varepsilon)\rceil}$.

Finally, the approximation holds because for any $\delta > 0$, we can select $\rho^{(\delta)}$ such that $1/(1-\rho^{(\delta)}) \leqslant 1 + \delta/2$. We note that

$$b^{(1+\delta)\lambda/\varepsilon} / b^{\lceil \lambda/((1-\rho^{(\delta)})\varepsilon)\rceil} \geqslant b^{((1+\delta)-1/(1-\rho^{(\delta)}))\lambda/\varepsilon - 1} \geqslant b^{(\delta/2)\lambda/\varepsilon - 1} \tag{3.77}$$

grows faster than any polynomial in $1/\varepsilon$, and specifically it must grow faster than $P^{\mathfrak{C}}(1/((1-\rho^{(\delta)})\varepsilon))$. Thus, for sufficiently small ε,

$$|\Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho^{(\delta)}}| \leqslant P^{\mathfrak{C}}(1/((1-\rho)\varepsilon))b^{\lceil \lambda/((1-\rho^{(\delta)})\varepsilon)\rceil} \leqslant b^{(1+\delta)\lambda/\varepsilon} \tag{3.78}$$

and we are done. □

Furthermore, given any constant $r \geqslant 0$, the probability mass of any $\bar{R}^{\mathfrak{C}}_{r\varepsilon}(\boldsymbol{\psi})$ (and thus the expected number of targets that fall within it) is easy to bound:

Lemma 14. *For any fixed $\lambda > 0$ and $\delta > 0$, there is a sufficiently small $\rho^{(\delta)} > 0$ such that for any sufficiently small $\varepsilon > 0$ and $\boldsymbol{\psi} \in \Psi^{\mathfrak{C}}_{\lambda,\varepsilon,\rho^{(\delta)}}$,*

$$\mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathfrak{C}}_{r\varepsilon}(\boldsymbol{\psi})] \leqslant (1+\delta) r^{\gamma} \lambda \varepsilon^{\gamma-1}\,. \tag{3.79}$$

Proof. We know that for any $\psi_k \in \mathcal{Q}$,

$$\mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathfrak{C}}_{\varepsilon}(\psi_k)] \leqslant (1+o(1))\varepsilon^{\gamma} \tag{3.80}$$

Therefore, using $r\varepsilon$ as the radius (which doesn't affect the fact that the $o(1)$ error term still goes to 0 as $\varepsilon \to 0$ and only depends on ε) and letting $\rho^{(\delta)}$ (as in the previous lemma) satisfy $1/(1-\rho^{(\delta)}) < 1 - \delta/2$, we can take a union bound. When $\varepsilon < \delta\lambda/8$, there are

$$\lceil \lambda/((1-\rho^{(\delta)})\varepsilon)\rceil + 1 \leqslant (1+\delta/2)\lambda/\varepsilon + 2 \leqslant (1+(3/4)\delta)\lambda/\varepsilon \tag{3.81}$$

possible values of k. We can then take ε sufficiently small so that $\mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathfrak{C}}_{r\varepsilon}(\psi_k)] \leqslant (1+\delta^*) r^{\gamma}\varepsilon^{\gamma}$, where δ^* is the value (dependent on δ only) such that $(1+\delta^*)(1+(3/4)\delta) = 1+\delta$); then by the union bound, we have

$$\begin{aligned}
\mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathfrak{C}}_{r\varepsilon}(\boldsymbol{\psi})] &\leqslant \bigcup_{k=0}^{\lceil \lambda/((1-\rho^{(\delta)})\varepsilon)\rceil} \mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathfrak{C}}_{r\varepsilon}(\psi_k)] && (3.82)\\
&\leqslant (\lceil \lambda/((1-\rho^{(\delta)})\varepsilon)\rceil + 1)(1+\delta^*) r^{\gamma}\varepsilon^{\gamma} && (3.83)\\
&\leqslant (1+\delta) r^{\gamma}\lambda\varepsilon^{\gamma-1} && (3.84)
\end{aligned}$$

as we wanted. □

We now define a new problem in which the goal is to find a representation sequence whose 2ε-cost-reachable set (if Π is nonsymmetric) or $(3/2)\varepsilon$-cost-reachable set (if Π is symmetric) contains as many target points as possible. We formally define:

Definition 20. Given length bound $\lambda > 0$, scale $\varepsilon > 0$, and approximation factor $\delta > 0$, we define the *Sequence Containment Problem* (SCP) as follows. First, let $\rho := \rho^{(\delta)}$ be sufficiently small so that Lemmas 13 and 14 hold (we can take the minimum of the values necessary for each), and let $\zeta > 0$ (the cost function regularization factor) be sufficiently small so that Lemma 11 holds (thus defining the cost function $\mathfrak{c}$). If Π is nonsymmetric, let $r = 2$; if Π is symmetric, let $r = 3/2$; then

$$\mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) := \max_{\boldsymbol{\psi} \in \Psi^{\mathfrak{c}}_{\lambda,\varepsilon,\rho}} |\{X_1, \ldots, X_n\} \cap \bar{R}^{\mathfrak{c}}_{r\varepsilon}(\boldsymbol{\psi})| \tag{3.85}$$

This replaces the optimization problem over the tricky and uncountably infinite Π (regarded as a set of trajectories) with an optimization problem over the finite set $\Psi^{\mathfrak{c}}_{\lambda,\varepsilon,\rho}$. Furthermore, the new problem is an upper bound for the old one:

Lemma 15. *For any dynamics Π, density function f (which influences the cost function), cost constraint $\lambda > 0$ and scale $\varepsilon > 0$,*

$$\mathrm{CBO}_\Pi(X_1, \ldots, X_n; \lambda) \leqslant \mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) \tag{3.86}$$

Proof. This follows from Lemmas 11 and 12: for any $\pi \in \Pi^{\mathfrak{c}}_\lambda$, there exists $\boldsymbol{\psi} \in \Psi^{\mathfrak{c}}_{\lambda,\varepsilon,\rho}$ such that $\boldsymbol{\psi} \sqsupset \pi$, and this implies $\bar{\pi} \subseteq \bar{R}^{\mathfrak{c}}_{r\varepsilon}(\boldsymbol{\psi})$. Thus,

$$|\{X_1, \ldots, X_n\} \cap \bar{\pi}| \leqslant |\{X_1, \ldots, X_n\} \cap \bar{R}^{\mathfrak{c}}_{r\varepsilon}(\boldsymbol{\psi})| \tag{3.87}$$

and so the maximum of the former over $\pi \in \Pi$ is at most the maximum of the latter over $\boldsymbol{\psi} \in \Psi^{\mathfrak{c}}_{\lambda,\varepsilon,\rho}$. □

We now prove Proposition 8. In order to find an upper bound (wvhp) to the CBO, we want to obtain a (wvhp) upper bound to $\mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta)$, which by Lemma 15 will then hold for the CBO. Since we can set $\varepsilon > 0$ to any (sufficiently small) value, we use $\varepsilon = n^{-\frac{1}{\gamma}}$ (which will get arbitrarily small as $n \to \infty$).

We now consider first fixing $\boldsymbol{\psi}$ (arbitrarily) before $X_1, \ldots, X_n \overset{iid}{\sim} f$ are chosen and then looking at $|\{X_1, \ldots, X_n\} \cap \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})|$ as a random variable (since it takes random inputs $X_1, \ldots, X_n$). We let $p^{(\boldsymbol{\psi})} := \mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})]$. Plugging into Lemma 14 yields

$$p^{(\boldsymbol{\psi})} = \mathbb{P}_{X\sim f}[X \in \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})] \leqslant (1+\delta)\lambda r^{\gamma} n^{-(1-\frac{1}{\gamma})} \tag{3.88}$$

$$\implies \mathbb{E}_{X_i \overset{iid}{\sim} f}[|\{X_1, \ldots, X_n\} \cap \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})|] \leqslant (1+\delta)\lambda r^{\gamma} n^{\frac{1}{\gamma}} \tag{3.89}$$

Additionally, with $\boldsymbol{\psi} \in \Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho}$ fixed, we can let $Z_i := Z_i^{(\boldsymbol{\psi})} := \mathbb{1}\{X_i \in \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})\}$ and

$$Z := Z^{(\boldsymbol{\psi})} := \sum_{i=1}^{n} Z_i^{(\boldsymbol{\psi})} = |\{X_1, \ldots, X_n\} \cap \bar{R}^{\mathscr{C}}_{r\varepsilon}(\boldsymbol{\psi})| \tag{3.90}$$

We define $p^* = (1+\delta)r^{\gamma}n^{-(1-\frac{1}{\gamma})}$. Then we know that $Z_i \overset{iid}{\sim} \text{Bern}(p^{(\boldsymbol{\psi})})$, and that for sufficiently small ε (which translates to sufficiently large n since $\varepsilon \to 0$ as $n \to \infty$) we have $p^{(\boldsymbol{\psi})} \leqslant p^*$ for all $\boldsymbol{\psi}$; thus, we can WLOG assume $p^{(\boldsymbol{\psi})} \leqslant p^*$ for all $\boldsymbol{\psi} \in \Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho}$.

We then define $Z_1^*, \ldots, Z_i^* \overset{iid}{\sim} \text{Bern}(p^*)$ and $Z^* = \sum_{i=1}^{n} Z_i^*$. Since $p^{(\boldsymbol{\psi})} \leqslant p^*$, for any $A > 0$,

$$\mathbb{P}[Z^{(\boldsymbol{\psi})} \geqslant A] \leqslant \mathbb{P}[Z^* \geqslant A] \tag{3.91}$$

Let $\mu^* := \mathbb{E}[Z^*] = (1+\delta)r^{\gamma}\lambda n^{\frac{1}{\gamma}}$.

Therefore we use Lemma 15, the union bound, equation (3.91), and Lemma 13

(in that order) to get that for all $A > 0$ (where the probabilities are over $X_i \overset{iid}{\sim} f$):

$$\mathbb{P}\big[\mathrm{CBO}_\Pi(X_1, \ldots, X_n; f, \lambda) \geqslant A\big] \leqslant \mathbb{P}\big[\mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) \geqslant A\big] \tag{3.92}$$

$$= \mathbb{P}[\exists\, \boldsymbol{\psi} \in \Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho} : Z^{(\boldsymbol{\psi})} \geqslant A] \tag{3.93}$$

$$\leqslant \sum_{\boldsymbol{\psi} \in \Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho}} \mathbb{P}[Z^{(\boldsymbol{\psi})} \geqslant A] \tag{3.94}$$

$$\leqslant |\Psi^{\mathscr{C}}_{\lambda,\varepsilon,\rho}|\, \mathbb{P}[Z^* \geqslant A] \tag{3.95}$$

$$\leqslant b^{(1+\delta)\lambda/\varepsilon}\, \mathbb{P}[Z^* \geqslant A] \tag{3.96}$$

$$= b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}}\, \mathbb{P}[Z^* \geqslant A] \tag{3.97}$$

for any sufficiently small $\varepsilon > 0$, i.e. sufficiently large n.

Since Z^* is a sum of iid Bernoulli random variables, we can apply the Chernoff bound. In particular, let $A = (1 + \xi)\mu^*$. Then the upper Chernoff bound says

$$\mathbb{P}[Z^* \geqslant A] = \mathbb{P}[Z^* \geqslant (1 + \xi)\mu^*] \tag{3.98}$$

$$\leqslant e^{-\frac{\xi^2}{2+\xi}\mu^*} \tag{3.99}$$

$$= e^{-\frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda n^{\frac{1}{\gamma}}} \tag{3.100}$$

We then note that equation (3.97), which is an upper bound on the probability that the CBO problem yields a result $\geqslant A$, is

$$b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}}\, \mathbb{P}[Z^* \geqslant A] \leqslant e^{\log(b)\,(1+\delta)\,\lambda n^{\frac{1}{\gamma}} - \frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda n^{\frac{1}{\gamma}}} \tag{3.101}$$

$$= e^{\left(\log(b) - \frac{\xi^2}{2+\xi} r^\gamma\right)\lambda(1+\delta)n^{\frac{1}{\gamma}}} \tag{3.102}$$

We note that since $\lambda(1 + \delta) > 0$, as long as $\log(b) - \frac{\xi^2}{2+\xi} r^\gamma < 0$, the above goes to 0 as $n \to \infty$ (and does so according to $e^{-cn^{\frac{1}{\gamma}}}$, i.e. with very high probability). We now analyze ξ as defined in (1.19): since $\log(b)/r^\gamma > \sqrt{\log(b)/r^\gamma} \iff \log(b)/r^\gamma >$

$1 \iff \log(b) > r^\gamma$, if $\log(b) > r^\gamma$ we get

$$\log(b) - \frac{\xi^2}{2+\xi} r^\gamma = \log(b) - \frac{9(\log(b)/r^\gamma)^2}{2 + 3(\log(b)/r^\gamma)} r^\gamma \tag{3.103}$$

$$\leqslant \log(b) - \frac{9(\log(b)/r^\gamma)^2}{5(\log(b)/r^\gamma)} r^\gamma \tag{3.104}$$

$$= -\frac{4}{5}\log(b) \tag{3.105}$$

Similarly, if $\log(b) \leqslant r^\gamma$ we get

$$\log(b) - \frac{\xi^2}{2+\xi} r^\gamma = \log(b) - \frac{9(\log(b)/r^\gamma)}{2 + 3\sqrt{\log(b)/r^\gamma}} r^\gamma \tag{3.106}$$

$$\leqslant \log(b) - \frac{9(\log(b)/r^\gamma)}{5} r^\gamma \tag{3.107}$$

$$= -\frac{4}{5}\log(b) \tag{3.108}$$

and hence we have in either case the bound

$$\log(b) - \frac{\xi^2}{2+\xi} r^\gamma \leqslant -\frac{4}{5}\log(b) \tag{3.109}$$

Then for sufficiently large n (since $\varepsilon = n^{-\frac{1}{\gamma}}$ this is equivalent to 'for sufficiently small ε') and letting $\{X_i\} := \{X_1, \ldots, X_n\}$ we have

$$\mathbb{P}[\mathrm{CBO}_\Pi(\{X_i\}; f, \lambda) \geqslant (1+\delta)\beta\lambda n^{\frac{1}{\gamma}}] \leqslant \mathbb{P}[\mathrm{SCP}_\Pi(\{X_i\}; \lambda, \varepsilon, \rho) \geqslant (1+\delta)\beta\lambda n^{\frac{1}{\gamma}}] \tag{3.110}$$

$$\leqslant |\Psi^{\not\in}_{\lambda,\varepsilon,\rho}|\ \mathbb{P}[Z^* \geqslant (1+\xi)\mu^*] \tag{3.111}$$

$$\leqslant b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}}\ e^{\left(-\frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda \varepsilon^{\gamma-1} n\right)} \tag{3.112}$$

$$\leqslant e^{-\frac{4}{5}(1+\delta)\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{3.113}$$

thus proving Propositions 7 and 8.

3.3 DSTSP lower bound

While the Cost-Balanced Orienteering problem can be bounded cleanly, it replaces the length of the trajectory with a cost function. Thus, the direct correspondence between Orienteering and TSP doesn't hold with CBO and we need a new technique. Let's denote

$$\Pi^{\{X_i\}} := \{\pi \in \Pi : X_i \in \bar{\pi} \text{ for all } i\} \tag{3.114}$$

i.e. the set of all TSP solution trajectories (not necessarily the minimum length). We then take some $\pi \in \Pi^{\{X_i\}}$ and want to show that with very high probability it must be at least a certain length. We do this by chopping π into segments of λ cost; if λ is sufficiently short, the cost function will be roughly constant over the span of any of these segments (since $\mathcal{c}_\zeta$ is Lipschitz continuous). Thus, the length of such a segment will be at least roughly $\lambda / \mathcal{c}_\zeta(x)$ where x is any point on the segment. Thus, we can let ℓ_j be the cost of the jth such segment, and let $j(i)$ be the segment that X_i falls in; we can then consider the sum over all i of the length of the segment that X_i falls in, which is

$$\sum_{i=1}^{n} \ell_{j(i)} \approx \sum_{i=1}^{n} \lambda / \mathcal{c}_\zeta(X_i) \approx n\lambda \mathbb{E}_{X\sim f}[\,\mathcal{c}_\zeta(X)^{-1}] \approx (1-\delta) n\lambda \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\,. \tag{3.115}$$

However, this may well overcount the total length since each segment is counted once per X_i that falls on it; but by Proposition 8 we know that no such segment can have more than $(1+\delta)\beta\lambda n^{\frac{1}{\gamma}}$ targets on it. Thus, the total length is at least this sum divided by $(1+\delta)\beta\lambda n^{\frac{1}{\gamma}}$; combining the two approximation factors δ (they can both be set arbitrarily small), we end with the conclusion that for any $\delta > 0$, for all sufficiently large n,

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{3.116}$$

with very high probability.

Let us fix $\lambda > 0$ generate $X_1, X_2, \ldots \overset{iid}{\sim} f$ (an infinite sequence of targets, of which we will look at trajectories visiting the first n and let $n \to \infty$). For simplicity we define

$$\mathrm{CBO}^{(n)} := \mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \tag{3.117}$$

Let β and ξ be as defined in Definition 8. We will be using $(1-\delta/2)^{-1}$ instead of $1+\delta$ for the approximation (this will make the TSP bounds more elegant to state), which are not equal but for any $\delta_1 > 0$, there is some $\delta_2 > 0$ such that $1 + \delta_1 = (1 - \delta_2/2)^{-1}$ and vice versa. Thus, with this alteration, Proposition 8 holds for all $\delta > 0$ for all sufficiently large n (where 'sufficiently large' may depend on δ),

$$\mathbb{P}[\mathrm{CBO}^{(n)} \leqslant (1-\delta/2)\beta\lambda\, n^{\frac{1}{\gamma}}] \geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}} \geqslant 1 - e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{3.118}$$

For the remainder of this section, we assume the bound on $\mathrm{CBO}^{(n)}$ holds; at the end we will incorporate the probability that it fails into our bound. Let $\pi := \pi^{(n)} \in \Pi$ be a TSP trajectory for targets $\{X_1, \ldots, X_n\}$ and let $t_1, t_2, \ldots, t_n \in [0, \ell(\pi)]$ satisfy $\bar{\pi}(t_i) = X_i$ (t_i is the time when π visits X_i, if there's more than one then choose arbitrarily). Without loss of generality we assume that $\max_i t_i = \ell(\pi)$ (a TSP trajectory has no need to continue once it has visited all n targets).

We want to partition π into cost-λ segments. However, π might not divide evenly into cost-λ segments, so we define the following values:

$$k := \lceil \ell^{\mathfrak{c}}(\pi)/\lambda \rceil \quad \text{and} \quad \lambda' := \ell^{\mathfrak{c}}(\pi)/k \tag{3.119}$$

Thus, $\lambda - 1/k \leqslant \lambda' \leqslant \lambda$. Furthermore, since the bound on $\mathrm{CBO}^{(n)}$ holds, we know that any cost-λ' trajectory can have at most $(1-\delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}}$ targets in it, and hence

to get all n points we need

$$k \geqslant \frac{n}{(1-\delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}}} = (1-\delta/2)\beta^{-1}\lambda^{-1}n^{1-\frac{1}{\gamma}} \tag{3.120}$$

such segments. Thus, as $n \to \infty \implies k \to \infty \implies \lambda' \to \lambda$. We define $0 = t'_0 < t'_1 < \cdots < t'_k = \ell(\pi)$ such that

$$t'_j := \min\left(t : \int_0^t \mathscr{C}_\zeta(\pi(\tau))\, d\tau = j\lambda'\right) \tag{3.121}$$

i.e. the trajectory reaches cost $j\lambda'$ at time t'_j. This also means that

$$\ell^{\mathscr{C}}(\pi_{[t'_{j-1},t'_j]}) = \lambda' \tag{3.122}$$

i.e. the cost π accumulates between any t'_{j-1} and t'_j is λ'. For any j, let $A_j := \{i : t_i \in [t'_{j-1}, t'_j)\}$ (and $A_k = \{i : t_i \in [t'_{k-1}, t'_k]\}$) so the whole interval $[0, \ell(\pi)]$ is included), i.e. π visits X_i in the time interval $[t'_{j-1}, t'_j)$. Because $\mathrm{CBO}^{(n)}$ is bounded and by (3.122), we know that for sufficiently large n, all $j \in [k]$ satisfy

$$|A_j| \leqslant (1-\delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}}\,. \tag{3.123}$$

We also let $\ell_j := \ell_j^{(\pi,\lambda)} = t'_j - t'_{j-1}$, corresponding to the length of the jth cost-λ' trajectory, and define $j(i)$ such that $i \in A_{j(i)}$ (π visits X_i in the $j(i)$th cost-λ' interval); note that it takes λ and not λ' as an input (and λ' is determined by λ and $\ell^{\mathscr{C}}(\pi)$ as described above). Note that the collection $\{A_j\}_{j\in[k]}$ partitions $[n]$.

Since $\mathscr{C}_\zeta(x) \geqslant \alpha_1$ everywhere and $\mathscr{C}_\zeta$ is α_2-Lipschitz continuous, for any $\eta > 0$ we can find λ such that for all $\lambda' \leqslant \lambda$, if π' is a trajectory with cost $\ell^{\mathscr{C}}(\pi') = \lambda'$ and

$X \in \bar{\pi}'$ then $\ell(\pi') \geqslant (1-\eta)\lambda'/\not{c}_\zeta(X)$. We then derive the following:

$$\ell(\pi) = \sum_{j=1}^{k} \ell_j \tag{3.124}$$

$$= \frac{1-\delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{j=1}^{k} (1-\delta/2)^{-1} \beta\lambda n^{\frac{1}{\gamma}} \ell_j \tag{3.125}$$

$$\geqslant \frac{1-\delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{j=1}^{k} |A_j| \ell_j \tag{3.126}$$

$$= \frac{1-\delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{i=1}^{n} \ell_{j(i)} \tag{3.127}$$

This then leads to the intuitive steps (note that we use $\lambda \approx \lambda'$ and remove the δ terms):

$$\ell(\pi) \gtrsim \frac{1}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{i=1}^{n} \lambda/\not{c}_\zeta(X_i) \text{ with very high probability} \tag{3.128}$$

$$\gtrsim \frac{1}{\beta} n^{1-\frac{1}{\gamma}} \mathbb{E}_{X\sim f}[\not{c}_\zeta(X)^{-1}] \text{ with very high probability} \tag{3.129}$$

$$\approx \frac{1}{\beta} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{3.130}$$

The two '$\gtrsim$' steps (3.128) and (3.129) need to be precisely stated and justified: (3.128) happens because the cost-λ' segment $\pi_{[t'_{j(i)-1}, t'_{j(i)}]}$ contains X_i but requires care in handling the approximation, while (3.129) is shown with a concentration bound (and because when n is large $\lambda' \approx \lambda$). The approximation in (3.130) is justified by using a regularization factor ζ which is sufficiently small to yield an approximation as close as we desire.

Before we precisely state and prove the results that will give us steps (3.128) and (3.129), we discuss our main tool for showing them, which is the *one-sided Bernstein condition* and *Bernstein's inequality* (see [21]). We will not define the one-sided Bernstein condition since we don't directly use the definition; instead, we give a (known) sufficient condition, namely that a random variable which is bounded above satisfies it:

Proposition 9. If $Y - \mathbb{E}[Y] \leqslant a'$ (guaranteed) for some $a' > 0$, then Y satisfies the one-sided Bernstein condition with parameter $a = a'/3$.

This is useful because it allows us to use the (one sided) *Bernstein's inequality*:

Proposition 10. Let $Y_1, \dots Y_n$ be iid random variables (each with expectation $\mathbb{E}[Y]$ and variance $\mathrm{Var}[Y]$) satisfying the one-sided Bernstein condition with parameter $a > 0$. Then, for any $\delta > 0$,

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}[Y] \geqslant \delta\Big] \leqslant e^{-\frac{n\delta^2/2}{\mathrm{Var}[Y]+a\delta}} \tag{3.131}$$

This is usually given in terms of Y bounded above because the one-sided Bernstein inequality is given as an upper tail bound (as it's typically used in that fashion); here we are trying to derive a lower tail bound of a sum of iid random variables which are bounded below (because they are all nonnegative) so we rephrase them as:

Proposition 11. Let $Y_1, \dots, Y_n$ be iid nonnegative random variables (each with expectation $\mathbb{E}[Y]$ and variance $\mathrm{Var}[Y]$). Then for any $\delta > 0$,

$$\mathbb{P}\Big[\sum_{i=1}^{n} Y_i \leqslant (1-\delta)n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n\mathbb{E}[Y]^2\delta^2/2}{\mathrm{Var}[Y]+\mathbb{E}[Y]^2\delta/2}} \tag{3.132}$$

Proof. By Proposition 9, if Y is nonnegative then $-Y$ satisfies the Bernstein condition with parameter $a = \mathbb{E}[Y]/3$. Then we can re-write

$$\sum_{i=1}^{n} Y_i \leqslant (1-\delta)n\mathbb{E}[Y] \iff \sum_{i=1}^{n}(-Y_i) \geqslant (1-\delta)n\mathbb{E}[-Y] \tag{3.133}$$

$$\iff \frac{1}{n}\sum_{i=1}^{n}(-Y_i) - \mathbb{E}[-Y] \geqslant \delta\,\mathbb{E}[Y] \tag{3.134}$$

Plugging this into Proposition 10 then yields the result. □

Let $Y_i := \mathfrak{c}_\zeta(X_i)^{-1} \leqslant 0$, and if $X \sim f$ we define $Y := \mathfrak{c}_\zeta(X)^{-1} \leqslant 0$ (the generic version of Y_i). We need to establish a few things:

Lemma 16. *For any approximation factor $\delta < 0$, there is a sufficiently small regularization factor $\zeta := \zeta^{(\delta)} > 0$ so that $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$ satisfy*

$$(1-\delta)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx \leqslant \mathbb{E}[Y] \leqslant \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx \quad (3.135)$$

$$\textit{and} \quad \mathrm{Var}[Y] \leqslant \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}} g(x)^{-\frac{2}{\gamma}}\,dx - (1-\delta)^2\left(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx\right)^2 \quad (3.136)$$

which are both finite when $\gamma \geqslant 2$. Additionally, even when $\gamma < 2$, we can bound the expected value and the variance above using the regularization factor itself:

$$(1-\delta)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx \leqslant \mathbb{E}[Y] \leqslant \alpha_1^{-1} \quad (3.137)$$

$$\textit{and} \quad \mathrm{Var}[Y] \leqslant \alpha_1^{-2}/4\,. \quad (3.138)$$

Proof. We recall that the lucrativity function (of which the cost function $\mathfrak{c}_\zeta$ is an approximation) is $\mathfrak{c}^*(x) = (f(x)g(x))^{\frac{1}{\gamma}}$ and that $\gamma \geqslant 2$. Let $Y^* = \mathfrak{c}^*(X)$ where $X \sim f$. Then we have

$$\begin{aligned}
\mathbb{E}[Y^*] &= \int (\mathfrak{c}^*)^{-1}\,df && (3.139)\\
&= \int_{\mathcal{X}_f} f(x)(f(x)g(x))^{-\frac{1}{\gamma}}\,dx && (3.140)\\
&= \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\,dx && (3.141)\\
&\leqslant \int_{\mathcal{X}_f} (1+f(x)) g_{\min}^{-\frac{1}{\gamma}}\,dx && (3.142)\\
&= (\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f)+1) g_{\min}^{-\frac{1}{\gamma}} && (3.143)\\
&< \infty && (3.144)
\end{aligned}$$

where (3.142) follows because $g(x) \geqslant g_{\min}$ for all x and

$$f(x)^{1-\frac{1}{\gamma}} \leqslant \max(1, f(x)) \leqslant 1 + f(x) \quad (3.145)$$

(and $\int_{\mathcal{X}_f} f(x)\,dx = 1$ by definition). Similarly,

$$\mathbb{E}[(Y^*)^2] = \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}} g(x)^{-\frac{2}{\gamma}}\,dx \tag{3.146}$$

$$\leqslant \int_{\mathcal{X}_f} (1+f(x)) g_{\min}^{-\frac{2}{\gamma}}\,dx \tag{3.147}$$

$$= (\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f)+1) g_{\min}^{-\frac{2}{\gamma}} \tag{3.148}$$

$$< \infty \tag{3.149}$$

Thus, both $\mathbb{E}[Y^*]$ and $\mathbb{E}[(Y^*)^2]$ are finite when $\gamma \geqslant 2$ (since we rely on $1-\frac{2}{\gamma} \geqslant 0$); but by definition $\mathscr{c}_\zeta(x) \geqslant \mathscr{c}^*(x)$ and hence if $\gamma \geqslant 2$, we have $\mathscr{c}_\zeta(x)^{-1} \leqslant \mathscr{c}^*(x)^{-1}$ and so $\mathbb{E}[Y] \leqslant \mathbb{E}[Y^*] < \infty$ and $\mathbb{E}[Y^2] \leqslant \mathbb{E}[(Y^*)^2] < \infty$ and hence both $\mathbb{E}[Y]$ and $\mathrm{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \leqslant \mathbb{E}[(Y^*)^2] - (1-\delta)^2\mathbb{E}[Y^*]^2$ are finite.

When $\gamma < 2$, we then use the fact that by construction $\mathscr{c}_\zeta(x) \geqslant \alpha_1$, and hence $\mathbb{E}[\,\mathscr{c}_\zeta(X)^{-1}] \leqslant \alpha_1^{-1}$ and $\mathrm{Var}[\,\mathscr{c}_\zeta(X)^{-1}] \leqslant \alpha_1^{-2}/4$ (since $\mathscr{c}_\zeta(X)^{-1} \in (0, \alpha_1^{-1}]$ its variance is at most $\alpha_1^{-2}/4$). Thus, we are done. □

Returning to the main proof, we first address step (3.128):

Lemma 17. *For any $\delta_1 > 0$, there is a sufficently small $\lambda > 0$ such that (when $X_i \overset{iid}{\sim} f$),*

$$\mathbb{P}\Big[\exists\, \pi \in \Pi^{\{X_i\}} \ s.t. \ \sum_{i=1}^{n} \ell_{j(i)} < (1-\delta_1)\lambda n \mathbb{E}[Y]\Big] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y]+(1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{3.150}$$

for all sufficiently large n.

Proof. For any $x \in \mathcal{X}_f$ let

$$\ell^*(x;\lambda) = \min_{\pi' \in \Pi} (\ell(\pi') : \ell^{\mathscr{c}}(\pi') = \lambda, x \in \bar{\pi}') \tag{3.151}$$

i.e. the length of the shortest possible cost-λ trajectory through x.

Recall that for any $\eta > 0$, there is some $\lambda_\eta^* > 0$ such that for any $0 < \lambda \leqslant \lambda_\eta^*$,

$$\mathbb{P}_{X\sim f}[\exists\, \pi' \in \Pi : X \in \bar{\pi}',\ \ell^{\mathscr{c}}(\pi') = \lambda \text{ and } \ell(\pi') < (1-\eta)\lambda/\,\mathscr{c}_\zeta(X)] \leqslant \eta \tag{3.152}$$

We define $\mathcal{X}(\eta) \subseteq \mathcal{X}_f$ to be the region in which

$$x \in \bar{\pi}',\ \ell^{\mathfrak{c}}(\pi') = \lambda \implies \ell(\pi') \geqslant (1-\eta)\lambda / \mathfrak{c}_\zeta(x) \tag{3.153}$$

for $\lambda \leqslant \lambda^*_\eta$ breaks down, i.e.

$$\mathcal{X}(\eta) = \{x \in \mathcal{X}_f : \exists\, \pi' \in \Pi : x \in \bar{\pi}',\ \ell^{\mathfrak{c}}(\pi') \leqslant \lambda^*_\eta \text{ and } \ell(\pi') < (1-\eta)\ell^{\mathfrak{c}}(\pi')/\mathfrak{c}_\zeta(x)\} \tag{3.154}$$

Then, if we have some $x \in \mathcal{X}_f$ and a trajectory π' s.t. $x \in \bar{\pi}'$ and $\ell^{\mathfrak{c}}(\pi') = \lambda < \lambda^*_\eta$,

$$\ell(\pi') \geqslant \begin{cases} (1-\eta)\lambda'/\mathfrak{c}_\zeta(x) & \text{if } x \notin \mathcal{X}(\eta) \\ 0 & \text{if } x \in \mathcal{X}(\eta) \end{cases} \tag{3.155}$$

Therefore, we define the random variables

$$Z_i := Z_i^{(\eta)} = \begin{cases} (1-2\eta)\lambda/\mathfrak{c}_\zeta(X_i) & \text{if } X_i \notin \mathcal{X}(\eta) \\ 0 & \text{if } X_i \in \mathcal{X}(\eta) \end{cases} \tag{3.156}$$

Since these are iid (depending only on iid X_i) let Z be the generic version. We know from (3.123) that

$$\lambda' \geqslant \lambda - 1/k \geqslant (1 - \beta n^{-(1-\frac{1}{\gamma})})\lambda \tag{3.157}$$

$$\implies (1-\eta)\lambda' \geqslant (1-2\eta)\lambda \quad \text{for} \quad n \geqslant (\beta/\eta)^{1+\frac{1}{\gamma-1}} \tag{3.158}$$

Since $\ell_{j(i)}$ is a cost-λ' trajectory passing through X_i, we know (fixing η) that for any sufficiently large n

$$\ell_{j(i)} \geqslant Z_i \text{ for all } i \tag{3.159}$$

Additionally, since $\mathbb{P}_{X\sim f}[X \in \mathcal{X}(\eta)] \leqslant \eta$ and because $\mathbb{E}[Y] = \mathbb{E}_{X\sim f}[\,\mathfrak{c}_\zeta(X)^{-1}]$ and

$\mathbb{E}[Y^2] = \mathbb{E}_{X\sim f}[\,\phi_\zeta(X)^{-2}]$ are finite, we have

$$\lim_{\eta\to 0}\mathbb{E}[Z^{(\eta)}] = \lambda\mathbb{E}[Y] \quad \text{and} \quad \lim_{\eta\to 0}\mathrm{Var}[Z^{(\eta)}] = \lambda\mathrm{Var}[Y] \tag{3.160}$$

Therefore, for any $\delta_1 > 0$ we can select some $\eta > 0$ such that

$$\mathbb{E}[Z^{(\eta)}] \geqslant (1-\delta_1/2)\lambda\mathbb{E}[Y] \text{ and } \mathbb{E}[Z^{(\eta)}] \leqslant (1+\delta_1)\lambda\mathrm{Var}[Y] \tag{3.161}$$

Fixing such η, we apply Proposition 11 (noting that Z_i's are nonnegative iid) to get

$$\mathbb{P}\Big[\sum_{i=1}^n Z_i \leqslant (1-\delta_1/2)n\mathbb{E}[Z]\Big] \leqslant e^{-\frac{n\mathbb{E}[Z]^2(\delta_1/2)^2/2}{\mathrm{Var}[Z]+\mathbb{E}[Z]^2(\delta_1/2)/3}} \tag{3.162}$$

$$\implies \mathbb{P}\Big[\sum_{i=1}^n Z_i \leqslant (1-\delta_1)\lambda n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y]+(1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{3.163}$$

$$\implies \mathbb{P}\Big[\sum_{i=1}^n \ell_{j(i)} \leqslant (1-\delta_1)\lambda n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y]+(1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{3.164}$$

for all $\lambda < \lambda_\eta^*$, which holds since we can apply

$$\mathbb{E}[Z]^2 \geqslant (1-\delta_1/2)^2\lambda\mathbb{E}[Y] \geqslant (1-\delta_1)\lambda\mathbb{E}[Y] \quad \text{and} \quad \mathrm{Var}[Z] \leqslant (1+\delta_1)\mathrm{Var}[Y] \tag{3.165}$$

(because of the $\mathrm{Var}[Z] > 0$ term, the first substitution on the numerator and the denominator makes the fraction smaller, hence the exponential of the negative of the fraction larger). □

But this means that with very high probability, given the CBO bound and applying (3.127) (and using $1+\delta_1 < (1+\delta_1/2)^2$)

$$\sum_{i=1}^n \ell_{j(i)} > (1-\delta_1/2)\lambda n\mathbb{E}[Y] \tag{3.166}$$

$$\implies \ell(\pi) \geqslant \frac{1-\delta_1/2}{\beta\lambda n^{\frac{1}{\gamma}}}\sum_{i=1}^n \ell_{j(i)} > (1-\delta_1)\beta^{-1}n^{1-\frac{1}{\gamma}}\mathbb{E}[Y] \tag{3.167}$$

with very high probability. Specifically (taking into account the probability that the

CBO bound holds) we have

$$\mathbb{P}\big[\mathrm{TSP}_{\Pi}(\{X_i\}) \geqslant (1-\delta_1)\beta^{-1}n^{1-\frac{1}{\gamma}}\mathbb{E}[Y]\big] \tag{3.168}$$

$$\geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta_1/2}} - e^{-\frac{n(1-\delta_1/2)\mathbb{E}[Y]^2\delta_1^2/32}{(1+\delta_1/2)\mathrm{Var}[Y]+(1-\delta_1/2)\mathbb{E}[Y]^2\delta_1/12}} \tag{3.169}$$

We then do a few simplifications on the exponent of the final term (ignoring the constant multiple of n and the negative to make the expression cleaner):

$$\frac{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}{(1+\frac{\delta_1}{2})\mathrm{Var}[Y]+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}} = \frac{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}{(1+\frac{\delta_1}{2})(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2)+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}} \tag{3.170}$$

$$= \left(\frac{(1+\frac{\delta_1}{2})(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2)+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}}{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}\right)^{-1} \tag{3.171}$$

$$= \left(\frac{1+\frac{\delta_1}{2}}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} + \frac{(1-\frac{\delta_1}{2})\frac{\delta_1}{12}-(1+\frac{\delta_1}{2})}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\right)^{-1} \tag{3.172}$$

This can be left as it is, but if we have $\mathbb{E}[Y^2]/\mathbb{E}[Y]^2$ and δ_1 is very small, we can write it in its most significant terms as

$$\left(\frac{1+\frac{\delta_1}{2}}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} + \frac{(1-\frac{\delta_1}{2})\frac{\delta_1}{12}-(1+\frac{\delta_1}{2})}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\right)^{-1} = \frac{\delta_1^2}{32(\mathbb{E}[Y^2]/\mathbb{E}[Y]^2-1)} + o(\delta_1^3) \tag{3.173}$$

$$= \frac{\delta_1^2}{32(\mathrm{Var}[Y]/\mathbb{E}[Y]^2)} + o(\delta_1^3) \tag{3.174}$$

Then, given a desired approximation factor $1-\delta$, we choose $\delta = \delta_1/2$ and a regularization term $\zeta > 0$ producing

$$\mathbb{E}[Y] \geqslant (1-\delta_2)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \tag{3.175}$$

where $\delta_2 = 1 - (1-\delta)/(1-\delta/2)$, i.e. such that

$$1 - \delta = (1 - \delta/2)(1 - \delta_2) = (1 - \delta_1)(1 - \delta_2) \tag{3.176}$$

Thus, our bound becomes

$$(1 - \delta_1)\beta^{-1} n^{1-\frac{1}{\gamma}} \mathbb{E}[Y] = (1 - \delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{3.177}$$

and substituting $\delta_1 = \delta/2$ into the very-high-probability bound yields

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \geqslant (1 - \delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx\right] \tag{3.178}$$

$$\geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/4}} - e^{-\frac{n(1-\delta/4)\mathbb{E}[Y]^2\delta^2/128}{(1+\delta/4)\mathrm{Var}[Y]+(1-\delta/4)\mathbb{E}[Y]^2\delta/24}} \tag{3.179}$$

$$\geqslant 1 - e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} - e^{-\left(\frac{\delta^2}{128(\mathrm{Var}[Y]/\mathbb{E}[Y]^2)} + o(\delta^3)\right)n} \tag{3.180}$$

when n is sufficiently large. When $\gamma \geqslant 2$, this can be made to be arbitrarily close to

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \geqslant (1 - \delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx\right] \tag{3.181}$$

$$\geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}} - e^{-\frac{n(1-\delta/2)\mathbb{E}[Y^*]^2\delta^2/32}{(1+\delta/2)\mathrm{Var}[Y^*]+(1-\delta/2)\mathbb{E}[Y^*]^2\delta/12}} \tag{3.182}$$

$$\geqslant 1 - e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} - e^{-\left(\frac{\delta^2}{32(\mathrm{Var}[Y^*]/\mathbb{E}[Y^*]^2)} + o(\delta^3)\right)n} \tag{3.183}$$

where $\mathbb{E}[Y^*]$ and $\mathrm{Var}[Y^*]$ are $\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx$ and $\int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}} g(x)^{-\frac{2}{\gamma}} \, dx$ respectively, with the downside that this may increase how large n needs to be in order to be 'sufficiently large' (in particular, it's achieved by taking $\delta_1 \to \delta$ and $\delta_2 \to 0$ to compensate, which requires cost regularization factor ζ to go to 0). For sufficiently large n the dominant error term is always $-e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}}$ since the other error term's exponent scales with n, but when δ is very small the value of n at which this term takes over may be extremely large.

Finally, we can state it in the following maximally simple terms: for any $\delta > 0$, there is some sufficiently small $\lambda > 0$ and sufficiently large $n^{(\delta)}$ such that for all

$n \geqslant n^{(\delta)}$,

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \geqslant 1-(1+\delta)e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}} \tag{3.184}$$

$$\geqslant 1-(1+\delta)e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{3.185}$$

with the caveat that as $\delta \to 0$, $n^{(\delta)}$ potentially goes to ∞.

Chapter 4

Symmetric Dynamics Algorithm

In this chapter we discuss our algorithm for the DSTSP with symmetric dynamics, and the bound it achieves for the TSP. This algorithm relies on a division of $\mathcal{X}_f$ into Symmetric Hierarchical Cell Structures (see Section 2.3), which are nested cell structures satisfying certain key properties which allow them to be used easily to generate a TSP tour; we call this division the *Symmetric Hierarchical Cell Structure cover* (Symmetric HCS cover), which then transforms the TSP into an abstract problem on a tree which we call the *Symmetric Hierarchical Collection Problem* (Symmetric HCP or SHCP); this problem then admits an efficiently-computable optimal algorithm and a clean upper bound.

Remark 16. *The optimal solution to the SHCP, translated via the SHCS cover to the TSP, does* not *constitute an optimal TSP tour; this is because the SHCP's objective function is really an upper bound to the TSP tour it translates to.*

We organize this chapter as follows:

- In Section 4.1 we define the Symmetric HCS cover, and show that if Assumption 10 holds (for every $x \in \mathcal{X}_f$, there exists an SHCS containing x), there is always a Symmetric HCS cover.
- In Section 4.2 we show how the DTSP within a Symmetric HCS induces the Symmetric HCP; solving the SHCP then yields a viable tour for the vehicle

through the given target points as well as an upper bound to the length of this tour. Finally, we show an upper bound to the SHCP (dependent on its parameters and the number of targets).

- In Section 4.3 we show how the Symmetric HCS cover and the SHCP bound derived in Section 4.2 combine to produce our (probabilistic) bound for the Symmetric DSTSP.

4.1 Symmetric Hierarchical Cell Structure Cover

In this section we show how the existence of a Symmetric Hierarchical Cell Structure (Symmetric HCS or SHCS) at any $x \in \mathcal{X}_f$ (see Assumption 10) allows us to construct a more intricate structure, the *Symmetric Hierarchical Cell Structure Cover* (SHCS cover), which is then used by our algorithm. In brief, an SHCS cover is a fixed, finite set of SHCS's which cover $\mathcal{X}_f$ and don't overlap too much. We define:

Definition 21. For $\rho \geqslant 0$, a *ρ-accurate Symmetric Hierarchical Cell Structure cover* (SHCS cover) of $\mathcal{X}_f$ is a set $S_1, \ldots, S_m$ of SHCS's such that:

- All SHCS's have the same parameters (ε_0, α, ζ, s).
- $\mathcal{X}_f \subseteq \bigcup_{j=1}^{m} S_j$.
- $\mathbb{P}_{X \sim f}[X \text{ is in more than one } S_j] \leqslant \rho$.

Any SHCS with overlap parameter α necessarily satisfies the condition for any overlap parameter $\alpha' \leqslant \alpha$; thus, if the SHCS's have different values of α, we can use the minimum.

Assumption 10 implies the existence of SHCS covers of arbitrarily good accuracy. First, we have the existence of a fixed-scale finite SHCS cover (possibly with bad accuracy ρ):

Lemma 18. *There exist α, s such that for all $\zeta > 0$, there is some ε_0 such that there exists a finite SHCS cover (of some accuracy parameter ρ) of $\mathcal{X}_f$.*

Proof. By Assumption 10 each $x \in \mathcal{X}_f$ is contained in the interior of some SHCS S_x. Thus, letting S_x^{int} be the interior of S_x, we know that $\mathcal{X}_f \subseteq \bigcup_x S_x^{\text{int}}$. But by Assumption 5, $\mathcal{X}_f$ is compact, and hence there is a finite subset of these S_x's whose interiors cover $\mathcal{X}_f$. □

Then, given a finite SHCS covering, for any $\rho > 0$ we can produce a ρ-accurate covering at arbitrarily small scales by taking the sub-SHCS's recursively until we reach the desired scale, and deleting any redundant ones.

Lemma 19. *If Assumption 10 is satisfied with parameters α, s, then for any $\rho > 0$ and $\zeta > 0$, there is some ε_0 such that a ρ-accurate ζ-regularized SHCS cover $S_1, \ldots, S_m$ can be constructed. Furthermore, for any $\varepsilon_0^* > 0$, there is some $\varepsilon_0 \leqslant \varepsilon_0^*$ for which it exists.*

Proof. We use the fact that by definition a ε_0-scale Hierarchical Cell Structure is composed of s^γ smaller ε_0/s-scale (sub)-Hierarchical Cell Structures, which are then each divisible again and so forth. We also use the fact that within each SHCS, the sub-SHCS's are all disjoint, and that any scale-ε SHCS is contained in some ball in $\mathcal{X}$ with radius $c_\Pi \varepsilon$.

By Lemma 18 we have a SHCS cover $S_1^{(\text{start})}, \ldots, S_{m^{(\text{start})}}^{(\text{start})}$ at some scale $\varepsilon_0^{(\text{start})}$ which are ζ-regular with efficiency and scaling parameters α, s; it may have a very large accuracy parameter $\rho > 0$ but since it is a cover we know that $\mathcal{X}_f \subseteq \bigcup_{j=1}^{m^{(\text{start})}} S_j^{(\text{start})}$. We then will use this to create an alternative covering $S_1, \ldots, S_m$ (at a smaller scale ε_0) which in addition to being ζ-regular is also ρ-accurate. For any Jordan-measurable subset $A \subseteq \mathcal{X}_f$, let $\text{Int}(A)$ denote its interior and

$$\text{Int}_\eta(A) := \{x \in A : d_{\mathcal{X}}(x, x') \geqslant \eta \text{ for all } x' \notin A\} \tag{4.1}$$

i.e. the η-interior of A (points at least η distance away from anything outside of A). We then note the following (all sets involved are Jordan-measurable):

a. $\text{Vol}_{\mathcal{X}}(A) = \text{Vol}_{\mathcal{X}}(\text{Int}(A)) = \lim_{\eta \to 0} \text{Vol}_{\mathcal{X}}(\text{Int}_\eta(A))$.

b. For any $A_1, \ldots, A_k \subseteq \mathcal{X}$, letting $A = \bigcap_{j=1}^{k} A_j$, we have $\mathrm{Int}_\eta(A) = \bigcap_{j=1}^{k} \mathrm{Int}_\eta(A_j)$ and $\mathrm{Int}(A) = \bigcap_{j=1}^{k} \mathrm{Int}(A_j)$. Since intersections of finitely many Jordan-measurable sets are Jordan-measurable, we also have $\mathrm{Vol}_{\mathcal{X}}(A) = \mathrm{Vol}_{\mathcal{X}}(\mathrm{Int}(A))$.

We now subdivide each $S_j^{(\mathrm{start})}$ them k times, for any integer k to get a cover of $m := m^{\mathrm{start}} s^{\gamma k}$ SHCS's at scale $\varepsilon_0 := \varepsilon_0^{(\mathrm{start})} s^{-k}$. Let these be denoted $S_1, \ldots, S_m$; let $S_1^{(j)}, \ldots, S_{s^{\gamma k}}^{(j)}$ denote the SHCS's which were created by subdividing $S_j^{(\mathrm{start})}$. For each $S_i^{(j)}$, we then remove it if there is some $j' < j$ such that $S_i^{(j)} \subseteq S_{j'}^{(\mathrm{start})}$, i.e. if it is completely contained within a different SHCS from our starting cover with a lower index. This is a simple way to remove redundant SHCS's, and preserves the covering property; to see this, we can perform this action starting with the descendants of $S_{m^{(\mathrm{start})}}^{(\mathrm{start})}$, then the descendants of $S_{m^{(\mathrm{start})}-1}^{(\mathrm{start})}$ and so forth. We now claim that as long as we set k sufficiently large, we can reduce the redundancy (that is, the region contained within multiple SHCS's) to an arbitrarily small set. Let

$$S_k(j, j') := \{x : x \in S_i^{(j)} \cap S_{i'}^{(j')} \text{ for some } i, i'\} \tag{4.2}$$

i.e. the region in which descendants of $S_j^{(\mathrm{start})}$ and $S_{j'}^{(\mathrm{start})}$ overlap, and its volume, given k subdivisions (after removing redundant descendants as per the above steps). Without loss of generality let $j' < j$. Note that $S_0(j, j') = S_j^{(\mathrm{start})} \cap S_{j'}^{(\mathrm{start})}$, and that

$$S_k(j, j') \subseteq S_{k'}(j, j') \text{ for all } k > k' . \tag{4.3}$$

We now let

$$\eta := 2c_\Pi \varepsilon_0 = 2c_\Pi \varepsilon_0^{(\mathrm{start})} s^{-k} \tag{4.4}$$

and claim that

$$S_k(j, j') \cap \mathrm{Int}_\eta(S_0(j, j')) = \varnothing \tag{4.5}$$

that is, that deleting redundant cells will progressively 'hollow out' the interior of

$S_0(j, j')$. This happens because for the root $\boldsymbol{q}$ of cell $S_i^{(j)}$ at x (i.e. $x = [\boldsymbol{q}]_{\mathcal{X}}$)

$$S_i^{(j)} \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}) \subseteq \mathcal{B}_{c_\Pi \varepsilon_0}(x) \tag{4.6}$$

This means that any cell $S_i^{(j)}$ rooted at x is necessarily deleted if $x \in \mathrm{Int}_{c_\Pi \varepsilon_0}(S_0(j, j'))$; and this additionally means that if $S_i^{(j)}$ *contains* x where $x \in \mathrm{Int}_{2c_\Pi \varepsilon_0}(S_0(j, j')) = \mathrm{Int}_\eta(S_0(j, j'))$, then $S_i^{(j)}$ is removed as well. Thus we get (4.5), which further implies

$$S_k(j, j') \subseteq S_0(j, j') \backslash \mathrm{Int}_\eta(S_0(j, j'))\,. \tag{4.7}$$

But since $S_0(j, j') = S_j^{(\mathrm{start})} \cap S_{j'}^{(\mathrm{start})}$, it is Jordan-measurable and hence the volume of the above goes to 0 as $\eta \to 0$, which happens as $k \to \infty$. But this means if we let the total region of overlaps be

$$S'_k := \{x \in \mathcal{X}_f : x \in S_i^{(j)} \cap S_{i'}^{(j')} \text{ for some } i, j, i', j' \text{ where } j \neq j'\} \tag{4.8}$$

we get that

$$S'_k = \bigcup_{j \neq j'} S_k(j, j') \tag{4.9}$$

$$\implies \mathrm{Vol}_{\mathcal{X}}(S'_k) \leqslant \sum_{j \neq j'} \mathrm{Vol}_{\mathcal{X}}(S_k(j, j')) \tag{4.10}$$

$$\implies \lim_{k \to \infty} \mathrm{Vol}_{\mathcal{X}}(S'_k) \leqslant \lim_{k \to \infty} \sum_{j \neq j'} \mathrm{Vol}_{\mathcal{X}}(S_k(j, j')) = 0 \tag{4.11}$$

Thus for any $\rho > 0$, we can set k sufficiently large so that

$$\mathbb{P}_{X \sim f}[X \in S'_k] \leqslant \rho \tag{4.12}$$

which means that we have a ρ-accurate cover.

Finally, once we have it at scale ε_0, we can scale it down by as many factors of s as we like, and in particular until the radius is smaller than the given ε_0^*. □

In our upper bound proofs the remainder of this chapter (and the corresponding

proofs in Chapter 6), we will assume that the SHCS cover we base our algorithm on is 0-accurate, i.e. there is no overlap at all between different SHCS's.

Finally, we can combine this with the result from the main text (in which the remaining $\leqslant n$ target points are good) to achieve the same bound; this is because we can collect the good target points and then the bad target points with only an additional constant (added for the arc between the last good target visited and the first bad target).

Remark 17. *For regular symmetric control-affine dynamics (on which we can invoke the Ball-Box Theorem) and continuous density functions f we can find SHCS's with $s = 2$. However, we make a broader definition allowing $s \neq 2$ so that targets distributed within sets of fractal dimension, which may scale more naturally with some other s (e.g. the Menger Sponge naturally scales with $s = 3$) may be analyzed in future work.*

4.2 The Symmetric Hierarchical Collection Problem

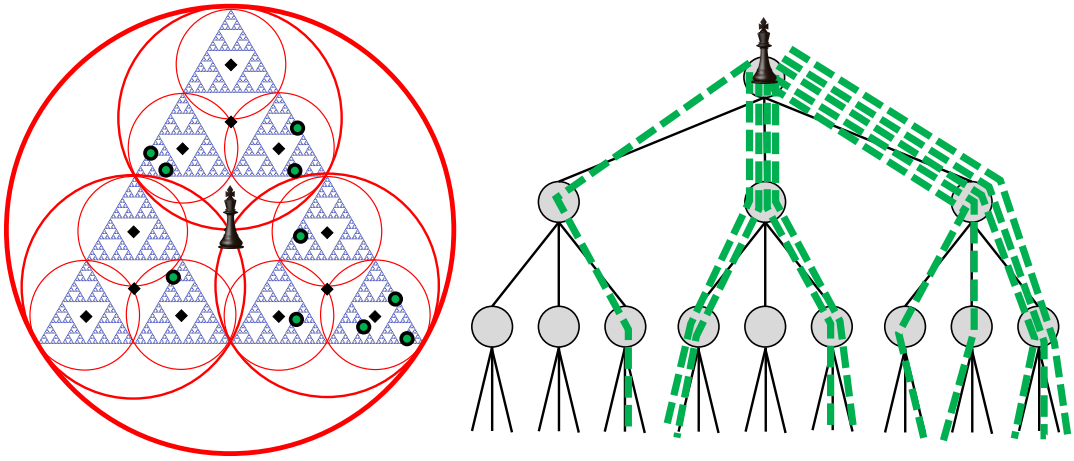

Figure 4-1: A Symmetric Hierarchical Cell Structure (on a fractal target set, chess king representing the vehicle, cells in red) with targets, and its corresponding Symmetric Hierarchical Collection Problem with targets represented by dashed lines. Subcells from left to right in the SHCS correspond to children from left to right in the SHCP.

Unlike in the general DSTSP algorithms, however, the symmetric property of

the control system allows a tile to be completely cleared of target points before the algorithm moves on to the next one, since after visiting the one target the vehicle can return to the anchor and go to the next one and so on.

For any $0 < \varepsilon_0 < \varepsilon^*$ (where ε^* is the value given in Assumption 8), $\boldsymbol{q}_0 \in \mathcal{Q}$, and $\delta > 0$, we define the *ε_0-scale hierarchical cell structure rooted at $\boldsymbol{q}_0$* as an infinite rooted tree $T(\boldsymbol{q}_0, \varepsilon_0)$ in the following way:

1. The nodes represent reachable sets / cells, each with an anchor $\boldsymbol{q}$ and a radius ε. We denote the node as $v_{(\boldsymbol{q},\varepsilon)}$, which represents the reachable set $\bar{R}_\varepsilon(\boldsymbol{q})$.

2. The children of $v_{(\boldsymbol{q},\varepsilon)}$ are

$$C(v_{(\boldsymbol{q},\varepsilon)}) = \{v_{(\boldsymbol{q}^{(1)},\varepsilon/2)}, \ldots, v_{(\boldsymbol{q}^{(b)},\varepsilon/2)}\} \quad \text{satisfying} \quad R_\varepsilon(\boldsymbol{q}) \subseteq \bigcup_{j=1}^{b} R_{\varepsilon/2}(\boldsymbol{q}^{(j)}) \tag{4.13}$$

   as given in Assumption 8. We also assume WLOG that

$$R_\varepsilon(\boldsymbol{q}) \cap R_{\varepsilon/2}(\boldsymbol{q}^{(j)}) \neq \varnothing \quad \text{for all} \quad j \tag{4.14}$$

   (if not, we remove $v_{(\boldsymbol{q}^{(j)},\varepsilon/2)}$ from the tree).

If $v' \in C(v)$ we say $v = P(v')$ (the parent).

We define the *layers* of the tree in the normal way, with the root $v_{(\boldsymbol{q}_0,\varepsilon_0)}$ being layer 0, its children being layer 1, and so forth (note that all nodes at layer k represent reachable sets with radius $\varepsilon_0/2^k$). We denote these layers as

$$T_k := \{v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0, \varepsilon_0) : \varepsilon = \varepsilon_0/2^k\}. \tag{4.15}$$

For any $x \in \mathcal{X}$ and $v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0, \varepsilon_0)$, we say '$x \in v_{(\boldsymbol{q},\varepsilon)}$' if $x \in \bar{R}_\varepsilon(\boldsymbol{q})$ (slightly abusing notation by conflating the node v and the region it represents).

Lemma 20. *For any $T(\boldsymbol{q}_0, \varepsilon_0)$ the following hold:*

i. $|T_k| \leqslant b^k$.

ii. For any $v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0, \varepsilon_0)$ and any $v_{(\boldsymbol{q}',\varepsilon/2)} \in C(v_{(\boldsymbol{q},\varepsilon)})$ (any parent-child pair),

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant 3\varepsilon/2\,. \tag{4.16}$$

iii. For any $v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0, \varepsilon_0)$,

$$d_\Pi(\boldsymbol{q}_0, \boldsymbol{q}) \leqslant 3\varepsilon_0 \tag{4.17}$$

Proof. We show these in order:

(i) $T_k = \bigcup_{v \in T_{k-1}} C(v)$ and $|C(v)| \leqslant b$, so $|T_k| \leqslant b|T_{k-1}|$; recursing back to $|T_0| = |\{\boldsymbol{q}_0\}| = 1$ yields the result.

(ii) Noting that $R_\varepsilon(\boldsymbol{q}) \cap R_{\varepsilon/2}(\boldsymbol{q}') \neq \varnothing$ (by (4.16), let $\boldsymbol{q}'' \in R_\varepsilon(\boldsymbol{q}) \cap R_{\varepsilon/2}(\boldsymbol{q}')$. Then by definition $d_\Pi(\boldsymbol{q}, \boldsymbol{q}'') \leqslant \varepsilon$ and $d_\Pi(\boldsymbol{q}', \boldsymbol{q}'') \leqslant \varepsilon/2$ and (since the control system Π is symmetric, d_Π is a metric) we get by the triangle inequality that

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant d_\Pi(\boldsymbol{q}, \boldsymbol{q}'') + d_\Pi(\boldsymbol{q}', \boldsymbol{q}'') \leqslant 3\varepsilon/2 \tag{4.18}$$

(iii) This follows from (ii): let $v_{(\boldsymbol{q},\varepsilon)} \in T_k$, so $\varepsilon = \varepsilon_0/2^k$. This means that, letting $\boldsymbol{q}_k := \boldsymbol{q}$, we can find $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k-1}$ such that $v_j := v_{(\boldsymbol{q}_j, \varepsilon_0/2^j)} \in C(v_{(\boldsymbol{q}_{j-1}, \varepsilon_0/2^{j-1})}) = C(v_{j-1})$ for all $j = 1, 2, \ldots, k$ (tracing the path from the root $v_0 = v_{(\boldsymbol{q}_0, \varepsilon_0)}$ to $v_{(\boldsymbol{q},\varepsilon)}$). Then by (ii) we have

$$d_\Pi(\boldsymbol{q}_0, \boldsymbol{q}_k) \leqslant \sum_{j=1}^{k} d_\Pi(\boldsymbol{q}_{j-1}, \boldsymbol{q}_j) \leqslant \sum_{j=1}^{k} 3\varepsilon_0 2^{-j} \leqslant 3\varepsilon_0 \tag{4.19}$$

Hence we are done. □

Let n target points $X_1, \ldots, X_n$ be in $\bar{R}_{\varepsilon_0}(\boldsymbol{q}_0)$ (the area covered by this tree essentially). Each target point $X_1, \ldots, X_n$ is contained (at least) one of the nodes at level k (at each k). Specifically, for any $i \in [n], k \in \mathbb{Z}_{\geqslant 0}$ let $v_k(X_i) := v_{(\boldsymbol{q}_k(X_i), \varepsilon_0/2^k)} \in T(\boldsymbol{q}_0, \varepsilon_0)$ satisfying the following properties:

- $X_i \in v_k(X_i)$ for all i, k (i.e. $X_i \in \bar{R}_{\varepsilon_0/2^k}(\boldsymbol{q}_k(X_i))$);

- $v_0(X_i) = v_{(\boldsymbol{q}_0,\varepsilon_0)}$ is the root for all i;
- $v_k(X_i) \in C(v_{k-1}(X_i))$ for all $i \in [n]$ and $k \in \mathbb{Z}_{>0}$.

Note that this means $v_k(X_i) \in T_k$. Such a set of nodes exists because of the structure of $T(\boldsymbol{q}_0, \varepsilon_0)$, specifically that the reachable sets of the nodes in $C(v)$ cover the reachable set of v.

This suggests a 'depth-first-search' type algorithm for visiting $X_1, \ldots, X_n$. For any $v \in T_k$ let

$$n(v) := |\{i : v_k(X_i) = v\}| \tag{4.20}$$

i.e. the number of target points X_i whose layer-k node is v. We travel from node to node along the tree $T(\boldsymbol{q}_0, \varepsilon_0)$. If we describe the vehicle as being 'at' node $v_{(\boldsymbol{q},\varepsilon_0/2^k)} \in T_k$, it refers to being at $\boldsymbol{q}$, from which we can do three things:

1. travel to its parent $P(v)$ (taking at most $3\varepsilon_0 2^{-k}$ time);
2. travel to a child $v' \in C(v)$ (taking at most $3\varepsilon_0 2^{-(k+1)}$ time);
3. visit some target point $X_i \in v$ and returning (taking at most $\varepsilon_0 2^{-(k-1)}$ time since $X_i \in \bar{R}_{\varepsilon_0/2^k}(\boldsymbol{q})$).

This allows us to state our TSP as an abstract problem similar to the classic TSP on graphs (but slightly different); since all distance upper bounds above are multiples of ε_0, we ignore it (and re-insert it after we find the solution).

Definition 22. The *Symmetric Hierarchical Collection Problem* (Symmetric HCP or SHCP) with cell branching factor $\bar{b}$ and scaling factor s is an abstract problem on an infinite rooted tree $T = (V, E)$ where every $v \in V$ has $\bar{b}$ children. We define:

1. $\gamma = \log_s(\bar{b})$, which defines the relationship between s and $\bar{b}$ (in the Symmetric Hierarchical Cell Structure this problem models, scaling the cells down by a factor of s results in a $\bar{b}$-factor increase in the number needed to cover the same region).

2. The root of T is v_0; the children of $v \in V$ are denoted as the set $C(v)$ and its parent is $P(v)$ and the edges incident to v is denoted E_v.

3. The *level* of a node is its distance from v_0 and is denoted $k(v)$ (so $k(v_0) = 0$ and for all $v \in V$ and $v' \in C(v)$ we have $k(v') = k(v) + 1$); the set of nodes of level k is denoted $L_k := \{v \in V : k(v) = k\}$.

4. Each edge $e = (v, v') \in E$ has a cost

$$w(e) = s^{-\min(k(v),k(v'))} \tag{4.21}$$

   i.e. if e connects level k to level $k+1$ then its cost is s^{-k}.

Finally, we have n *targets* $\tau_1, \ldots, \tau_n$; these are not nodes on the tree but are infinitely long paths starting at the root and continuing down the tree. We say $v \in \tau_i$ if v is in the path τ_i, and $v_k(\tau_i)$ is the level-k node of τ_i, and we assume that $\tau_i \neq \tau_j$ (i.e. they don't represent the same infinite path) for all $i, j \in [n]$.

The problem is solved by a *player* who moves around T and collects targets. When at $v \in V$, the player chooses one of two actions, for an associated cost:

1. move along some $e \in E_v$ for cost $w(e)$ (i.e. going up costs $s^{-(k-1)}$, going down costs s^{-k});

2. collect a target τ_i such that $v \in \tau_i$ for cost $2 \cdot s^{-k}$.

Then, starting at v_0, the goal is to collect all n targets τ_i and return to v_0 for minimum total cost. We denote the minimum cost to collect $\tau_1, \ldots, \tau_n$ as $\mathrm{HCP}(\tau_1, \ldots, \tau_n; \bar{b}, s)$, and the maximum cost to collect n targets (i.e. $\tau_1, \ldots, \tau_n$ are adversarially chosen) as

$$\mathrm{HCP}^*(n; \bar{b}, s) := \max_{\tau_1,\ldots,\tau_n} \mathrm{HCP}(\tau_1, \ldots, \tau_n; \bar{b}, s)\,. \tag{4.22}$$

The targets $\tau_1, \ldots, \tau_n$ in the SHCP represent which cells the targets $X_1, \ldots, X_n$ fall into in the symmetric DTSP (we refer here to the DTSP and not the DSTSP

because the targets will be treated as given, not randomly distributed, for now).

Remark 18. *The SHCP isn't exactly equivalent to the original symmetric DTSP because*

a. all the costs are derived from upper bounds to the equivalent costs in the DTSP;

b. the DTSP doesn't require starting or ending at any particular point; and (most importantly)

c. it restricts possible solutions to only those that follow the hierarchical structure.

The general tradeoff that this problem explores is that collecting targets gets cheaper as the player moves down the tree T, but this movement has costs and the targets can get more dispersed as the tree branches out (note that all the targets are present at the root, though collecting them is the most expensive there). This mirrors the TSP on the cell structure in the following way:

Proposition 12. Let $S : \mathcal{Q} \times (0, \varepsilon^*]$ be a hierarchical cell structure with cell branching factor $\bar{b}$, scaling factor s, and efficiency factor α (see Definition 11) and let $X_1, \ldots, X_n$ be targets in $S(\boldsymbol{q}_0, \varepsilon_0)$. We then set up an infinite rooted tree $T = (V, E)$ with $\bar{b}$ children at every node with the following correspondence: every node $v \in V$ represents a cell $S(\boldsymbol{q}, \varepsilon)$ for appropriate values of $\boldsymbol{q}, \varepsilon$, such that

i. the root v_0 represents $S(\boldsymbol{q}_0, \varepsilon_0)$;

ii. if v represents $S(\boldsymbol{q}, \varepsilon)$, then its children $v_1, \ldots, v_{\bar{b}} \in C(v)$ represent (in some order) the sets $S(\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\bar{b})}, \varepsilon/s)$ which cover $S(\boldsymbol{q}, \varepsilon)$ (as given in Definition 11).

Let $\boldsymbol{q}(v)$ be the anchor of the cell represented by v, and $\varepsilon(v)$ be its radius.

Each target τ_i corresponding to X_i is a path down the tree starting at the root v_0 and always choosing a node representing $S(\boldsymbol{q}, \varepsilon)$ containing X_i (if there are multiple such paths, choose one arbitrarily).

Then, the following bound holds:

$$\mathrm{TSP}_{\Pi}(X_1, \ldots, X_n) \leqslant \varepsilon_0 \mathrm{HCP}(\tau_1, \ldots, \tau_n; b, s) \leqslant \varepsilon_0 \mathrm{HCP}^*(n; b, s)\,. \tag{4.23}$$

Proof. The first inequality follows because $\varepsilon(v) = \varepsilon_0 s^{-k(v)}$ (at each level the scale is divided by s). Thus, we can equate a solution to the SHCP instance (moving on the tree and collecting targets) with an equivalent solution on the TSP: at each step where we are at $v \in V$ in the SHCP, we are at $\boldsymbol{q}(v) \in \mathcal{Q}$ in the TSP. If $v' \in C(v)$, then

$$d_\pi(\boldsymbol{q}(v), \boldsymbol{q}(v')) \leqslant \varepsilon(v) = \varepsilon_0 s^{-k(v)} \tag{4.24}$$

i.e. moving from $\boldsymbol{q}(v)$ to $\boldsymbol{q}(v')$ takes time at most $\varepsilon_0 w(e)$ where $e = (v, v') \in E$. Furthermore,

$$v \in \tau_i \implies X_i \in S(\boldsymbol{q}(v), \varepsilon(v)) \implies d_\pi(\boldsymbol{q}(v), X_i) \leqslant \varepsilon_0 s^{-k(v)} \tag{4.25}$$

and hence visiting X_i from $\boldsymbol{q}(v)$ and then returning takes at most $2\varepsilon_0 s^{-k(v)}$ time, or, in other words at most ε_0 times the equivalent cost in the SHCP. Thus every solution to the SHCP yields a corresponding TSP trajectory whose length is at most ε_0 times the cost, thus showing the first inequality.

The second inequality then follows from the definition of $\mathrm{HCP}^*(n; \bar{b}, s)$. □

We now define a few notions which will help us discuss the SHCP.

Definition 23. For $v \in T$, $n_v := |\{i : v \in \tau_i\}|$, i.e. the number of targets τ_i passing through $v \in T$.

A *plan* θ for a SHCP instance $\tau_1, \ldots, \tau_n$ is a list of actions which, when performed starting at the root v_0, collects all the targets τ_i and returns to v_0. The cost of the plan is $c_{\mathrm{cost}}(\theta)$, the set of vertices entered is V_θ, and θ is *optimal* if there is no plan θ' such that

i. $c_{\mathrm{cost}}(\theta') < c_{\mathrm{cost}}(\theta)$; or

ii. $c_{\text{cost}}(\theta') = c_{\text{cost}}(\theta)$ and $|V_{\theta'}| < |V_\theta|$.

The extra condition for 'optimality' (to enter as few vertices as possible) is not really a feature of the optimization problem represented by the SHCP in general but is made for convenience, particularly for Lemma 22. We now consider two questions: (i) what is the optimal algorithm for the SHCP?; (ii) what is the upper bound $\text{HCP}^*(n; b, s)$?

Lemma 21. *In any optimal plan, every edge $e \in E$ is crossed either 0 or 2 times.*

Proof. Let $e = (v, v')$, where v is on the same side of e as the root v_0 and v' is on the opposite side. First, since the player must start and end at v_0, every edge is crossed an even number of times. Then, if an edge is crossed more than twice, it must happen an even number $2m$ times. Let these happen at steps $s_1 < s_2 < \cdots < s_{2m}$, and let S_i denote the actions taken between s_i and s_{i+1} (and S_0, S_{2m} denote respectively the actions before s_1 beginning at v_0 and the actions after s_{2m} ending at v_0). Note that $S_1, S_3, \ldots, S_{2m-1}$ all begin and end at v' while $S_2, S_4, \ldots, S_{2m-2}$ all begin and end at v, and finally S_0 begins at v_0 and ends at v and S_{2m} begins at v and ends at v_0. Then we can produce a modified solution which performs S_0, crosses e, does $S_1, S_3, \ldots, S_{2m-1}$ without crossing e, then crosses back over e and performs $S_2, S_4, \ldots, S_{2m}$ and ends at v_0. Since costs do not depend on the step at which they are performed, the costs of $S_0, S_1, \ldots, S_{2m}$ all remain the same, and all targets are collected, but instead of crossing e a total of $2m$ times it is crossed only twice, thus reducing the cost of the plan. □

Lemma 22. *A vertex $v \neq v_0$ is entered in an optimal plan iff $n_v \geqslant \frac{s}{s-1}$.*

Proof. We first prove that if $n_v < \frac{s}{s-1}$ then an optimal plan never enters v. Suppose we have a plan θ that enters v; it must do so from $P(v)$, and let $k = k(P(v))$. We then consider a modification θ' of this plan which, rather than entering v, collects all τ_i going through v at $P(v)$. The only difference between θ and θ' is that θ travels to v and around its subtree and makes its collections there, while θ' makes all collections at $P(v)$, so we can compare only the costs of these actions. Then:

- θ' incurs costs of $2s^{-k}n_v$ to collect all the relevant targets;

- letting $k' \geqslant 1$ be the number of levels deeper than k that θ goes, the costs incurred are at least

$$2s^{-(k+k')}n_v + 2\sum_{j=0}^{k'-1} s^{-(k+j)} \tag{4.26}$$

(collection costs, minimized if they are maximally deep, and minimum possible movement costs to reach level $k + k'$ and return).

But then

$$n_v < \frac{s}{s-1} \implies n_v < 1 + \frac{n_v}{s} \tag{4.27}$$

and we can then collapse (4.26) because:

$$2s^{-(k+k')}n_v + 2\sum_{j=0}^{k'-1} s^{-(k+j)} = 2s^{-k}\Big(s^{-k'}n_v + \sum_{j=0}^{k'-1} s^{-j}\Big) \tag{4.28}$$

$$> 2s^{-k}\Big(s^{-(k'-1)}n_v + \sum_{j=0}^{k'-2} s^{-j}\Big) \tag{4.29}$$

$$> \ldots > 2s^{-k}n_v \tag{4.30}$$

where each step takes the last entry in the sum and the n_v term and applies (4.27). But this means that θ' is a lower-cost plan than θ, so any plan that enters v can be improved by avoiding v and collecting earlier.

Now suppose $n_v \geqslant \frac{s}{s-1}$, and suppose we have a plan θ which does not enter n_v, and let $k = k(v)$. Let v' be the closest ancestor to v that θ enters, and let $k' = k(v) - k(v')$ (how many levels above v plan θ gets). Then all the targets τ_i that go through v are most efficiently collected at v' and incur cost

$$2s^{-k(v')} = 2s^{-(k-k')} \tag{4.31}$$

Alternatively, we can consider plan θ' which instead of collecting at v' go to v and

collect there. This incurs cost

$$2s^{-k(v)}n_v + 2\sum_{j=0}^{k'-1} s^{-(k(v')+j} = 2s^{-k}n_v + 2\sum_{j=0}^{k'-1} s^{-(k-k'+j)} \tag{4.32}$$

$$= 2s^{-k}\Big(n_v + \sum_{j=1}^{k'} s^{-j}\Big) \tag{4.33}$$

$$\leqslant 2s^{-k}s^{k'} = 2s^{-(k-k')} \tag{4.34}$$

where the sum collapses because at each step we apply

$$n_v \geqslant \frac{s}{s-1} \implies s\,n_v \geqslant s + n_v \tag{4.35}$$

Hence, as before, the modified plan θ' costs at most as much as θ (and visits more vertices) so θ cannot be optimal. □

Thus, we have our optimal algorithm: assuming $n \geqslant \frac{s}{s-1}$ (if not, the optimal algorithm is to simply collect all targets at the root), we consider the subset of vertices

$$V_{\geqslant \frac{s}{s-1}} := \Big\{v \in V : n_v \geqslant \frac{s}{s-1}\Big\} \tag{4.36}$$

which must include the root and must be connected since

$$n_v \geqslant \frac{s}{s-1} \implies n_{P(v)} \geqslant n_v \geqslant \frac{s}{s-1}\,. \tag{4.37}$$

Then the optimal plan is to do a depth-first-search tour of this subgraph (which is a finite tree) using every edge exactly twice and collecting every target τ_i at the deepest vertex at which it is present.

Proposition 13. For any $n, \bar{b}, s$ such that $s \geqslant 2$ and $\gamma = \log_s(\bar{b}) \geqslant 2$,

$$\mathrm{HCP}^*(n; \bar{b}, s) \leqslant 6sn^{1-\frac{1}{\gamma}} \tag{4.38}$$

Proof. We consider a plan that does the following: it takes a tour to each vertex at some level k^*; at each $v \in L_{k*}$, it collects all targets $\tau_i \in v$. Note that this plan has a

fixed cost: it must traverse all edges within the first k^* levels twice, and each target collected costs $2s^{-k^*}$.

Thus the total cost of movement is

$$2\sum_{k=0}^{k^*-1} \bar{b}(\bar{b}/s)^k = 2\bar{b}\frac{(\bar{b}/s)^{k^*}-1}{(\bar{b}/s)-1} = 2\frac{(\bar{b}/s)^{k^*}-1}{(1/s)-(1/\bar{b})} \tag{4.39}$$

because between levels k and $k+1$ there are $\bar{b}^{k+1}$ edges, each with cost s^{-k}, which must be traversed twice, and the total cost of collection is $2ns^{-k^*}$ since we need to collect n targets at a cost of $2s^{-k^*}$ each.

Now we let $k^* = \lceil \log_{\bar{b}} n \rceil - 1$. This is the deepest level at which there must still be at least two targets at the same vertex. In this case, we have

$$(\bar{b}/s)^{k^*} \leqslant (\bar{b}/s)^{\log_{\bar{b}} n} = ns^{-\log_{\bar{b}} n} = n^{1-\frac{1}{\gamma}} \tag{4.40}$$

as $s^{-\log_{\bar{b}} n} = s^{-\log_{s^\gamma} n} = n^{-\frac{1}{\gamma}}$. Furthermore, since $s \geqslant 2$ and $\gamma \geqslant 2$, we know that $\bar{b} \geqslant 2s$ and so $(1/s) - (1/\bar{b}) \geqslant 1/(2s)$ and so our movement cost is bounded by

$$2\frac{(\bar{b}/s)^{k^*}-1}{(1/s)-(1/\bar{b})} \leqslant 4sn^{1-\frac{1}{\gamma}}\,. \tag{4.41}$$

Our collection cost is additionally

$$2ns^{-k^*} \leqslant 2sns^{-\log_{\bar{b}} n} = 2sn^{1-\frac{1}{\gamma}}\,. \tag{4.42}$$

Adding these together gives a total cost of $\leqslant 6sn^{1-\frac{1}{\gamma}}$. □

Remark 19. *While the abstraction provided by the SHCP is useful for providing an upper bound and an algorithm for the Symmetric DSTSP, a number of common-sense improvements can be made for the most common control systems. These include the following:*

- *The abstraction ignores the possibility that cells which are far from each other in the structure (i.e. to go from one to the other requires moving far back up the tree and then back down) might be very close in the space. These 'horizontal'*

connections can make it possible to traverse through a number of cells of the same level k without having to return to larger cells close to the root.

- *As discussed in Remark 18, you can ignore the requirement to start and end at the root (which we made in order to have nicer theorems and proofs).*

- *If you have a good point-to-point motion planner (specifically, configuration-to-configuration), you can first use an SHCP algorithm to determine the order in which to visit the targets and the configuration $\boldsymbol{q} \in [X_i]^{\mathcal{Q}}$ in which to visit each target X_i and then use the point-to-point motion planner for each pair of consecutive configurations. If the point-to-point motion planner is computationally efficient enough you can even use the cell structure to determine the configurations for each target point (depending on which cell it ends up in, each target point needs to be in a certain set of configurations to be reachable from the anchor of the cell) and then use a standard TSP approximation algorithm on the complete graph where the edge weights are the point-to-point distances.*

4.3 The Symmetric HCP and the TSP

We now consider what happens if the target space $\mathcal{X}_f$ can be covered (with no overlap) by a finite set of m SHCS's of a given scale ε_0 (not dependent on the number of targets n). In this case, we can consider an algorithm that solves the SHCP induced in each SHCS separately, visiting all the targets within the given SHCS, then strings these together to produce a tour which visits all the target points. Such a tour can be thought of as consisting of an alternation between two types of trajectories: (i) trajectories within a single SHCS, visiting all the targets and returning to the root; (ii) trajectories between SHCS roots. We note that the total tiime needed for type (ii) segments is fixed and doesn't depend on n (it consists of a tour through the roots of the SHCS's). Thus, if ε_0 is fixed, we can treat it as an (additive) constant and focus only on type (i).

We then solve each individual SHCS, which we denote as $S_1, S_2, \ldots, S_m$. To get

an upper bound to the tour, we apply the upper bound given by Proposition 13 for the induced SHCP; note that n will be replaced by the number of targets n_j which fall into S_j. The sum of these bounds over all j then yields an upper bound on the total time needed for type (i) segments in the tour. However, note that n_j is now a random variable, since each random target may either fall in or out of S_j.

This yields a probabilistic problem. Suppose there are m bins, and n balls will be thrown in them iid according to some probability vector $\boldsymbol{p} = (p_1, \ldots, p_m)$; let n_j be the number of balls thrown into the jth bin, and note that its distribution is $\mathrm{Bin}(p_j, n)$. Let

$$\nu := 1 - \frac{1}{\gamma} \in (0,1) \quad \text{and} \quad Y_j := n_j^\nu \text{ and } Y = \sum_{j=1}^{m} Y_j \tag{4.43}$$

This models an upper bound to the Symmetric DSTSP in the following way: we cover the space with m ε_0-scale hierarchical cell structures $S_1, \ldots, S_m$ with scaling factor s and cell branching factor $\bar{b} = s^\gamma$; for each S_j, let $S'_j \subseteq S_j$ so that $S'_1, \ldots, S'_m$ partition $\mathcal{X}_f$ (each $x \in \mathcal{X}_f$ is in exactly one S'_j). Then let $p_j := \mathbb{P}_{X\sim f}[X \in S'_j]$; thus, each target $X_i \sim f$ falls into one of the m 'bins' according to probability vector $\boldsymbol{p}$, so we can let n_j be the number of targets in $S'_j \subseteq S_j$. Then, by Proposition 13, the time it takes to visit all the n_j targets in S'_j is at most

$$6s\varepsilon_0 n^{1-\frac{1}{\gamma}} = 6s\varepsilon_0 n_j^\nu = 6s\varepsilon_0 Y_j \tag{4.44}$$

To visit all n targets thus takes at most

$$C + \sum_{j=1}^{m} 6s\varepsilon_0 Y_j = C + 6s\varepsilon_0 Y \tag{4.45}$$

where C is a constant (in n) which depends on the choices of the S_j, representing the time it takes for the vehicle to tour the roots of the hierarchical cell structures. This is an upper bound for the Symmetric DSTSP trajectory length because it represents (an upper bound to) the length of the path generated by the hierarchical cell algorithm,

but the algorithm is not necessarily optimal.

We now show a few properties of Y:

Lemma 23. $\mathbb{E}[Y_j] \leqslant (p_j n)^\nu$ *and* $\mathbb{E}[Y] \leqslant \sum_{j=1}^m (p_j n)^\nu = \left(\sum_{j=1}^m p_j^\nu\right) n^\nu$

Proof. Note that $\cdot^\nu$ is a concave function because $\nu \in (0,1)$, and that $\mathbb{E}[n_j] = p_j n$. Thus, by Jensen's Inequality,

$$\mathbb{E}[Y_j] = \mathbb{E}[n_j^\nu] \leqslant (\mathbb{E}[n_j])^\nu = (p_j n)^\nu \tag{4.46}$$

and the result from $\mathbb{E}[Y]$ follows by summing the $\mathbb{E}[Y_j]$. □

Given the logic above (and some analysis on what m and $\boldsymbol{p}$ are as $\varepsilon_0 \to 0$), this alone gives an upper bound to the expected value of expected trajectory length in the Symmetric DSTSP. However, we also want concentration bounds.

We will consider Y as a Doob martingale: for any $i = 0, 1, \ldots, n$, let $Z_i \in [m]$ be the bin that ball i falls into, and consider placing them one-by-one in index order (i.e. deciding Z_1 first, then Z_2, etc). We then define our (centered) Doob martingale:

$$Y_j(i) = \mathbb{E}[Y_j \,|\, Z_1, \ldots, Z_i] - \mathbb{E}[Y_j] \quad \text{and} \quad Y(i) = \sum_{j=1}^m Y_j(i) = \mathbb{E}[Y \,|\, Z_1, \ldots, Z_i] - \mathbb{E}[Y] \tag{4.47}$$

Then note that by definition $Y_j(0) = 0$ and $Y_j(n) = Y_j - \mathbb{E}[Y_j]$ (and similarly with Y), and that $Y(i)$ is a martingale with respect to i, i.e.

$$\mathbb{E}[Y(i) \,|\, Y(i-1)] = Y(i-1) \tag{4.48}$$

(as is well-known from the construction as a conditional expectation of a process, which is known as a Doob martingale). For convenience, we also let $n_j(i)$ denote the number of balls in bin j after Z_i has been decided, i.e.

$$n_j(i) = \sum_{i'=1}^{i} \mathbf{1}\{Z_{i'} = j\}\,. \tag{4.49}$$

We then show a few results concerning $Y(i)$:

Lemma 24. $|Y(i) - Y(i-1)| \leqslant 1$.

Proof. Consider first what happens when Z_i is changed to Z_i'. Let $Z^{(n)} = (Z_1, \ldots, Z_n)$ and let $Z_{-i}^{(n)}(Z_i') = Z_1, \ldots, Z_{i-1}, Z_i', Z_{i+1}, Z_n$ i.e. $Z^{(n)}$ with Z_i replaced by Z_i. Then let $Y(Z^{(n)})$ be the value of Y given $Z_1, \ldots, Z_n$. For simplicity let $Z_i = j$ and $Z_i' = j'$.

$$|Y(Z^{(n)}) - Y(Z_{-i}^{(n)}(Z_i'))| = |n_j^\nu + n_{j'}^\nu - (n_j - 1)^\nu - (n_{j'} + 1)^\nu| \tag{4.50}$$

$$= |(n_j^\nu - (n_j - 1)^\nu) - ((n_{j'} + 1)^\nu - n_{j'}^\nu)| \leqslant 1 \tag{4.51}$$

Each of the two differences are in $[0, 1]$ and hence their difference is in $[-1, 1]$. Thus changing the value of any Z_i can only change the final value by 1 (in either direction). Thus, for any Z_i, Z_i',

$$|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']| = \mathbb{E}_{Z_{i+1}, \ldots, Z_n}[|Y(Z^{(n)}) - Y(Z_{-i}^{(n)}(Z_i'))|] \tag{4.52}$$

$$\leqslant 1 \tag{4.53}$$

But then we get that

$$|Y(i) - Y(i-1)| = |\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}]| \tag{4.54}$$

$$= |\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}_{Z_i'}\big[\mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']\big]| \tag{4.55}$$

$$\leqslant \mathbb{E}_{Z_i'}\big[|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']|\big] \tag{4.56}$$

$$\leqslant 1 \tag{4.57}$$

□

Thus, we can immediately apply Azuma-Hoeffding to get a bound; unfortunately, to get a good result this requires that $\gamma > 2$ (i.e. $\nu = 1 - \frac{1}{\gamma} > \frac{1}{2}$). However, we will give the result and sharpen it afterwards:

Lemma 25. *Given the definitions above,*

$$\mathbb{E}\left[Y(n) \geqslant n^{\nu} \sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^{m} p_j^{\nu})^2}{2n}} = e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^{m} p_j^{\nu})^2} \tag{4.58}$$

Furthermore, $\sum_{j=1}^{m} p_j^{\nu} \geqslant 1$*, so we can even find a bound independent of* $\boldsymbol{p} = (p_1, \ldots, p_m)$*:*

$$\mathbb{E}\left[Y(n) \geqslant n^{\nu} \sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}} \tag{4.59}$$

Proof. This is a direct application of the Azuma-Hoeffding Inequality with the bounded differences as shown in Lemma 24.

We know that $\sum_{j=1}^{m} p_j^{\nu} \geqslant 1$ because $\boldsymbol{p}$ is a probability vector and can therefore be expressed as a convex mixture of the basis vectors $\mathbf{1}_1, \ldots, \mathbf{1}_m$. Then we note that the function $\theta(\boldsymbol{p}) = \sum_{j=1}^{m} p_j^{\nu}$ is concave and $\theta(\mathbf{1}_j) = 1$ for all j, and hence by Jensen's Inequality

$$\theta(\boldsymbol{p}) \geqslant \sum_{j=1}^{m} p_j \theta(\mathbf{1}_j) = 1 \tag{4.60}$$

and thus we have the result we wanted. □

Proposition 14. Given $Y = \sum_{j=1}^{m} n_j^{\nu}$, we have

$$\mathbb{P}\left[Y \geqslant 2n^{\nu} \sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^{m} p_j^{\nu})^2} \tag{4.61}$$

which implies the following bound independent of $\boldsymbol{p}$:

$$\mathbb{P}\left[Y \geqslant 2n^{\nu} \sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}} \tag{4.62}$$

Proof. Using the definition of the centered martingale $Y(i)$, we have $Y(n) = Y - \mathbb{E}[Y]$,

thus giving us:

$$Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu \implies Y - n^\nu \sum_{j=1}^m p_j^\nu \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{4.63}$$

$$\implies Y - \mathbb{E}[Y] \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{4.64}$$

$$\implies Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{4.65}$$

which in turn means by Lemma 25

$$\mathbb{P}\left[Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant \mathbb{P}\left[Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^m p_j^\nu)^2} \tag{4.66}$$

The second bound independent of $\boldsymbol{p}$ follows as in the proof of Lemma 25. □

Since we have $-\frac{1}{2}n^{2\nu-1}$ in the exponent, this is a very high probability concentration bound when $\nu > 1/2$, which happens iff $\gamma > 2$. When $\gamma \leqslant 2$, this bound is not high probability (let alone very high probability), so we need more precise methods.

One place where we lost some potential benefit is in Lemma 24: while this is the best single constant bound that holds under all circumstances, it ignores the fact that when i is not very close to n, there are a lot of $Z_{i'}$ for $i < i'$ which still need to be allocated; this in turn means that each bin has some expected number of balls still to be added, so adding a ball to any given bin *now* won't change $\mathbb{E}[n_j^\nu \mid Z_1, \ldots, Z_i]$ by very much (as the differences between k^ν and $(k+1)^\nu$ get smaller and smaller as k gets larger). First, we need the following lemma:

Lemma 26. *Let $W \sim \mathrm{Bin}(p, n)$ for some $p \in (0,1)$ and n, and $\nu \in (0,1)$; then*

$$\mathbb{E}[(W+1)^\nu] - \mathbb{E}[W^\nu] \leqslant e^{-\frac{3}{28}np} + 2\nu(np)^{\nu-1} \tag{4.67}$$

Proof. Since $\nu \in (0,1)$, $\cdot^\nu$ is concave, and so the first-order approximation is an upper

bound, yielding

$$(W+1)^\nu - W^\nu \leqslant W^\nu + \nu W^{\nu-1} - W^\nu = \nu W^{\nu-1} \tag{4.68}$$

While this is fine for $W > 0$, we have to be careful for the case of $W = 0$, since the exponent $\nu - 1$ is negative; however, when $W = 0$, we have $(W+1)^\nu - W^\nu = 1$, so we have the bound

$$(W+1)^\nu - W^\nu \leqslant \min(\nu W^{\nu-1}, 1) \tag{4.69}$$

since for $W \geqslant 1$, we have $\nu W^{\nu-1} \leqslant \nu < 1$.

We then note the following: for $W \leqslant np/2$, we have $(W+1)^\nu - W^\nu \leqslant 1$, and for $W > np/2$, we have $(W+1)^\nu - W^\nu \leqslant \nu(np/2)^{\nu-1}$. Thus, we have

$$\begin{aligned}
\mathbb{E}[(W+1)^\nu] - \mathbb{E}[W^\nu] &= \mathbb{E}[(W+1)^\nu - W^\nu] && (4.70)\\
&\leqslant \mathbb{P}[W \leqslant np/2] + \mathbb{P}[W > np/2](\nu(np/2)^{\nu-1}) && (4.71)\\
&\leqslant \mathbb{P}[W \leqslant np/2] + 2\nu(np)^{\nu-1} && (4.72)
\end{aligned}$$

We can then use Bernstein's Inequality to bound $\mathbb{P}[W \leqslant np/2]$, as W is a binomial random variable and hence a sum of iid Bernoulli random variables. In particular, we let these iid Bernoullis be W_i ($i \in [n]$) and we let

$$W_i' = p - W_i \quad \text{(which also means } W_i = p - W_i') \tag{4.73}$$

which are zero-mean and $|W_i'| \leqslant 1$ always. Thus,

$$W = \sum_{i=1}^n W_i = np - \sum_{i=1}^n W_i' \tag{4.74}$$

i.e. $W \leqslant np/2 \iff \sum_{i=1}^n W_i' \geqslant np/2$. Thus, since the W_i' have variance $p(1-p)$, we

have

$$\mathbb{P}\left[\sum_{i=1}^{n} W_i' \geqslant \frac{np}{2}\right] \leqslant e^{-\frac{(np)^2/8}{np(1-p)+np/6}} = e^{-\frac{np/8}{(1-p)+1/6}} \leqslant e^{-\frac{3}{28}np} \tag{4.75}$$

and hence we can put it all together and get

$$\mathbb{E}[(W+1)^\nu] - \mathbb{E}[W^\nu] \leqslant e^{-\frac{3}{28}np} + 2\nu(np)^{\nu-1} \tag{4.76}$$

□

Lemma 27. *For the martingale defined above, WLOG* $\boldsymbol{p}$ *be sorted, i.e.* $0 < p_1 \leqslant \cdots \leqslant p_m$ *(any* 0 *entries can be ignored and* $\boldsymbol{p}$ *reduced). Then*

$$|Y(i) - Y(i-1)| \leqslant \min(e^{-\frac{3}{28}(n-i)p_1} + 2\nu((n-i)p_1)^{\nu-1}, 1) \tag{4.77}$$

Proof. We consider the situation when Z_i is about to be decided, i.e. $Z_1, \ldots, Z_{i-1}$ are known and $Z_{i+1}, \ldots, Z_n$ are in the future. How much can $\mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}]$ differ from $\mathbb{E}[Y \mid Z_1, \ldots, Z_i]$? We note that

$$\left|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}]\right| \tag{4.78}$$

$$= \left|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}_{Z\sim\boldsymbol{p}}[\mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z]]\right| \tag{4.79}$$

$$\leqslant \max_{Z\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z'\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z'] \tag{4.80}$$

(note that we are working with Y which is not centered, but the centering subtracts $\mathbb{E}[Y]$ from both sides and can be ignored). That is, we can bound $|Y(i) - Y(i-1)|$ given $Z_1, \ldots, Z_{i-1}$ by taking the difference between the choice of Z_i that maximizes $Y(i)$ and the choice of Z_i that minimizes $Y(i)$. Since $Z_1, \ldots, Z_{i-1}$ are unknown, we also have to maximize over them, i.e. we want to bound

$$\max_{Z_1,\ldots,Z_{i-1}} (\max_{Z\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z'\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z']) \tag{4.81}$$

We now break this down further: setting $Z = j$ really adds 1 to the final n_j, and

$Z' = j'$ does the same to $n_{j'}$; the remaining $j'' \neq j, j'$ are unaffected and cancel out. Note also that $Z_1, \ldots, Z_{i-1}$ only matter insofar as they affect $n_1(i-1), \ldots, n_m(i-1)$, so we will condition on those instead. Thus, letting $W_j \sim \mathrm{Bin}(p_j, n-i)$, we have :

$$\max_{Z_1,\ldots,Z_{i-1}} \left(\max_{Z\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z'\in[m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z']\right) \tag{4.82}$$

$$= \max_{n_j(i), n_{j'}(i)} \left(\mathbb{E}[(W_j + n_j(i) + 1)^\nu] - \mathbb{E}[(W_j + n_j(n))^\nu]\right. \tag{4.83}$$

$$\left. - \left(\mathbb{E}[(W_{j'} + n_{j'}(i) + 1)^\nu] - \mathbb{E}[(W_{j'} + n_{j'}(n))^\nu]\right)\right) \tag{4.84}$$

$$\leqslant \max_{n_j(i)} \mathbb{E}[(W_j + n_j(i) + 1)^\nu] - \mathbb{E}[(W_j + n_j(n))^\nu] \tag{4.85}$$

$$= \max_j \mathbb{E}[(W_j + 1)^\nu] - \mathbb{E}[W_j^\nu] \tag{4.86}$$

where the expectation is over the $n-i$ values $Z_{i+1}, \ldots, Z_n$ yet to be determined. This holds because $(a_1+1)^\nu - a_1^\nu \geqslant (a_2+1)^\nu - a_2^\nu$ if $a_1 \leqslant a_2$; and that also means that this is maximized if j is selected so p_j is minimized, i.e. we use p_1.

Then, by Lemma 26 (and the previously-derived upper bound of 1), the result follows, with $n-i$ because this is the number of $Z_{i'}$ yet to be determined. □

We now adjust the Azuma-Hoeffding argument for the new difference bound:

Proposition 15. Given the definitions above, let $\nu \leqslant 2/3$ and (wlog) let p_1 be the smallest nonzero value in $\boldsymbol{p} = (p_1, \ldots, p_m)$. Then, for all

$$n \geqslant \frac{280}{3} \log(1/\nu) p_1^{-1} \tag{4.87}$$

we have the following concentration bounds:

$$\mathbb{E}\left[Y(n) \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant \begin{cases} e^{-\frac{p_1 n^{2\nu} (\sum_{j=1}^m p_j^\nu)^2}{13 + \frac{8}{2\nu-1} n^{2\nu-1}}} & \text{for } \nu \in (1/2, 2/3) \\ e^{-\frac{(2/9) p_1 n (\sum_{j=1}^m p_j^{1/2})^2}{127 - \log(1/p_1) + \log(n)}} & \text{for } \nu = 1/2 \\ e^{-\frac{p_1 n^{2\nu} (\sum_{j=1}^m p_j^\nu)^2}{\frac{560}{3}\log(1/\nu) + 4 + 18\nu^2 \frac{1}{1-2\nu} (\frac{280}{3}\log(1/\nu))^{2\nu-1}}} & \text{for } \nu \in (0, 1/2) \end{cases} \tag{4.88}$$

As before, $\sum_{j=1}^{m} p_j^\nu \geqslant 1$, so we can remove it and the bound remains valid (if potentially somewhat looser). Writing in asymptotic notation (but noting that we have firm constants by the above) this is

$$\mathbb{E}\left[Y(n) \geqslant n^\nu \sum_{j=1}^{m} p_j^\nu\right] \leqslant \begin{cases} e^{-\Theta(p_1 n(2\nu-1)(\sum_{j=1}^{m} p_j^\nu)^2)} & \text{for } \nu \in (1/2, 2/3) \\ e^{-\Theta(p_1 (n/\log(n))(\sum_{j=1}^{m} p_j^\nu)^2)} & \text{for } \nu = 1/2 \\ e^{-\Theta(p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2(\log(1/\nu)+(1-2\nu)^{-1})^{-1})} & \text{for } \nu \in (0, 1/2) \end{cases} \tag{4.89}$$

Proof. Let

$$c_i = \min(e^{-\frac{3}{28}(n-i)p_1} + 2\nu((n-i)p_1)^{\nu-1}, 1) \tag{4.90}$$

We now want to compute $\sum_{i=1}^{n} c_i^2$. We first define

$$a_i = c_{n-i} = \min(e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1}, 1) \tag{4.91}$$

where i now goes from 0 to $n-1$. Then we know that

$$\sum_{i=0}^{n-1} a_i^2 = \sum_{i=1}^{n} c_i^2 \tag{4.92}$$

Let $z = ip_1$; then we want to compare $e^{-\frac{3}{28}z}$ to $2\nu z^{\nu-1}$. In particular we want to find z^* sufficiently large such that

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \text{ for all } z \geqslant z^* \tag{4.93}$$

We compute:

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \tag{4.94}$$

$$\iff -\frac{3}{28}z \leqslant (\nu-1)\log(z) + \log(2\nu) \tag{4.95}$$

$$\iff \frac{3}{28}z \geqslant (1-\nu)\log(z) - \log(2\nu) \tag{4.96}$$

Then note that if $z \geqslant \frac{28}{3}\frac{1}{1-\nu}$,

$$\frac{d}{dz}\frac{3}{28}z = \frac{3}{28} \geqslant (1-\nu)\frac{1}{z} = \frac{d}{dz}(1-\nu)\log(z) - \log(2\nu) \tag{4.97}$$

Thus, if $z^* \geqslant \frac{28}{3}\frac{1}{1-\nu}$ and

$$\frac{3}{28}z^* \geqslant (1-\nu)\log(z^*) - \log(2\nu) \tag{4.98}$$

then we can conclude that for all $z \geqslant z^*$,

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1}\,. \tag{4.99}$$

Noting that $1 < \frac{28}{3}\frac{1}{1-\nu} \leqslant 28$ (as $\nu \leqslant 2/3$), we let $z^* = \frac{280}{3}\log(1/\nu)$. Then

$$\frac{3}{28}z^* = 10\log(1/\nu) \geqslant (1-\nu)\log\left(\frac{280}{3}\log(1/\nu)\right) - \log(2\nu) \tag{4.100}$$

which holds for $\nu \in (0, 2/3)$, and $z^* \geqslant 28$ for this range as well. This means that for all $z \geqslant z^*$ (and $\nu \in (0, 2/3)$) we have

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1}\,. \tag{4.101}$$

Then, recalling that $z = ip_1$, we have for all $i \geqslant z^*/p_1 = \frac{280}{3}\log(1/\nu)p_1^{-1}$,

$$e^{-\frac{3}{28}ip_1} \leqslant 2\nu(i\,p_1)^{\nu-1} \implies e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1} \leqslant 3\nu(i\,p_1)^{\nu-1} \tag{4.102}$$

Thus, we have the following:

$$a_i = \min(e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1}, 1) \leqslant \begin{cases} 1 & \text{if } i \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} \\ 3\nu(i\,p_1)^{\nu-1} & \text{if } i \geqslant \frac{280}{3}\log(1/\nu)p_1^{-1} \end{cases} \tag{4.103}$$

We now assume that $n \geqslant \frac{280}{3}\log(1/\nu)p_1^{-1}$. In that case,

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + \sum_{i=\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} 9\nu^2(i\,p_1)^{2\nu-2} \tag{4.104}$$

$$\leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2}\int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\,dt \tag{4.105}$$

We now have to split this according to $\nu > 1/2$, $\nu = 1/2$, and $\nu < 1/2$. This yields the following: for $\nu \in (1/2, 2/3)$,

$$9\nu^2 p_1^{2\nu-2}\int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\,dt \tag{4.106}$$

$$= 9\nu^2 p_1^{2\nu-2}\frac{1}{2\nu-1}\left((n-1)^{2\nu-1} - \left(\frac{280}{3}\log(1/\nu)p_1^{-1}\right)^{2\nu-1}\right) \tag{4.107}$$

$$\leqslant 4p_1^{-1}\frac{1}{2\nu-1}\left(n^{2\nu-1} - \left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1}\right) \tag{4.108}$$

and hence we can conclude that (again, for $\nu \in (1/2, 2/3)$)

$$\sum_{i=0}^{n-1} a_i^2 \leqslant p_1^{-1}\left(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}\left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1}\right) + 2 + \frac{4}{2\nu-1}p_1^{-1}n^{2\nu-1} \tag{4.109}$$

$$= O(p_1^{-1}\log(1/\nu) + 1) + O(p_1^{-1}(2\nu-1)^{-1})n^{2\nu-1} \tag{4.110}$$

Then, for $\nu = 1/2$, we have

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2}\int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\,dt \tag{4.111}$$

$$= \frac{280}{3}\log(2)p_1^{-1} + 2 + (9/4)p_1^{-1}\left(\log(n-1) - \log\left(\frac{280}{3}\log(2)\right) + \log(1/p_1)\right) \tag{4.112}$$

$$\leqslant \left(\frac{280}{3}\log(2) - (9/4)\log\left(\frac{280}{3}\log(2)\right)\right)p_1^{-1} + 2 + (9/4)p_1^{-1}\big(\log(n) - \log(1/p_1)\big) \tag{4.113}$$

$$< (56 - (9/4)\log(1/p_1))p_1^{-1} + 2 + (9/4)p_1^{-1}\log(n) \tag{4.114}$$

And finally, for $\nu \in (0, 1/2)$, we have

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2}\int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\,dt \tag{4.115}$$

$$\leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2}\int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{\infty} t^{2\nu-2}\,dt \tag{4.116}$$

$$= \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2}\frac{1}{1-2\nu}\Big(\frac{280}{3}\log(1/\nu)p_1^{-1}\Big)^{2\nu-1} \tag{4.117}$$

$$= \left(\frac{280}{3}\log(1/\nu) + 2 + 9\nu^2\frac{1}{1-2\nu}\Big(\frac{280}{3}\log(1/\nu)\Big)^{2\nu-1}\right)p_1^{-1} \tag{4.118}$$

(note that this is constant in n).

We then plug it into Azuma-Hoeffding (keeping in mind that $\sum_{i=1}^{n} c_i^2 = \sum_{i=0}^{n-1} a_i^2$. We get: for $\nu \in (1/2, 2/3)$,

$$\mathbb{E}\left[Y(n) \geqslant 2n^\nu \sum_{j=1}^{m} p_j^\nu\right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{2\sum_{i=1}^n a_i^2}} \tag{4.119}$$

$$\leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{2p_1^{-1}(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}(\frac{280}{3}\log(1/\nu))^{2\nu-1}) + 4 + \frac{8}{2\nu-1}p_1^{-1}n^{2\nu-1}}} \tag{4.120}$$

$$= e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{2(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}(\frac{280}{3}\log(1/\nu))^{2\nu-1}) + 4p_1 + \frac{8}{2\nu-1}n^{2\nu-1}}} \tag{4.121}$$

$$\leqslant e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{9 + 4p_1 + \frac{8}{2\nu-1}n^{2\nu-1}}} \tag{4.122}$$

$$\leqslant e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{13 + \frac{8}{2\nu-1}n^{2\nu-1}}} \tag{4.123}$$

$$= e^{-\Theta(p_1 n(2\nu-1)(\sum_{j=1}^m p_j^\nu)^2)} \tag{4.124}$$

where the second-to-last step is simply from numerically taking the maximum of

$$\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}\Big(\frac{280}{3}\log(1/\nu)\Big)^{2\nu-1} \tag{4.125}$$

over the domain $\nu \in (1/2, 2/3)$.

For $\nu = 1/2$, we have

$$\mathbb{E}\left[Y(n) \geqslant 2n^{1/2}\sum_{j=1}^{m} p_j^{1/2}\right] \leqslant e^{-\frac{n(\sum_{j=1}^m p_j^{1/2})^2}{2\sum_{i=1}^n a_i^2}} \tag{4.126}$$

$$\leqslant e^{-\frac{n(\sum_{j=1}^m p_j^{1/2})^2}{2((56-(9/4)\log(1/p_1))p_1^{-1}+2+(9/4)p_1^{-1}\log(n))}} \tag{4.127}$$

$$= e^{-\frac{p_1 n(\sum_{j=1}^m p_j^{1/2})^2}{(28-(9/2)\log(1/p_1))+4p_1+(9/2)\log(n)}} \tag{4.128}$$

$$\leqslant e^{-\frac{(2/9)p_1 n(\sum_{j=1}^m p_j^{1/2})^2}{(126-\log(1/p_1))+p_1+\log(n)}} \tag{4.129}$$

$$\leqslant e^{-\frac{(2/9)p_1 n(\sum_{j=1}^m p_j^{1/2})^2}{127-\log(1/p_1)+\log(n)}} \tag{4.130}$$

$$= e^{-\Theta(p_1(n/\log(n))(\sum_{j=1}^m p_j^{\nu})^2)} \tag{4.131}$$

Finally, for $\nu \in (0, 1/2)$ we have

$$\mathbb{E}\left[Y(n) \geqslant 2n^{\nu}\sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^m p_j^{\nu})^2}{2\sum_{i=1}^n a_i^2}} \tag{4.132}$$

$$\leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^m p_j^{\nu})^2}{(\frac{560}{3}\log(1/\nu)+4+18\nu^2\frac{1}{1-2\nu}(\frac{280}{3}\log(1/\nu))^{2\nu-1})p_1^{-1}}} \tag{4.133}$$

$$= e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^{\nu})^2}{\frac{560}{3}\log(1/\nu)+4+18\nu^2\frac{1}{1-2\nu}(\frac{280}{3}\log(1/\nu))^{2\nu-1}}} \tag{4.134}$$

$$= e^{-\Theta(p_1 n^{2\nu}(\sum_{j=1}^m p_j^{\nu})^2(\log(1/\nu)+(1-2\nu)^{-1})^{-1})} \tag{4.135}$$

□

Note that the dependence on n in the above bounds (holding ν, p_1 fixed) is the following: $e^{-\Theta(n)}$ for $\nu \in (1/2, 2/3)$; $e^{-\Theta(n/\log(n))}$ for $\nu = 1/2$; and $e^{-\Theta(n^{2\nu})}$ for $\nu \in (0, 1/2)$.

Barring fractals (which are of interest for future work) we have integer $\gamma \geqslant 2$ and hence $1 - \frac{1}{\gamma} = \nu \in (0, 2/3) \implies \nu = 1/2$, i.e. when considering control-affine systems and full-dimensional continuous distributions f of target points, our main concern is $\nu = 1/2$. For $\nu \geqslant 2/3$, Proposition 14 provides a much cleaner very high probability concentration bound (whose dependence on n is $e^{-\Theta(n^{2\nu-1})} \geqslant e^{-\Theta(n^{1/6})}$), without any fuss about how large n has to be or the p_1 term.

4.4 Symmetric HCS covering algorithm

We now consider a division of $\mathcal{X}_f$ into m different ε_0-scale hierarchical cell structures, as given in Assumption 10; we call their roots $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$. Let $x_j = [\boldsymbol{q}_j]_{\mathcal{X}}$. These will remain fixed while the number of targets $n \to \infty$. Let $S_1, S_2, \ldots, S_m$ be the top-level cells of the SHCS's that cover $\mathcal{X}_f$ (at scale ε_0), and let $x_j = [\boldsymbol{q}_j]_{\mathcal{X}}$ for all j; we define

$$p_j := \mathbb{P}_{X \sim f}[X \in S(\boldsymbol{q}_j, \varepsilon_0)] \tag{4.136}$$

Note that p_j depend on which SHCS's we use, which in turn depends on ε_0. For this proof, we will assume that the sets $S(\boldsymbol{q}_j, \varepsilon_0)$ are disjoint (see Appendix D for a proof that this simplifying assumption does not affect the result). Thus, $\sum_{j=1}^m p_j = 1$, and distributing $X \sim f$ puts it into a cell S_j according to probability vector $\boldsymbol{p} = (p_1, \ldots, p_m)$. Let $j(X)$ denote the index of the cell X falls into.

Then we wish to show that each SHCS covers a certain amount of probability mass, except a negligible fraction:

Lemma 28. *For any $\rho_1 > 0$, let a ρ_1-*good *cell S_j be defined as one such that*

$$p_j \geqslant (1 - \rho_1)\alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^{\gamma}\,. \tag{4.137}$$

Then, for any $\rho_1, \rho_2 \geqslant 0$, for any sufficiently small $\zeta < 0$ there is some ε_0^* such that for any $0 < \varepsilon_0 \leqslant \varepsilon_0^*$ there are $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$ satisfying the conditions of Assumption 10 such that

$$\mathbb{P}_{X \sim f}[S_{j(X)} \textit{ is } \rho_1\textit{-good}] \geqslant 1 - \rho_2\,. \tag{4.138}$$

Proof. Consider some threshold $\rho^* > 0$ and call a cell S_j ρ^*-*sufficient* if $\check{f}^{(\zeta)}(x_j) \geqslant \rho^*$. We claim the following:

i. For any $\rho_2 > 0$, there is some $\rho^* > 0$ such that for all sufficiently small $\zeta > 0$, $\mathbb{P}_{X \sim f}[\check{f}^{(\zeta)}(X) \leqslant (3/2)\rho^*] \leqslant \rho_2$.

ii. For any $\rho^* > 0$, there is some ε_0^* such that if $\check{f}^{(\zeta)}(X) > (3/2)\rho^*$, then $S_{j(X)}$ must be ρ^*-sufficient.

iii. For any $\rho^*, \rho_1, \zeta > 0$, there is a sufficiently small ε_0^* such that for all $0 < \varepsilon_0 \leqslant \varepsilon_0^*$, any ρ^*-sufficient cell is ρ_1-good.

Claim (i) holds because f is continuous almost everywhere, and therefore $\lim_{\zeta\to 0} \check{f}^{(\zeta)}(x) = f(x)$ almost everywhere; thus we pick ρ^* sufficiently small so that $\mathbb{P}_{X\sim f}[f(X) \leqslant (3/2)\rho^*] \leqslant p_2/2$, and then letting $\zeta \to 0$ we get

$$\lim_{\zeta\to 0} \mathbb{P}_{X\sim f}[\check{f}^{(\zeta)}(X) \leqslant (3/2)\rho^*] = \mathbb{P}_{X\sim f}[f(X) \leqslant (3/2)\rho^*] \leqslant p_2/2 \tag{4.139}$$

which means that for sufficiently small $\zeta > 0$, we have $\mathbb{P}_{X\sim f}[\check{f}^{(\zeta)}(X) \leqslant (3/2)\rho^*] \leqslant \rho_2$.

Claim (ii) holds because by definition

$$X \in S_{j(X)} \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_j) \subseteq \mathcal{B}_{c_\Pi \varepsilon_0}(x_j) \tag{4.140}$$

where $\mathcal{B}$ denotes a ball in the metric on $\mathcal{X}$ (with a given radius and center). Since by definition $\check{f}^{(\zeta)}$ is $(1/\zeta)$-Lipschitz continuous, as long as $c_\Pi \varepsilon_0/\zeta \leqslant \rho^*/2$ (i.e. $\varepsilon_0 \leqslant \rho^*\zeta/(2c_\Pi)$) we have

$$|\check{f}^{(\zeta)}(X) - \check{f}^{(\zeta)}(x_j)| \leqslant \rho^*/2 \tag{4.141}$$

$$\implies \check{f}^{(\zeta)}(x_j) \geqslant \rho^* \tag{4.142}$$

for any X such that $\check{f}^{(\zeta)}(X) > (3/2)\rho^*$, which by definition means that $S_{j(X)}$ is ρ^*-sufficient.

Claim (iii) holds because $\check{f}^{(\zeta)}$ is Lipschitz continuous, $S_j \subseteq \mathcal{B}_{c_\Pi\varepsilon_0}(x_j)$, and $\check{f}^{(\zeta)}(x_j) \geqslant \rho^*$; thus, by making ε_0 sufficiently small, $\check{f}^{(\zeta)}(X) \geqslant (1 - \rho_1/2)\check{f}^{(\zeta)}(x_j)$ for every

$X \in \mathcal{B}_{c_\Pi \varepsilon_0}(x_j)$, and $\mathrm{Vol}_{\mathcal{X}}(S_j) \geqslant (1-\rho_1/2)\alpha \check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma$. Thus,

$$\begin{aligned}
\mathbb{P}_{f\sim X}[X \in S_j] &= \int_{S_j} f(x)\,dx && (4.143)\\
&\geqslant \int_{S_j} \check{f}^{(\zeta)}(x)\,dx && (4.144)\\
&\geqslant \int_{S_j} (1-\rho_1/2)\check{f}^{(\zeta)}(x_j)\,dx && (4.145)\\
&= (1-\rho_1/2)\check{f}^{(\zeta)}(x_j)\mathrm{Vol}_{\mathcal{X}}(S_j) && (4.146)\\
&\geqslant (1-\rho_1/2)^2\alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma && (4.147)\\
&\geqslant (1-\rho_1)\alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma\,. && (4.148)
\end{aligned}$$

Finally, claims (i)-(iii) show our result because

$$\begin{aligned}
\mathbb{P}_{X\sim f}[S_{j(X)} \text{ is } \rho_1\text{-good}] &\geqslant \mathbb{P}_{X\sim f}[S_{j(X)} \text{ is } \rho^*\text{-sufficient}] && (4.149)\\
&\geqslant \mathbb{P}_{X\sim f}[\check{f}^{(\zeta)}(X) > (3/2)\rho^*] && (4.150)\\
&\geqslant \rho_2 && (4.151)
\end{aligned}$$

and we are done. □

For the remainder of this proof, we will assume that all cells are ρ_1-good. We will show in Appendix D that this assumption can be made without loss of generality, i.e. that the same very high probability bound can be achieved.

Given covering SHCS's with roots $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$, we define

$$C_{\varepsilon_0} = \mathrm{TSP}_\Pi(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m) \tag{4.152}$$

i.e. the amount of TSP tour time needed to visit all the roots of the cells, which depends on ε_0 but is constant with regards to n. Note that this requires the vehicle to visit particular *configurations* rather than just targets, and note that computing it exactly can be intractable; however, we only need any constant length and, if we have good point-to-point (or configuration-to-configuration) motion planning for our

vehicle, we can even get a good constant approximation by computing the distances between all pairs of $\boldsymbol{q}_{j_1}, \boldsymbol{q}_{j_2}$ and applying a standard TSP approximation algorithm.

We now want to apply Proposition 14 or Proposition 15 (depending on whether $\gamma \geqslant 3$ since $\nu = 1 - \frac{1}{\gamma}$). To do this, we need to compute $\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}}$. We note that if we define a discrete random variable $Z \sim \boldsymbol{p}$ in $[m]$, this is equivalent to

$$\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} = \mathbb{E}_{Z\sim\boldsymbol{p}}\Big[p_Z^{-\frac{1}{\gamma}}\Big] \tag{4.153}$$

Let $X \sim f$; then we can interpret $Z = j(X)$. For any $\delta > 0$, we can choose $\rho_1 > 0$ and $\zeta > 0$ and ε_0 small enough that Lemma 28 applies and

$$\int_{\mathcal{X}_f} f(x)\check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}}\check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}}\,dx \approx \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \tag{4.154}$$

(this approximation can be made arbitrarily close by setting ζ small) to get

$$\begin{aligned}
\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} &= \mathbb{E}_{Z\sim\boldsymbol{p}}\Big[p_Z^{-\frac{1}{\gamma}}\Big] && (4.155)\\
&= \mathbb{E}_{X\sim f}\Big[p_{j(X)}^{-\frac{1}{\gamma}}\Big] && (4.156)\\
&\leqslant \mathbb{E}_{X\sim f}\Big[(1-\rho_1)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}\check{f}^{(\zeta)}(X)^{-\frac{1}{\gamma}}\check{g}^{(\zeta)}(X)^{-\frac{1}{\gamma}}\Big] && (4.157)\\
&= (1-\rho_1)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}\int_{\mathcal{X}_f} f(x)\check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}}\check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}}\,dx && (4.158)\\
&\leqslant (1+\delta)\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx && (4.159)
\end{aligned}$$

By Proposition 13, if n_j is the number of targets that fall into S_j, the time needed to visit all of them (starting and ending at the root $\boldsymbol{q}_j$) is at most $6s\varepsilon_0 n_j^{1-\frac{1}{\gamma}}$ and hence a TSP trajectory can be constructed by combining a TSP trajectory through the roots $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$ (taking a constant C_{ε_0} time) and these tours within cells; each time you

arrive at a $\boldsymbol{q}_j$, you execute a tour of the n_j targets within S_j. Thus,

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant C_{\varepsilon_0} + 6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} \tag{4.160}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + 12s\varepsilon_0 n^{1-\frac{1}{\gamma}} \sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} \tag{4.161}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + 12s\varepsilon_0 n^{1-\frac{1}{\gamma}} (1-\rho_1)^{-\frac{1}{\gamma}} \alpha^{-\frac{1}{\gamma}} \varepsilon_0^{-1} \int_{\mathcal{X}_f} f(x) \check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \, dx \tag{4.162}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + (1+\delta)12s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{4.163}$$

Exactly what 'wvhp' means here is dependent on what regime from Propositions 14 and 15 the exponent $\nu = 1 - \frac{1}{\gamma}$ falls in, but in all cases probability of failure is upper bounded (in the limit) by some $c_1 e^{-c_2 x^{c_3}}$ where $c_1, c_2, c_3 > 0$. Note that while C_{ε_0} increases as ε_0 gets smaller, and that ε_0 might have to be set small to achieve approximation error $\delta > 0$, it is constant with regard to n and hence negligible as $n \to \infty$.

Remark 20. *A fun consequence of this argument is that, in the Euclidean TSP on $\mathbb{R}^d$, if the target points $X_1, \ldots, X_n$ are drawn from a bounded set $\mathcal{X}^* \subseteq \mathbb{R}^d$ with Hausdorff dimension $\gamma \geqslant 1$, then the worst-case tour length scales according to $\Theta(n^{1-\frac{1}{\gamma}})$.*

The requirement that $\gamma > 1$ comes from the fact that if $\gamma < 1$, the time required to move between hierarchical cell structures, which is $O(1)$, dominates over $n^{1-\frac{1}{\gamma}}$. Thus, if $\gamma < 1$ the Euclidean TSP trajectory length defaults back to $O(1)$, i.e. bounded above by a constant even as $n \to \infty$.

Chapter 5

Nonsymmetric Dynamics Algorithm

In this chapter we present our algorithm for nonsymmetric dynamics, which, like the symmetric dynamics algorithm, is based on tiling the space with cells inscribed in reachable sets. The difference, however, is that a vehicle with symmetric dynamics can visit all the target points in a cell in a short period of time, as after each visit it can return to the anchor and then proceed to the next one; a vehicle with nonsymmetric dynamics can only be sure of visiting a single target point in each cell before the drift pushes it out and into the next cell. Thus, the algorithm for nonsymmetric dynamics must have a strategy of when and how often to return to each cell in order to to ensure that every target point is visited. As in the nonsymmetric dynamics algorithm, we will be working with cells of various different scales in order to avoid wasting too much time traveling through empty cells.

Since we can only achieve the correct order-of-growth for the shortest tour length for nonsymmetric dynamics, we make the simplifying assumption that the targets are uniformly distributed over a cube. Since we may have targets on a curved manifold, we define this in relation to the Nonsymmetric Hierarchical Cell Structure.

5.1 Tiling Algorithm

5.1.1 Nonsymmetric HCS cover

Under Assumption 11, we know that for every $x \in \mathcal{X}_f$, there is a Nonsymmetric Hierarchical Cell Structure containing x. We want to use this to produce a Nonsymmetric HCS cover:

Definition 24. A *Nonsymmetric Hierarchical Cell Structure* cover (Nonsymmetric HCS cover, or NHCS cover) is a finite set of NHCS's for which the union of their top-level cells covers $\mathcal{X}_f$.

Note that unlike the symmetric case, since we are only trying to show the order of growth, we are not concerned with overlaps. Thus, we can bound the performance of our algorithm above by considering how the algorithm performs when *every* NHCS contains n targets; if each NHCS has a $\Theta(n^{1-\frac{1}{\gamma}})$ growth rate for the length of the DSTSP tour through n random targets within it (with very high probability), then (also accounting for a fixed extra time to the tour for traveling between cells) the whole tour will have a $\Theta(n^{1-\frac{1}{\gamma}})$ length with very high probability.

Lemma 29. *If Assumption 11 holds, then a NHCS cover exists.*

Proof. For any $x \in \mathcal{X}_f$, let S_x denote the interior of the top-level cell of the NHCS containing x, which is open. Then clearly

$$\mathcal{X}_f \subseteq \bigcup_{x \in \mathcal{X}_f} S_x \,; \tag{5.1}$$

but since $\mathcal{X}_f$ is compact by Assumption 5, there is a finite subset of the S_x which also cover $\mathcal{X}_f$, which is our finite set of NHCS's. □

5.1.2 Nonsymmetric HCS and Passes

We now consider the Nonsymmetric HCS. For convenience we repeat Definition 14 here. An ε-scale NHCS with structural constants $\alpha_1, \alpha_2 > 0$ is a structure consisting

of an infinite sequence of *levels* $k = 0, 1, \ldots$. Each level k then consists of the following elements:

- $2^{k\gamma}$ *cells* which are subsets of the workspace

$$S_1^{(k)}, \ldots, S_{2^{k\gamma}}^{(k)} \subseteq \mathcal{X}\,. \tag{5.2}$$

  We denote the set of level-k cells as $\mathcal{S}^{(k)}$.

- $2^{k(\gamma-1)}$ *tracks*, which are trajectories

$$\pi_1^{(k)}, \ldots, \pi_{2^{k(\gamma-1)}}^{(k)} : [0, \alpha_1] \to \mathcal{Q} \tag{5.3}$$

  where each $\pi_i^{(k)}$ is associated with some control $\boldsymbol{u}_i^{(k)} \in \mathcal{U}$ and some starting configuration $\boldsymbol{q}_i^{(k)}$ and satisfies

$$\pi_i^{(k)}(t) := \exp(t\, h_{\boldsymbol{u}_i^{(k)}})(\boldsymbol{q}_i^{(k)}) \text{ for all } k \text{ and } i \in 1, 2, \ldots, 2^{k(\gamma-1)}\,. \tag{5.4}$$

  This means that $\pi_i^{(k)}$ is the trajectory produced when applying control $\boldsymbol{u}_i^{(k)}$ starting from $\boldsymbol{q}_i^{(k)}$ for α_1 time.

These satisfy the following relationships with each other:

- For each k, the level-k cells are all disjoint.

- For each k, every level-k cell $S_i^{(k)}$ is partitioned into 2^γ level-$(k+1)$ cells which we call its *sub-cells*.

- For each k, every level-k track $\pi_i^{(k)}$ is associated with 2^k different level k cells; WLOG we can assign the indices of the cells associated with $\pi_i^{(k)}$ as $S_{j;i}^{(k)}$ where $j = 1, 2, \ldots, 2^k$. We denote the set of all cells associated with track $\pi_i^{(k)}$ as

$$\mathcal{S}_i^{(k)} := \{S_{j;i}^{(k)} : j \in [2^k]\}\,. \tag{5.5}$$

- For each $S \in \mathcal{S}_i^{(k)}$, there is some $t \in [0, \alpha_1]$ such that

$$S \subseteq \bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t); \boldsymbol{u}_i^{(k)}) \,. \tag{5.6}$$

- Each $S \in \mathcal{S}^{(k)}$ at level k has volume

$$\mathrm{Vol}_{\mathcal{X}}(S) \geqslant \alpha_2 \varepsilon^\gamma / 2^{k\gamma} \,. \tag{5.7}$$

  and the single level-0 cell has nonzero volume.

We now define the following operation on an NHCS:

Definition 25. Given an NHCS with cell sets $\mathcal{S}^{(k)}$ for each level k, a *pass at level k* is a trajectory that visits one target in each (nonempty) level-k cell in $\mathcal{S}^{(k)}$. Any visited target is considered as removed from the problem.

This is our fundamental operation for visiting all targets which fall within the NHCS. We show that it can be done (at a given level k) in a limited time:

Proposition 16. Given an NHCS, there is some constant C such that a pass at any level k can be executed in time $\leqslant C 2^{k(\gamma-1)}$.

Proof. As above, we let $\boldsymbol{q}_i^{(k)}$ be the starting point of track $\pi_i^{(k)}$, which operates under fixed control $\boldsymbol{u}_i^{(k)}$.

We now consider the cells $S_{1;i}^{(k)}, \ldots, S_{2^k;i}^{(k)} \in \mathcal{S}_i^{(k)}$. Each is contained in some deviation set $\bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t))$; we denote the time in question as $t_{j;i}^{(k)}$, i.e.

$$S_{1;i}^{(k)} \subseteq \bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t_{j;i}^{(k)}), \boldsymbol{u}_i^{(k)}) \,. \tag{5.8}$$

Without loss of generality we index these in order, i.e.

$$0 \leqslant t_{1;i}^{(k)} \leqslant t_{2;i}^{(k)} \leqslant \cdots \leqslant t_{2^k;i}^{(k)} \leqslant \alpha_1 \,. \tag{5.9}$$

The general idea is that we can go on track $\pi_i^{(k)}$ and deviate at each $\pi_i^{(k)}(t_{j;i}^{(k)})$ to visit a target in $S_{j;i}^{(k)}$ before returning to $\pi_i^{(k)}(t_{j;i}^{(k)} + \varepsilon/2^k)$. By the definition of a

deviation set, the time needed to start at $\pi_i^{(k)}(t_{j;i}^{(k)})$ and visit any target

$$x \in S_{j;i}^{(k)} \subseteq \bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t_{j;i}^{(k)}), \boldsymbol{u}_i^{(k)}) \tag{5.10}$$

and then return to $\pi_i^{(k)}(t_{j;i}^{(k)} + \varepsilon/2^k)$ is at most $2\varepsilon/2^k$. Being back on $\pi_i^{(k)}$ will then allow us to proceed to $\pi_i^{(k)}(t_{j+1;i}^{(k)})$ by applying control $\boldsymbol{u}_i^{(k)}$.

However, one problem with the above is that it is very possible that

$$t_{j;i}^{(k)} + \varepsilon/2^k > t_{j+1;i}^{(k)} \tag{5.11}$$

i.e. the return point for the current deviation set is after the starting point for the next one. However, we claim that there is some constant (integer) Δ independent of k such that

$$t_{j;i}^{(k)} + \varepsilon/2^k \leqslant t_{j+\Delta;i}^{(k)}\,. \tag{5.12}$$

This means that we can always go to the Δ-th next cell. We show this claim as follows: first, we note that for any $t \in [t_{j;i}^{(k)}, t_{j;i}^{(k)} + \varepsilon/2^k]$,

$$\bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t), \boldsymbol{u}_i^{(k)}) \subseteq \bar{R}_{2\varepsilon/2^k}(\pi_i^{(k)}(t)) \subseteq \bar{R}_{3\varepsilon/2^k}(\pi_i^{(k)}(t_{j;i}^{(k)})) \tag{5.13}$$

Then, by Assumption 6 we have a constant C' such that

$$\mathrm{Vol}_{\mathcal{X}}(\bar{R}_{3\varepsilon/2^k}(\pi_i^{(k)}(t_{j;i}^{(k)}))) \leqslant C'(\varepsilon/2^k)^\gamma = C'\varepsilon^\gamma/2^{k\gamma}\,. \tag{5.14}$$

Now we look at all j' such that $t_{j';i} \in [t_{j;i}^{(k)}, t_{j;i}^{(k)} + \varepsilon/2^k]$. We know that for such j',

$$S_{j';i}^{(k)} \subseteq \bar{D}_{\varepsilon/2^k}(\pi_i^{(k)}(t_{j';i}^{(k)}), \boldsymbol{u}_i^{(k)}) \subseteq \bar{R}_{3\varepsilon/2^k}(\pi_i^{(k)}(t_{j;i}^{(k)})) \tag{5.15}$$

and all of them are disjoint. But then note that each $S_{j';i}^{(k)}$ satisfies

$$\mathrm{Vol}_{\mathcal{X}}(S_{j';i}^{(k)}) \geqslant \alpha_2\varepsilon^\gamma/2^{k\gamma}\,. \tag{5.16}$$

But this means by volume that at most $\Delta = \lceil C'/\alpha_2 \rceil$ of these $S^{(k)}_{j';i}$ can fit in $\bar{R}_{3\varepsilon/2^k}(\pi^{(k)}_i(t^{(k)}_{j;i}))$.

Thus, we know that $t^{(k)}_{j;i} + \varepsilon/2^k \leqslant t^{(k)}_{j+\Delta;i}$; thus we can get all of the cells in $\mathcal{S}^{(k)}_i$ with Δ runs down the track (the first time we get a target in each of $S^{(k)}_{1;i}, S^{(k)}_{1+\Delta;i}, S^{(k)}_{1+2\Delta;i} \dots$, the second time we get a target in each of $S^{(k)}_{2;i}, S^{(k)}_{2+\Delta;i}, S^{(k)}_{2+2\Delta;i} \dots$, and so on up to $S^{(k)}_{\Delta;i}, S^{(k)}_{\Delta+\Delta;i}, S^{(k)}_{\Delta+2\Delta;i} \dots$). Each run down the track takes at most $2\alpha_1$ time (the track originally takes $\leqslant \alpha_1$ time by (5.3), and we need to potentially double it to account for extra time taken during deviations); by Assumption 4, for each 'reset' of the track (return to $\boldsymbol{q}^{(k)}_i$) requires at most a constant C'' time.

[Note: if the end of the track $\pi^{(k)}_i(\alpha_1 + 2\varepsilon/2^{k\gamma})$ (the end because we may still need to visit a target starting at $\pi^{(k)}_i(\alpha_1)$) and/or the start $\boldsymbol{q}^{(k)}_i = \pi^{(k)}_i(0)$ fall outside $\mathcal{X}_f$, we can use the last time it falls outside and the first time it falls inside so the assumption still applies.]

Thus, we can repeat track $\pi^{(k)}_i$ at most Δ times and visit one target from each (nonempty) cell in $\mathcal{S}^{(k)}_i$, which takes a total time of at most

$$\Delta(\alpha_1 + 2\varepsilon/2^{k\gamma} + C'') \leqslant \Delta(\alpha_1 + \varepsilon + C'') = C \tag{5.17}$$

including an extra segment needed to move to the next track. Repeating this for all $2^{k(\gamma-1)}$ tracks then completes a pass in at most $C2^{k(\gamma-1)}$ time, visiting at least one target from each (nonempty) cell in $\mathcal{S}^{(k)} = \bigcup_i \mathcal{S}^{(k)}_i$. □

5.2 Uniform targets on a Nonsymmetric HCS

We now consider the simplified problem of uniformly random iid targets on a Nonsymmetric HCS. By Proposition 16, the vehicle can pass through all cells at level $k \geqslant 0$ in time $\leqslant C\, 2^{k(\gamma-1)}$ and visit at least one target from each of $2^{k\gamma}$ (nonempty) cell at level k. We call this a *pass*. We will show in Corollary 2 that this simplified case is sufficient for the general upper bound we want to show.

Since in this section we are only concerned about the order-of-growth, we can ignore constant multiplicative factors (so long as we know that their supremum is

finite). Thus, we can consider the cost of a pass at level k to be $2^{k(\gamma-1)}$. Furthermore, since our primary goal is to find a probabilistic upper bound to the length of the shortest tour (via designing an algorithm which achieves a length of at most the upper bound), we make the pessimistic assumption that we can visit exactly one target in each nonempty cell per pass (targets that we have already visited are considered removed from the problem).

To emulate this problem, we define the *Nonsymmetric Hierarchical Collection Problem* (Nonsymmetric HCP, or NHCP). Like its symmetric counterpart, is an abstract combinatorial problem designed to represent the general characteristics of our algorithmic framework for solving a DSTSP instance. For the nonsymmetric DSTSP, this means representing the problem of deciding how many passes to make before each merge step in the algorithmic framework given above.

5.2.1 The Nonsymmetric Hierarchical Collection Problem

We define the Nonsymmetric Hierarchical Collection Problem as follows; for simplicity we assume that $n = (2^\gamma)^\ell = 2^{\ell\gamma}$ for some integer ℓ. For n which does not satisfy this, we note that we can add (random) 'phantom' targets to round up to the nearest power of 2^γ, which increases the number of targets by at most a factor of 2^γ.

As in the Symmetric HCP, the nesting cell structure is represented by a rooted tree T with branching factor $\bar{b} = 2^\gamma$, where each vertex represents a cell and its children represent its sub-cells; the target points $X_1, \ldots, X_n$ are represented by targets $\tau_1, \ldots, \tau_n$ which are each an infinite path down the tree, starting from the root, representing the sequence of (diminishing) cells containing X_i. However, a major difference in the Symmetric HCP and the Nonsymmetric HCP is that the Symmetric HCP allows the player to move around the cells according to the cell structure as they want, visiting as many target points in each cell as they want; by contrast, the Nonsymmetric HCP only allows the player to visit the cells in a very fixed way with little flexibility, and with the ability to visit only one target point in each cell (however, they can repeat ths cycle as many times as they want). This means the main action of the player is not to move around the tree T, but to execute a *pass*:

at each step, the player picks a particular level (i.e. all the vertices at some depth j) and visits one target (of their choice) at each vertex; this

Definition 26. The Nonsymmetric Hierarchical Collection Problem (NHCP), parameterized by γ, is a combinatorial problem with the following elements:

i. A rooted tree $T = (V, E)$ with branching factor 2^γ; we denote by V_j the set of vertices at depth j (with the root at depth 0).

ii. n targets $\tau_1, \ldots, \tau_n$; each τ_i represents an infinite path through T, starting from the root. We denote by $v_j(\tau_i)$ the vertex at level j that τ_i passes through.

A (single-player) game is played on this structure. At each step, the player may make a *pass* at any level j that they select: they collect one target of their choice which passes through each $v \in V_j$ (if no target τ_i passes through v, no target is collected at that vertex), for a cost of $2^{(\gamma-1)j}$. The objective is to collect all the targets for the minimum cost. We denote by

$$\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \tag{5.18}$$

the minimum cost needed to collect all n targets τ_i.

The *Stochastic NHCP* is the above but the targets $\tau_1, \ldots, \tau_n$ are iid distributed uniformly, i.e. $v_{j+1}(\tau_i)$ is always equally likely to be any of the 2^γ children of $v_j(\tau_i)$.

Note that 'passes' get more efficient in a sense at deeper levels: since T has branching factor 2^γ, at level j there are $2^{\gamma j}$ vertices and hence the potential to collect $2^{\gamma j}$ targets in a single pass, for a cost of only $2^{(\gamma-1)j}$. However, the trade-off is that at deeper levels (with a limited number of targets n) many of the cells will be empty, which becomes inefficient. Thus the player needs to strike a balance in order to collect targets for minimum cost.

One feature of the Nonsymmetric HCP which makes it very unlike the Symmetric HCP is that the cost is maximized when the targets $\tau_1, \ldots, \tau_n$ are bunched up in the same vertices, rather than spread out (in fact, the minimum cost of the NHCP

corresponds roughly to the maximum cost of the SHCP, since both are realized when the targets are spread out, in which case the SHCP algorithm performs actions which closely resemble the 'pass' of the NHCP). This reflects the fact that a symmetric vehicle can remain inside the cell and collect all the targets there quickly, while a nonsymmetric vehicle must make repeated passes, which essentially iterates through all the vertices at that level, to get them all (in which case it may as well just do its passes at level 0).

One effect of this is that the guaranteed $\Theta(n^{1-\frac{1}{\gamma}})$ upper bound of the SHCP does not apply to the NHCP; for example, if at each level the targets within each vertex are divided into at most $2^{\gamma-1}$ sub-vertices, then a pass at any level can collect at most 1 target per unit of cost it incurs, and therefore the targets can be collected only by incurring a total of n cost (by e.g. taking n passes at level 0). This target point distribution can be achieved (in the worst case) if the targets are distributed on a lower-dimensional subset of $\mathcal{X}$; see the proof of Proposition 20 for an example where this sort of target point placement causes the shortest tour to have length $\Theta(n)$.

5.2.2 Relationship between the DTSP and NHCP

We now show the relationship between the Dynamic TSP and the Nonsymmetric HCP. In particular, we show how a DTSP instance maps to an NHCP instance:

Definition 27. For a Nonsymmetric Hierarchical Cell Structure S, we build an infinite rooted tree T as follows: each cell of the NHCS corresponds to a vertex of T, with the root vertex v_0 corresponds to the top-level cell in the NHCP; then the 2^γ children of each vertex v correspond to the 2^γ sub-cells of the cell corresponding to v. This trivially means that each level-k cell corresponds to a vertex at depth k of T.

Then, given some $x \in S$, its *cell path* is the path through T corresponding to all the cells containing x (starting at the root v_0 and going to the child of v_0 corresponding to the sub-cell containing x, and so forth). We call this $\tau(x)$; note that if $x \neq x'$ then $\tau(x) \neq \tau(x')$ since at some level they will fall in different cells.

Then for targets $x_1, \ldots, x_n$ on S, the *equivalent NHCP instance* of the DTSP is $\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma)$ where $\tau_i = x_i$.

We now show how a DTSP instance relates to its equivalent NHCP instance:

Proposition 17. Given targets $x_1, \ldots, x_n$ in a Nonsymmetric Hierarchical Cell Structure, let $\tau_1, \ldots, \tau_n$ denote the cell paths corresponding to $x_1, \ldots, x_n$. Then there is some constant $C > 0$ such that

$$\mathrm{TSP}_\Pi(x_1, \ldots, x_n) \leqslant C \cdot \mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma)\,. \tag{5.19}$$

Proof. This follows from the rules of making a pass as outlined above: there must be a sequence of passes in the NHCP which allow all the targets to be collected with $\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma)$ cost; we can then run that sequence of passes (the same levels and visiting targets x_i corresponding to the same □

Thus, if we have some function $\alpha(n)$ (not to be confused with the efficiency parameter from the Symmetric HCS) such that $\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \leqslant \alpha(n)$, then $\mathrm{TSP}_\Pi(x_1, \ldots, x_n) = C \cdot \alpha(n)$. This yields the following corollary:

Corollary 2. If $\tau_1, \ldots, \tau_n$ are iid uniformly distributed random paths on T and there is some constant C' such that $\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \leqslant C' n^{1-\frac{1}{\gamma}}$ with very high probability, then for any Π, f satisfying the assumptions in Section 2.1, there is some constant C such that

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant C n^{1-\frac{1}{\gamma}} \tag{5.20}$$

with very high probability when $X_1, \ldots, X_n \overset{iid}{\sim} f$.

Proof. We first show this to be true when f is the uniform distribution on some Nonsymmetric HCS S. Then having $\tau_1, \ldots, \tau_n$ iid uniformly distributed random paths on T corresponds exactly to the cell paths of $X_1, \ldots, X_n$; then Proposition 17 means that if $\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \leqslant C' n^{1-\frac{1}{\gamma}}$ with very high probability, we automatically get $\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant C n^{1-\frac{1}{\gamma}}$ with very high probability.

Then to generalize this to all distributions f, we find a NHCS cover (by Lemma 29 one always exists), which covers $\mathcal{X}_f$ with a finite collection of NHCS's $S_1, \ldots, S_m$; then we can use this cover to partition $\mathcal{X}_f$ and solve the problem on each piece separately and combine them in the end. If the distribution f is not uniform on each piece, we can add random 'phantom' targets to make it uniform, increasing the number of target points by a multiplicative factor (which does not affect the order-of-growth since it applies the multiplicative factor to $n^{1-\frac{1}{\gamma}}$). For a full proof, see Appendix A.1.2 and in particular Proposition 23. □

5.2.3 Analysis of the Nonsymmetric HCP

Now that we have established the relationship between the NHCP (with uniformly iid random targets) and the DSTSP, we need to analyze the NHCP. We want to show the following:

Proposition 18. If $\tau_1, \ldots, \tau_n$ are iid uniformly randomly over the tree T (i.e. $v_{j+1}(\tau_i)$ is always equally likely to be any of the 2^γ children of $v_j(\tau_i)$), then

$$\mathbb{P}[\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \leqslant 6 \cdot 2^{\gamma-1} \cdot n^{1-\frac{1}{\gamma}}] \geqslant 1 - 4 \cdot e^{-n^{1-\frac{1}{\gamma}}} \tag{5.21}$$

with very high probability.

Note that in this section ℓ refers to a level in the tree T, not the length of a trajectory. Then Proposition 18 and Corollary 2 together show the nonsymmetric upper bound in Theorem 1.

Lemma 30. *If $n \geqslant 2^{\gamma\ell}$ for integer ℓ, then for any $\tau_1, \ldots, \tau_n$,*

$$\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \geqslant 2^{(\gamma-1)\ell}. \tag{5.22}$$

Consequently, we can establish the bound:

$$\mathrm{NHCP}(\tau_1, \ldots, \tau_n; \gamma) \geqslant 2^{-(\gamma-1)} n^{1-\frac{1}{\gamma}}. \tag{5.23}$$

Proof. Trivially, adding more targets increases the cost to collect all targets, so we may WLOG assume that $n = 2^{\gamma\ell}$.

Suppose we make a pass at level j; what is the minimum possible cost per target collected?

- When $j \leqslant \ell$, we collect at most $2^{\gamma j}$ targets for a cost of $2^{(\gamma-1)j}$, yielding a cost of at least

$$2^{(\gamma-1)j}/2^{\gamma j} = 2^{(\gamma-1)j-\gamma j} = 2^{-j} \tag{5.24}$$

  per target collected.

- When $j > \ell$, we collect at most $n = 2^{\gamma\ell}$ targets (since those are all the targets) for a cost of at least

$$2^{(\gamma-1)j}/2^{\gamma\ell} = 2^{(\gamma-1)j-\gamma\ell} = 2^{-\ell+(\gamma-1)(j-\ell)} \tag{5.25}$$

  per target collected.

This is then clearly minimized at $j = \ell$, yielding a minimum possible $2^{-\ell}$ cost per target collected by any pass. Furthermore, this cost-per-target can only increase as the number of targets diminishes. Thus, the minimum possible cost of collecting all $n = 2^{\gamma\ell}$ targets is

$$2^{-\ell}n = 2^{(\gamma-1)\ell} \tag{5.26}$$

and we are done with the first part. The second part follows from the first because

we can set $\ell = \lfloor \log_2(n)/\gamma \rfloor \geqslant \log_2(n)/\gamma - 1$ to get

$$\ell \geqslant \frac{\log_2 n}{\gamma} - 1 \tag{5.27}$$

$$\implies \gamma\ell \geqslant \log_2 n - \gamma \tag{5.28}$$

$$\implies (\gamma - 1)\ell \geqslant \left(1 - \frac{1}{\gamma}\right) \log_2 n - (\gamma - 1) \tag{5.29}$$

$$\implies 2^{(\gamma-1)\ell} \geqslant 2^{-(\gamma-1)} n^{1-\frac{1}{\gamma}} \tag{5.30}$$

thus completing the second part. □

Note that this minimum possible might be achieved if each target is present at a unique level-ℓ vertex, but it may be unachievable if e.g. the targets are concentrated in some much smaller subset of the vertices.

Remark 21. *While this shows the minimum possible cost to collect all targets in the Nonsymmetric HCP, it does* not *reflect the minimum possible cost to collect all targets in the DSTSP; this is because the NHCP make the safe assumption that only one target can be collected in a vertex (corresponding to a cell) per pass, and that empty vertices cannot be skipped. Hence, like the Symmetric HCP, it is only an upper bound for the corresponding DSTSP instance.*

This demonstrates that if we find a strategy which (with very high probability with random targets) collects all the target points with cost $\leqslant Cn^{1-\frac{1}{\gamma}}$ for some constant $C > 0$, we will have achieved a solution whose cost is, with very high probability, within a constant factor of the optimal solution of the Nonsymmetric HCP.

We now consider strategies for the Nonsymmetric HCP. Given that the targets are randomly distributed, there are two basic approaches to showing that it can be solved for relatively low cost with very high probability: describe a strategy that always successfully collects all the targets, and show that it has very high probability to achieve low cost; or describe a strategy that always has low cost, and show that it has a very high probability of successfully collecting the target points. We will opt for the latter:

Definition 28. Given γ and the number of targets n, the *exponential moment suppression strategy* (EMS strategy) makes the following passes:

- *Initialization:* 1 pass at level $\ell := \lceil \frac{\log_2(n)}{\gamma} \rceil$;
- *Main sequence:* 2^γ passes at levels $j = \ell - 1, \ell - 2, \ldots, 1, 0$;
- *Cleanup:* $2^{(\gamma-1)\ell}$ passes at level 0.

Since the passes are taken in reverse order of levels (deeper passes first), as long as a target is taken from each nonempty cell on every pass, it makes no difference which targets are taken (different targets passing through the same vertex $v \in V_j$ may pass through different vertices at level $j' > j$, but at level $j'' < j$ they must all pass through the ancestor of v, so which one is taken and which is left for future passes is irrelevant).

Note that this is a 'blind' strategy: it doesn't take into account where the targets $\tau_1, \ldots, \tau_n$ actually are, but instead goes through a fixed sequence of passes. This makes the cost easy to bound, leaving us to show that it successfully collects all the targets with very high probability.

Lemma 31. *The EMS strategy collects all the targets with cost* $\leqslant 6 \cdot 2^{\gamma-1} \cdot n^{1-\frac{1}{\gamma}}$

Proof. We start by noting that $2^{(\gamma-1)\ell} \leqslant 2^{\gamma-1} n^{1-\frac{1}{\gamma}}$ since by definition $n > 2^{(\gamma-1)(\ell-1)}$. We then consider all costs as multiples of $2^{(\gamma-1)\ell}$:

- *Initialization:* 1 pass at level ℓ, which costs $2^{(\gamma-1)\ell}$;
- *Main Sequence:* 2^γ passes at each level $\ell - 1, \ell - 2, \ldots, 0$. Since $\gamma \geqslant 2$, we know that the cost of a pass at level $\ell - j$ is at most

$$2^{(\gamma-1)(\ell-j)} = 2^{-(\gamma-1)j} \cdot 2^{(\gamma-1)\ell} \leqslant 2 \cdot 2^{-(\gamma-1)} \cdot 2^{-j} \cdot 2^{(\gamma-1)\ell} \tag{5.31}$$

(since $(\gamma - 1)j \geqslant (\gamma - 1) + j - 1$ as $\gamma - 1, j$ are both positive integers), hence the cost of 2^γ passes is at most $4 \cdot 2^{-j} \cdot 2^{(\gamma-1)\ell}$. Then the total cost of the main

sequence passes is bounded by:

$$\sum_{j=1}^{\ell} 2^{\gamma} \cdot 2^{(\gamma-1)(\ell-j)} \leqslant \sum_{j=1}^{\ell} 4 \cdot 2^{-j} \cdot 2^{(\gamma-1)\ell} \leqslant 4 \cdot 2^{(\gamma-1)\ell} \tag{5.32}$$

- *Cleanup:* $2^{(\gamma-1)\ell}$ passes at level 0, which cost 1 per pass and hence a total cost of $2^{(\gamma-1)\ell}$.

Thus, summing these yields a total cost of at most $6 \cdot 2^{(\gamma-1)\ell}$ which, as noted, is at most $6 \cdot 2^{\gamma-1} \cdot n^{1-\frac{1}{\gamma}}$. □

We now need to show that, with high probability, the EMS strategy successfully collects all the targets. Note that since the EMS strategy never makes a pass at a deeper level than ℓ, we can ignore layers of T below that depth and treat it as a depth-ℓ (finite) rooted tree with branching factor 2^{γ}. Furthermore, since $2^{\gamma\ell} \geqslant n$ by definition, to get a lower bound on the probability that the EMS strategy collects all targets we can WLOG assume $n = 2^{\gamma\ell}$. This then means that T has $n = 2^{\gamma\ell}$ leaves, and we can consider the targets as being distributed iid uniformly among them (which then specifies their vertices at higher layers).

However, we really want the number of targets at each leaf to be independent of the number of targets at the other leaves. To do this, we *Poissonize* the number of targets: instead of distributing n targets at random, we first choose $N \sim \text{Pois}(n)$ (the Poisson distribution with parameter n) and then distribute N targets iid uniformly at random among the n leaves. We refer to this as the *Poissonized target distribution* or *Poissonized targets*. We now show that this doesn't substantially change the probability of collecting all the targets:

Lemma 32. *If the EMS strategy fails to collect all the targets with probability $\leqslant p$ when there are $N \sim \text{Pois}(n)$ iid targets, then its probability of failure is $\leqslant 4p$ with n iid targets.*

This implies that if the EMS strategy succeeds with very high probability on Poissonized targets, it will also succeed with very high probability on the original problem with n targets distributed uniformly at random.

Proof. We note that $\mathbb{P}[N \geqslant n] \geqslant 1/4$ if $N \sim \text{Pois}(n)$. Let p' be the probability of failure when n targets are distributed; this means that when N targets are distributed, with at least 1/4 probability (after $N \sim \text{Pois}(n)$ is chosen but the targets $\tau_1, \ldots, \tau_N$ are not yet distributed) the probability of failure of the EMS strategy is p'; thus the probability of failure of the EMS strategy under the Poissonized target distribution is $\geqslant (1/4)p'$ (inequality because (i) the chance that $N \geqslant n$ may be 1/4, (ii) distributing more than n targets can yield a greater than p' chance of failure, and (iii) distributing less than n targets but more than 0 can still yield a nonzero chance of failure). Thus,

$$\mathbb{P}_{N \sim \text{Pois}(n)}[\text{EMS strategy fails}] \leqslant p \implies p' \leqslant 4p \tag{5.33}$$

and we are done. □

In fact, there is good reason to believe that Poissonized targets are actually worse (from the standpoint of success probability of the EMS strategy) than a fixed number n of targets, since fixing the number of targets anticorrelates the number of targets in different buckets; but the above is sufficient for us.

We now note that since we are assuming (WLOG) that $n = 2^{\gamma\ell}$, the expected number of targets in the Poissonized case is $\mathbb{E}[N] = n$, and the unique properties of the Poisson distribution mean that if $N \sim \text{Pois}(n)$ targets are distributed uniformly and iid at random into n vertices, each vertex independently gets Pois(1) targets. For any vertex $v \in V_j$, we denote $N_v := |\{\tau_i : v_j(\tau_i) = v\}|$, i.e. the number of targets which pass through v. This means:

Lemma 33. *For $v \in V_\ell$ (leaves of T since we ignore layers below ℓ), we have $\{N_v\}_{v \in V_\ell} \overset{iid}{\sim} \text{Pois}(1)$.*

Proof. As stated, this follows from the properties of the Poisson distribution. □

We now define the *exponential moment* of a random variable:

Definition 29. For a random variable Y, we define $\mu^{(\exp)}(Y) := \mathbb{E}[e^Y]$.

This has the following properties: let $(z)_+ := \max(z, 0)$; then,

Lemma 34. *If $W_1, \ldots, W_k$ are independent random variables, then*

$$\mu^{(\exp)}\left(\sum_{j=1}^{k} W_j\right) = \prod_{j=1}^{k} \mu^{(\exp)}(Y_j)\,. \tag{5.34}$$

Furthermore, for any random W and any $k \geqslant 0$,

$$\mu^{(\exp)}((W-k)_+) \leqslant \frac{\mu^{(\exp)}(W) - 1}{e^k} + 1\,. \tag{5.35}$$

Proof. Equation (5.34) follows trivially from the fact that the expected value of the product of independent random variables is the product of their expected values.

To prove (5.35), let $w_j = \mathbb{P}[W = j]$ for all nonnegative integers j. Then, by definition:

$$\mu^{(\exp)}(W) - 1 = \sum_{j=0}^{\infty} w_j(e^j - 1). \tag{5.36}$$

Note that when $W \leqslant k$, $(W-k)_+ = 0$; otherwise, $(W-k)_+ = W - k$. This means that $(W-k)_+ = 0$ with probability $w_0 + w_1 + \cdots + w_k$, and for all $j > k$,

$$\mathbb{P}[(W-k)_+ = j - k] = w_j. \tag{5.37}$$

Thus, putting this together, we get that:

$$\mu^{(\exp)}((W-k)_+) - 1 = \sum_{j=0}^{k} w_j(e^0 - 1) + \sum_{j=k+1}^{\infty} w_j(e^{j-k} - 1) = \sum_{j=k+1}^{\infty} w_j(e^{j-k} - 1). \tag{5.38}$$

However, this means that

$$e^k\big(\mu^{(\exp)}((W-k)_+) - 1\big) = \sum_{j=k+1}^{\infty} w_j(e^j - e^k) \leqslant \sum_{j=0}^{\infty} w_j(e^j - 1) = \mu^{(\exp)}(W) - 1 \tag{5.39}$$

where the inequality is due to both the addition of (nonnegative) missing terms *and*

the fact that we are subtracting e^k from each term on the left and only subtracting 1 on the right. Rearranging this inequality gives us

$$\mu^{(\exp)}((W-k)_+) \leqslant \frac{\mu^{(\exp)}(W)-1}{e^k} + 1. \tag{5.40}$$

□

We note that the proof of property (5.35) works in much the same way (with an integral instead of a sum) even if W is not an integer random variable and k is not an integer; however, we only need the restricted form to obtain our desired result.

The above are important for the following reasons: the number of targets in a level-j vertex v is the sum of the number of targets in its children, which are independent when the target distribution is Poissonized; and making a pass at level j removes a target from each vertex $v \in V_j$ so long as it still has at least one, so if it had W targets and we make k passes we end up with $(W-k)_+$ targets.

We now can show that the EMS strategy works with very high probability:

Proposition 19. If $N \sim \text{Pois}(n)$ targets $\tau_1, \ldots, \tau_N$ are uniformly distributed in T with branching factor 2^γ, then

$$\mathbb{P}[\text{EMS strategy collects all targets}] \geqslant 1 - e^{-2^{(\gamma-1)\ell}} \geqslant 1 - e^{-n^{1-\frac{1}{\gamma}}}. \tag{5.41}$$

In other words, the EMS strategy collects all targets with very high probability.

Proof. Let $Y_j(v)$ be a random variable denoting the number of targets remaining at vertex $v \in V_j$ after the 2^γ main sequence passes at level j; for the special case of $j = \ell$, let $Y_\ell(v)$ denote the number of targets remaining at vertex v after the initialization step. We note that by symmetry, $\mu^{(\exp)}(Y_j(v))$ is the same for all $v \in V_j$, and hence we let $y_j := \mu^{(\exp)}(Y_j(v))$, which holds for all $v \in V_j$.

We then make the following claim: $y_j \leqslant e$ for all $j = \ell, \ell-1, \ldots, 1, 0$. We show this by induction (starting from base case $j = \ell$ and proceeding backwards to $j = 0$).

Base case $j = \ell$: Note that for all $v \in V_\ell$, we have $N_v \sim \text{Pois}(1)$ and hence, by definition of the exponential moment,

$$\mu^{(\exp)}(N_v) = \sum_{k=0}^{\infty} \frac{e^k}{e \cdot k!} = e^{-1} \cdot \sum_{k=0}^{\infty} \frac{1}{k!} e^k = e^{e-1} \tag{5.42}$$

as the value of the summation in the preceding line is the classic infinite-sum definition of the exponential. Thus (by (5.35) from Lemma Lemma 34):

$$y_\ell = \mu^{(\exp)}((N_v - 1)_+) \leqslant \frac{e^{e-1} - 1}{e} + 1 < e \tag{5.43}$$

(for those who are curious, $\frac{e^{e-1}-1}{e} + 1 \approx 2.683$ which is close to but still less than e). Thus, $y_\ell \leqslant e$, satisfying the base case.

Inductive step $j < \ell$: We need to show that if $y_{j+1} \leqslant e$, then $y_j < e$. For any $v \in V_j$, let Z_v denote the number of targets in v just before the 2^γ main sequence passes at level j; then since the Poissonized target distribution makes the number of targets in v's 2^γ children independent (the original number of targets in each level-$(j+1)$ vertex is independent, and since the strategy doesn't account for the positions of the targets when deciding what passes to make they remain independent) and each has (by definition) exponential moment $y_{j+1} \leqslant e$, we have

$$\mu^{(\exp)}(Z_v) = y_{j+1}^{2^\gamma} \leqslant e^{2^\gamma} \tag{5.44}$$

which then means that (since we make 2^γ passes)

$$y_j = \mu^{(\exp)}((Z_v - 2^\gamma)_+) \leqslant \frac{e^{2^\gamma} - 1}{e^{2^\gamma}} + 1 < 2 < e \tag{5.45}$$

as we wanted.

Therefore, we have shown our claim that $y_j \leqslant e$ for all $j = \ell, \ell - 1, \ldots, 1, 0$; in particular, we are interested in having $y_0 < e$. Now we let Y_0 denote the number of targets left after the main sequence is done (before the cleanup phase), and N_{final} be the number of targets left at the end of the game (after the cleanup phase). We thus

know that $\mu^{(\exp)}(Y_0) = y_0 \leqslant e$, and $N_{\text{final}} = (Y_0 - 2^{(\gamma-1)\ell})_+$. Therefore,

$$\mu^{(\exp)}(N_{\text{final}}) \leqslant \frac{e-1}{e^{2^{(\gamma-1)\ell}}} + 1\,. \tag{5.46}$$

The EMS strategy fails if $N_{\text{final}} > 0$; suppose it does so with probability p. Then a trivial lower bound for the exponential moment is

$$\mu^{(\exp)}(N_{\text{final}}) \geqslant (1-p) + ep = 1 + (e-1)p \tag{5.47}$$

Thus, putting it together yields

$$1 + (e-1)p \leqslant 1 + \frac{e-1}{e^{2^{(\gamma-1)\ell}}} \implies p \leqslant e^{-2^{(\gamma-1)\ell}} \tag{5.48}$$

and hence the probability that the EMS strategy gets all the targets is at least $1-p \geqslant 1-e^{-2^{(\gamma-1)\ell}}$. For the final step, we note that in this proof we were actually distributing $N \sim \text{Pois}(2^{\gamma\ell})$ targets, and we possibly had fewer than that; and furthermore by the definition of ℓ we have $2^{(\gamma-1)\ell} \geqslant n^{1-\frac{1}{\gamma}}$ and hence we have $1-e^{-2^{(\gamma-1)\ell}} \leqslant 1-e^{-n^{1-\frac{1}{\gamma}}}$. $\square$

Finally, to prove Proposition 18 we use the EMS strategy: by Lemma 31 the EMS strategy has cost at most $2^{\gamma-1} \cdot 6 \cdot n^{1-\frac{1}{\gamma}}$; by Proposition 19, when the targets are Poissonized the probability of failure is at most $e^{-n^{1-\frac{1}{\gamma}}}$; and by Lemma 32 the probability of failure in the original case (when there are always n targets) is at most 4 times the probability of failure in the Poissonized case. Thus, we are done.

Remark 22. *The EMS strategy can be quickly converted to a strategy that always collects all targets but has variable costs: follow the EMS strategy all the way to the cleanup step, and then simply make sufficient cleanup passes at level* 0 *to collect all targets. Since the EMS strategy works with very high probability, the expected cost of the modified EMS strategy will be* $\Theta(n^{1-\frac{1}{\gamma}})$ *(as the number of cleanup passes at level* 0 *is at most linear even in the worst case).*

Chapter 6

Adversarial Targets

We now show Theorem 2, which deals with the length of the tour when the targets are placed (within a bounded $\mathcal{X}^* \subseteq \mathcal{X}$) to maximize the shortest tour length. We refer to this as *adversarial* target placement. While our adversarial target lower bound will hold for non-symmetric dynamics, as in Chapter 4 our main focus here is on symmetric Π.

As discussed in Section 1.3, we consider two related questions which follow naturally from Theorem 1:

1. Given symmetric dynamics Π and bounded (but full dimensional) target region $\mathcal{X}^*$, what target point density f over $\mathcal{X}^*$ maximizes the expected length of the optimal TSP trajectory when $X_1, \ldots, X_n \overset{iid}{\sim} f$ as $n \to \infty$?

2. Given symmetric dynamics Π, a bounded set $\mathcal{X}^* \subseteq \mathcal{X}$ and a number of target points n, what is $\sup_{X_1,\ldots,X_n \in \mathcal{X}^*} \mathrm{TSP}_\Pi(X_1, \ldots, X_n)$?

Although we have lower and upper bounds on the tour length which are tight up to a fixed constant factor for all f, we cannot claim to know exactly what f maximizes the DSTSP tour length (either in expectation or with high probability). However, we know that both the lower and upper bounds to our tour length are of the form

$$cn^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{6.1}$$

where c is in both cases a constant which does not depend on f. Therefore, instead of question 1, we will answer the related question of: given symmetric dynamics Π and bounded (but full dimensional) target region $\mathcal{X}^*$, what target point density f over $\mathcal{X}^*$ maximizes

$$\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{6.2}$$

and what value does this maximum take?

Question 2 then asks for a deterministic upper bound on the tour length, with targets chosen by an adversary to maximize tour length. While adversarially-chosen targets are by definition going to produce a longer tour length than any random distribution, we are interested in knowing whether, as $n \to \infty$, it is possible to do substantially better than the worst-case f from question 1. Note that question 1 requires a single probability distribution to be used for all values of n as $n \to \infty$, while question 2 allows a different set of points to be selected depending on n. While this seems like it might allow more flexibility for the adversarial target points, we will show that roughly the same bound applies to the adversarial target points.

6.1 Adversarial Targets Lower Bound

In order to show the lower bound from Theorem 2, we will find density f_g over $\mathcal{X}^*$ which maximizes the lower bound from Theorem 1, which is equivalent to maximizing $\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx$; this clearly also serves as a lower bound to the longest tour from non-random target points. It also maximizes the upper bound from Theorem 1 thanks to the constant factor gap.

We thus have to solve the following problem:

$$\begin{aligned} &\text{maximize} \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \quad (\text{where } f : \mathcal{X}^* \to \mathbb{R} \text{ is integrable}) \\ &\text{subject to} \int_{\mathcal{X}^*} f(x)dx = 1, \text{ and } f(x) \geqslant 0 \text{ for all } x \in \mathcal{X}^* \end{aligned} \tag{6.3}$$

(i.e. the constraint is that f is a probability density function over $\mathcal{X}^*$).

Lemma 35. *Problem* (6.3) *is maximized by* $f^*(x) \propto g(x)^{-1}$*, i.e.*

$$f^*(x) := \frac{g(x)^{-1}}{\int_{\mathcal{X}^*} g(y)^{-1} dy}. \tag{6.4}$$

Proof. First, we note that f^* trivially satisfies the constraints in (6.3): it is normalized so it integrates to 1 over $\mathcal{X}^*$, and it is nonnegative since $g(x)$ (and hence $g(x)^{-1}$) is nonnegative. We also note that it exists since $g(x) \geqslant g_{\min}$ over $\mathcal{X}^*$ and $\mathcal{X}^*$ has finite volume (since it is bounded), so $\int_{\mathcal{X}^*} g(y)^{-1}\, dy \leqslant g_{\min}^{-1} \mathrm{Vol}_{\mathcal{X}}(\mathcal{X}^*) < \infty$.

We then note that scaling $g(x)$ by a constant c yields

$$\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} (c\, g(x))^{-\frac{1}{\gamma}}\, dx = c^{-\frac{1}{\gamma}} \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx \tag{6.5}$$

Thus it does not affect what f maximizes the objective function; hence we may assume without loss of generality that $\int_{\mathcal{X}^*} g(x)^{-1}\, dx = 1$ (of course we will have to put in the correct scale when evaluating the optimal value).

We will now show Lemma 35 using Hölder's Inequality. We first define:

$$\tilde{f}(x) := f(x)^{1-\frac{1}{\gamma}} \text{ and } \tilde{g}(x) = g(x)^{-\frac{1}{\gamma}} \tag{6.6}$$

and note that since by definition $f(x), g(x) \geqslant 0$ for all x, we know that $\tilde{f}(x), \tilde{g}(x) \geqslant 0$ for all x, so we can ignore the absolute value function in the statement of Hölder's Inequality. We then define the constant $\eta := \frac{\gamma}{\gamma - 1}$. Note that $\frac{1}{\eta} + \frac{1}{\gamma} = 1$, as required by Hölder's Inequality. Thus:

$$\int_{\mathcal{X}^*} \tilde{f}(x)\, \tilde{g}(x) dx \leqslant \Big(\int_{\mathcal{X}^*} \tilde{f}(x)^{\eta} dx \Big)^{\frac{1}{\eta}} \Big(\int_{\mathcal{X}^*} \tilde{g}(x)^{\gamma} dx \Big)^{\frac{1}{\gamma}}. \tag{6.7}$$

But, using the definitions from above,

$$\tilde{f}(x)^{\eta} = \big(f(x)^{1-\frac{1}{\gamma}} \big)^{\frac{\gamma}{\gamma-1}} = f(x) \text{ and } \tilde{g}(x)^{\gamma} = \big(g(x)^{-\frac{1}{\gamma}} \big)^{\gamma} = g(x)^{-1} \tag{6.8}$$

so we can rewrite equation (6.7) as

$$\int_{\mathcal{X}^*} \tilde{f}(x)\, \tilde{g}(x) dx \leqslant \Big(\int_{\mathcal{X}^*} f(x) dx \Big)^{\frac{1}{\eta}} \Big(\int_{\mathcal{X}^*} g(x)^{-1} dx \Big)^{\frac{1}{\gamma}}; \tag{6.9}$$

however, by the condition that f is a probability density function and our assumption (without loss of generality) about $g(x)^{-1}$, we know that

$$\int_{\mathcal{X}^*} f(x) dx = \int_{\mathcal{X}^*} g(x)^{-1} dx = 1 \tag{6.10}$$

implying that the right hand side of expression (6.9) is just 1. Thus,

$$\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}}\, g(x)^{-\frac{1}{\gamma}} dx = \int_{\mathcal{X}^*} \tilde{f}(x)\, \tilde{g}(x) dx \leqslant 1 \tag{6.11}$$

for any probability density function f. But, using f^* as defined in the statement of Lemma 35, it is trivial to see that

$$\int_{\mathcal{X}^*} f^*(x)^{1-\frac{1}{\gamma}}\, g(x)^{-\frac{1}{\gamma}} dx = 1 \tag{6.12}$$

thus showing that f^* is the maximizing density function. □

Remark 23. *Density f^* makes the lucrativity function constant over $\mathcal{X}^*$, as for all $x \in \mathcal{X}^*$,*

$$\mathfrak{c}^*(x) = (f^*(x) g(x))^{\frac{1}{\gamma}} = \Big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \Big)^{-\frac{1}{\gamma}}. \tag{6.13}$$

Lemma 36. *The optimal value of the problem given in* (6.3) *is*

$$\max_f \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx = \Big(\int_{\mathcal{X}^*} g(x)^{-1}\, dx \Big)^{-\frac{1}{\gamma}} \tag{6.14}$$

Proof. Plugging f^* into expression (6.3) gives

$$\int_{\mathcal{X}^*} f^*(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx = \int_{\mathcal{X}^*} \Big(\frac{g(x)^{-1}}{\int_{\mathcal{X}^*} g(y)^{-1} dy} \Big)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \tag{6.15}$$

$$= \int_{\mathcal{X}^*} \frac{g(x)^{-1}}{\big(\int_{\mathcal{X}^*} g(y)^{-1} dy \big)^{1-\frac{1}{\gamma}}} dx \tag{6.16}$$

$$= \frac{\int_{\mathcal{X}^*} g(x)^{-1} dx}{\big(\int_{\mathcal{X}^*} g(y)^{-1} dy \big)^{1-\frac{1}{\gamma}}} \tag{6.17}$$

$$= \Big(\int_{\mathcal{X}^*} g(x)^{-1} dx \Big)^{\frac{1}{\gamma}}. \tag{6.18}$$

□

Thus, if we distribute the points $X_1, X_2, \ldots, X_n$ according to f^* (iid), by Theorem 1, for any $\delta > 0$

$$\mathrm{TSP}_{\Pi}(X_1, \ldots, X_n) \leqslant (1+\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(x)^{-1} dx \Big)^{\frac{1}{\gamma}} \quad wvhp \tag{6.19}$$

holds for sufficiently large n. Thus, the deterministic $X_1, \ldots X_n$ in $\mathcal{X}^*$ which maximize the tour length achieves at least this length, giving us the lower bound in Theorem 2.

6.2 Adversarial Targets Upper Bound

We now consider the upper bound for the DTSP with adversarial targets. As with randomly-distributed targets, this differs significantly for symmetric and nonsymmetric dynamics; however, unlike with randomly-distributed targets, there is no general upper bound with order-of-growth $\Theta(n^{1-\frac{1}{\gamma}})$ for adversarial targets with nonsymmetric dynamics (see Proposition 20).

Remark 24. *For nonsymmetric dynamics, any distribution f satisfying the assumptions in Section 2.1 yields a very high probability upper bound with order-of-growth $\Theta(n^{1-\frac{1}{\gamma}})$ (though the dependence on f may be complicated and result in arbitrarily large constants). However, this is only guaranteed due to the assumption that f is*

full-dimensional and stays the same as $n \to \infty$, while adversarial targets can be arranged on a lower-dimensional subset of $\mathcal{X}$.

6.2.1 Adversarial targets with nonsymmetric dynamics

Proposition 20. There are examples of nonsymmetric dynamics Π and bounded $\mathcal{X}^* \subseteq \mathcal{X}$ such that

$$\sup_{X_1,\dots,X_n \in \mathcal{X}^*} \mathrm{TSP}_\Pi(X_1,\dots,X_n) = \Theta(n)\,. \tag{6.20}$$

Proof sketch of Proposition 20. We consider a Dubins car on $\mathbb{R}^2$ with a turning circle of radius 1, and we let $\mathcal{X}^* = [-1,1]^2$. Let $\mathcal{A} = \{\boldsymbol{x} : \|\boldsymbol{x}\|_2 = 1/1000\}$ (the circle centered at $\mathbf{0}$ with radius 1/1000), and let the n targets $X_1,\dots,X_n$ be uniformly spaced on $\mathcal{A}$, i.e. $X_i = (1/1000)(\cos(2\pi i/n), \sin(2\pi i/n))$. We then claim that the vehicle can't visit all n target points in less than $2n$ time. This is because the curvature bound on the Dubins car is significantly smaller than the curvature of the circle $\mathcal{A}$, so the shortest curve that can intersect 3 distinct points on $\mathcal{A}$ has length $\approx 2\pi$ (i.e. it can intersect $\mathcal{A}$ twice in a short period of time but then must circle back around all the way to return), meaning the shortest path that can visit n points has length at least (approximately) $(2\pi/3)n$; since $2 < (2\pi/3)$, we can use $2n$ as a firm lower bound for the length of the shortest tour that can visit n distinct points on $\mathcal{A}$ (which of course is a lower bound to the length of the shortest tour that can visit $X_1,\dots,X_n$), for sufficiently large n. □

Incidentally, this $\Theta(n)$ order-of-growth is the worst possible under the assumptions in Section 2.1, specifically because of Assumption 4: since any two configurations $\boldsymbol{q}, \boldsymbol{q}' \in [\mathcal{X}^*]^{\mathcal{Q}}$ satisfies $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant C$ for some fixed C (which depends on $\mathcal{X}^*$), given any $X_1,\dots,X_n$ we may arbitrarily choose $\boldsymbol{q}_i \in [X_i]^{\mathcal{Q}}$ for all i and then visit them in any order to achieve a tour of length $\leqslant Cn$.

Remark 25. *The worst-case placement of the target points $X_1,\dots,X_n$ achieves a worse order-of-growth than a random placement of target points according to density*

function f satisfying the assumptions in Section 2.1 because f is required to represent a continuous distribution on a full-dimensional set; if f were allowed to represent the uniform distribution on the circle $\mathcal{A}$ from the proof of Proposition 20, which is not full-dimensional, the worse order-of-growth $\Theta(n)$ would be achieved.

6.2.2 Adversarial targets with symmetric dynamics

We now prove the upper bound in Theorem 2 using the Symmetric Hierarchical Cell Structure algorithm developed in Chapter 4. Consider a Symmetric HCS cover over $\mathcal{X}^*$ with scaling parameter s and overlap parameter α at scale ε_0, which we know exists when Meta-Assumption 1 holds (see Proposition 4). The SHCS induces a specific algorithm, which by Proposition 13 generates a tour of length bounded above by

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} + C_{\varepsilon_0} \tag{6.21}$$

where n_j is the number of targets that falls into cell j and C_{ε_0} is a constant denoting the time needed to travel between cells. Let us denote $p_j = n_j/n$ and $\boldsymbol{p} = (p_1, \ldots, p_m)$; since $\sum_j n_j = n$, we know that $\boldsymbol{p}$ is a probability vector. Ignoring the constant C_{ε_0}, we can re-write the above as

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} = 6s\varepsilon_0 n^{1-\frac{1}{\gamma}} \sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} \tag{6.22}$$

Note that in contrast with Section 4.3, here we can simply distribute the target points according to $n_1, \ldots, n_m$ so there is no need for probabilistic analysis. Nevertheless, it will be useful to view $\boldsymbol{p}$ as a probability vector and, as before, set $Z \sim \boldsymbol{p}$ and to consider

$$\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} = \mathbb{E}_{Z\sim\boldsymbol{p}}[p_Z^{-\frac{1}{\gamma}}] \tag{6.23}$$

We then consider the probability distribution induced over the cells by $X \sim f^*$ where f^* is the worst-case probability distribution over $\mathcal{X}^*$, i.e. $f^*(x) \propto g(x)^{-1}$. Let $j(x)$ be the index of the cell $x \in \mathcal{X}^*$ falls into; if $\boldsymbol{p}^*$ is induced this way, then $Z \sim \boldsymbol{p}^*$ is equivalent to $Z = j(X)$ where $X \sim \boldsymbol{p}^*$. This probability distribution is

$$p_j^* = \mathbb{P}_{X \sim f^*}[X \in S_j] = \int_{S_j} f^*(x)\,dx = \frac{\int_{S_j} g(x)^{-1}\,dx}{\int_{\mathcal{X}^*} g(y)^{-1}\,dy}. \tag{6.24}$$

Therefore, we can re-write

$$\mathbb{E}_{Z \sim \boldsymbol{p}^*}[(p_Z^*)^{-\frac{1}{\gamma}}] = \mathbb{E}_{Z \sim \boldsymbol{p}^*}\left[\left(\frac{\int_{S_Z} g(x)^{-1}\,dx}{\int_{\mathcal{X}^*} g(y)^{-1}\,dy}\right)^{-\frac{1}{\gamma}}\right] \tag{6.25}$$

$$= \left(\int_{\mathcal{X}^*} g(y)^{-1}\,dy\right)^{\frac{1}{\gamma}} \mathbb{E}_{Z \sim \boldsymbol{p}^*}\left[\left(\int_{S_Z} g(x)^{-1}\,dx\right)^{-\frac{1}{\gamma}}\right] \tag{6.26}$$

Finally, we will analyze $\int_{S_j} g(x)^{-1}\,dx$ for any $j \in [m]$. Note that we can make ε_0 arbitrarily small because a HCS at scale ε_0 with scale factor s and overlap parameter α can be changed to a HCS at scale ε_0/s with the same parameters (and then to an HCS at scale ε_0/s^2 and so forth) by taking its sub-cells as the new cells. Therefore, letting $\check{g}^{(\eta)}$ be the η-regularized lower approximation of g, for any $\delta_1, \delta_2 > 0$ we can set ε_0 sufficiently small so that for any cell center $\boldsymbol{q}_j$ and configuration $\boldsymbol{q}'$ such that $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon_0$,

$$\mathrm{Vol}_{\mathcal{X}}(S_j) \geqslant (1-\delta_1)\alpha \mathrm{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon_0}(\boldsymbol{q}_j)) \geqslant (1-\delta_2)(1-\delta_1)\alpha \check{g}^{(\eta)}([\boldsymbol{q}]_{\mathcal{X}})\varepsilon_0^\gamma \tag{6.27}$$

$$\implies \check{g}^{(\eta)}(x) \leqslant (1-\delta_1)^{-1}(1-\delta_2)^{-1}\alpha^{-1}\varepsilon_0^{-\gamma}\mathrm{Vol}_{\mathcal{X}}(S_j) \tag{6.28}$$

$$\implies \check{g}^{(\eta)}(x)^{-1} \geqslant (1-\delta_1)(1-\delta_2)\alpha\varepsilon_0^\gamma \mathrm{Vol}_{\mathcal{X}}(S_j)^{-1} \tag{6.29}$$

for all $x \in S_j$ (since $S_j \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q})$). Furthermore, for any $\delta_3 > 0$, we can set η sufficiently small so that for all j,

$$\int_{S_j} g(x)^{-1}\,dx \geqslant (1-\delta_3)\int_{S_j} \check{g}^{(\eta)}(x)^{-1}\,dx \tag{6.30}$$

But then we can conclude that for any j,

$$\int_{S_j} g(x)^{-1}\,dx \geqslant (1-\delta_3)\int_{S_j} \check{g}^{(\eta)}(x)^{-1}\,dx \tag{6.31}$$

$$\geqslant (1-\delta_1)(1-\delta_2)(1-\delta_3)\alpha\varepsilon_0^{\gamma}\int_{S_j}\mathrm{Vol}_{\mathcal{X}}(S_j)^{-1}\,dx \tag{6.32}$$

$$= (1-\delta_1)(1-\delta_2)(1-\delta_3)\alpha\varepsilon_0^{\gamma} \tag{6.33}$$

and hence we can flip it around and get

$$\Big(\int_{S_j} g(x)^{-1}\,dx\Big)^{-\frac{1}{\gamma}} \leqslant (1-\delta_1)^{-\frac{1}{\gamma}}(1-\delta_2)^{-\frac{1}{\gamma}}(1-\delta_3)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}\,. \tag{6.34}$$

Since this holds for all j, we know that

$$\mathbb{E}_{Z\sim\boldsymbol{p}^*}\left[\Big(\int_{S_Z} g(x)^{-1}\,dx\Big)^{-\frac{1}{\gamma}}\right] \leqslant (1-\delta_1)^{-\frac{1}{\gamma}}(1-\delta_2)^{-\frac{1}{\gamma}}(1-\delta_3)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}\,. \tag{6.35}$$

We can then plug this in to get a bound on $\mathbb{E}_{Z\sim\boldsymbol{p}}[(p^*_Z)^{-\frac{1}{\gamma}}] = \sum_{j=1}^m (p_j^*)^{1-\frac{1}{\gamma}}$. Then we plug the bound into expression (6.22) to get

$$6s\varepsilon_0 n^{1-\frac{1}{\gamma}}\sum_{j=1}^m (p_j^*)^{1-\frac{1}{\gamma}} \leqslant 6s\alpha^{-\frac{1}{\gamma}}n^{1-\frac{1}{\gamma}}\Big(\int_{\mathcal{X}^*} g(y)^{-1}\,dy\Big)^{\frac{1}{\gamma}}\big((1-\delta_1)(1-\delta_2)(1-\delta_3)\big)^{-\frac{1}{\gamma}} \tag{6.36}$$

One wrinkle is that while this is true of $\boldsymbol{p}^*$ induced by f^*, our n_j's must be integers which means they cannot in general exactly represent $\boldsymbol{p}^*$. However, as $n\to\infty$, the $p_j = n_j/n$ can come arbitrarily close, and for any δ_4 there is some $n^{(\delta_4)}$ such that for all $n \geqslant n^{(\delta_4)}$, there are some $n_1,\ldots,n_m$ such that if $p_j = n_j/n$ then

$$\sum_{j=1}^m p_j^{1-\frac{1}{\gamma}} \leqslant (1+\delta_4)\sum_{j=1}^m (p_j^*)^{1-\frac{1}{\gamma}} \tag{6.37}$$

Then, for any $\delta_5 > 0$, we can set $\delta_1,\delta_2,\delta_3,\delta_4 > 0$ such that

$$1+\delta_5 = (1-\delta_1)^{-\frac{1}{\gamma}}(1-\delta_2)^{-\frac{1}{\gamma}}(1-\delta_3)^{-\frac{1}{\gamma}}(1+\delta_4) \tag{6.38}$$

which then gives that for $n \geqslant n^{(\delta_4)}$, distributing the n_j according to this approximation of $p_j^* n$, we get that

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} \leqslant (1+\delta_5) 6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \Big)^{\frac{1}{\gamma}} \tag{6.39}$$

Thus, for sufficiently large n, the HCS induces an algorithm which collects all the target points in at most $(1+\delta_5)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \big)^{\frac{1}{\gamma}} + C_{\varepsilon_0}$ length; and for any $\delta > 0$ we can choose $0 < \delta_5 < \delta$ and choose n sufficiently large that

$$(1+\delta_5) 6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \Big)^{\frac{1}{\gamma}} + C_{\varepsilon_0} \leqslant (1+\delta) 6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \Big)^{\frac{1}{\gamma}} \tag{6.40}$$

Since our specific algorithm can then always collect the target points in at most this amount of time, the optimal algorithm will do at least as well, and hence

$$\sup_{X_1,\ldots,X_n \in \mathcal{X}^*} \mathrm{TSP}_{\Pi}(X_1, \ldots, X_n) \leqslant (1+\delta) 6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \Big(\int_{\mathcal{X}^*} g(y)^{-1}\, dy \Big)^{\frac{1}{\gamma}} \tag{6.41}$$

and we have finished proving the upper bound to Theorem 2.

Remark 26. *Since $\sum_j p_j = 1$ by definition and $1 - \frac{1}{\gamma} < 1$, the optimal values of $p_1, \ldots, p_m$ to maximize $\sum_{j=1}^m p_j^{1-\frac{1}{\gamma}}$ are actually $p_1 = \cdots = p_m = \frac{1}{m}$, which yields*

$$\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} = m^{\frac{1}{\gamma}}\,. \tag{6.42}$$

This aligns with the intuition that to make the target points hard to visit, they should be spread out over the cells (at a given scale) evenly. While this was not used in the analysis as m is hard to compute directly, it is useful to know in order to find n sufficiently large so that every $p_j = \frac{n_j}{n}$ can approximate $\frac{1}{m}$. This 'sufficiently large' value increases with m, which increases as the scale ε_0 decreases.

Chapter 7

Conclusion

In this chapter we conclude and discuss possible directions for future work.

In this work we gave a very general formulation of the Dynamic Stochastic Traveling Salesman Problem (DSTSP), capturing a wide range of dynamic systems (with a specific focus on control-affine systems, though not limited to such), which extends previous formulations to systems on manifold (rather than Euclidean) workspaces. To do this, we develop a novel technique of discretizing the workspace, adapted from Arias-Castro et al. ([11], 2005) and Adler, Karaman ([20], 2016), as well as a novel Dynamic TSP algorithm for symmetric dynamically-constrained vehicles based on Hierarchical Cell Structures. We also showed that our techniques work even in cases where the natural properties needed for the system to be well-behaved are only approximate.

Under this expanded formulation, we showed novel results describing not only how the length of the tour scales with the number n of iid targets, but also the effect of the density function f of the target point distribution on the length of the fastest tour. In particular, we defined the *agility function* g induced by the dynamics over the workspace and showed for all dynamics in our formulation a very-high-probability lower bound on the tour length that both scale according to

$$n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{7.1}$$

where $\mathcal{X}_f$ is the support of f, and for *symmetric* dynamics a matching upper bound which scales according to (7.1). For symmetric dynamics, this yields very-high-probability lower and upper bounds that differ by only a constant in terms of f and n.

We then use these results to analyze the case of *adversarial* target points, in which the targets are distributed by an adversary on some bounded region $\mathcal{X}^*$ with the objective of maximizing the tour length. For symmetric vehicles, we show similar (but deterministic) matching lower and upper bounds which scale according to

$$n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}^*} g(x)^{-1} \, dx \right)^{\frac{1}{\gamma}} \tag{7.2}$$

with the lower bound also applying to non-symmetric vehicles.

7.1 Future work

While we establish lower and upper bounds to the tour lengths of the DSTSP and the Adversarial DTSP which are tight up to a constant factor in both f and n, many interesting lines of future work remain.

Our work establishes a $12s\alpha^{-\frac{1}{\gamma}}\beta$ multiplicative gap between the lower and upper bounds for symmetric dynamics, where s and α are parameters measuring the effectiveness of our cell-tiling algorithm (and $s = 2$ for control-affine systems, see Proposition 4), and β (see Definition 8) is a parameter describing the difficulty of covering larger reachable sets with smaller ones at small scales. In particular, depending on the dynamics, $\alpha \leqslant 1$ is the degree to which the tiles fill the reachable sets in which they are inscribed, and generally measures how efficiently the reachable sets can be used to tile the workspace. However, it is not settled whether α really represents how efficiently the vehicle can visit the targets, or whether it is a product of our specific Symmetric Hierarchical Cell Structure algorithm.

Theorem 1 almost completely extends the result of Beardwood et al. ([9], 1959) on the Euclidean Stochastic TSP to general symmetric dynamic constraints. In par-

ticular, the Beardwood-Halton-Hammersley Theorem states that for the Euclidean Stochastic TSP on $\mathbb{R}^d$ with targets $X_1, X_2, \ldots \overset{iid}{\sim} f$, there is some constant β_d (depending only on d) such that

$$\mathrm{TSP}(X_1, \ldots, X_n) \to \beta_d n^{1-\frac{1}{d}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{d}}\, dx \tag{7.3}$$

almost surely as $n \to \infty$. Our result extends the $n^{1-\frac{1}{d}}$ to $n^{1-\frac{1}{\gamma}}$ (Euclidean paths in $\mathbb{R}^d$ can be formulated as a dynamic system with $\gamma = d$) and extends $\int f(x)^{1-\frac{1}{d}}\, dx$ to $\int f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx$ (when considering only the length of Euclidean paths, g is constant and can therefore be moved to the constant outside the integral); however, it leaves open the possibility that the constant outside fluctuates between our derived lower and upper bounds and doesn't go to a particular value in the range $[\beta^{-1}, 12s\alpha^{-\frac{1}{\gamma}}]$. Showing that there is some constant β_Π for any symmetric dynamic constraints Π and $X_1, X_2, \ldots \overset{iid}{\sim} f$

$$\lim_{n\to\infty} \frac{\mathrm{TSP}_\Pi(X_1, \ldots, X_n)}{n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx} = \beta_\Pi \tag{7.4}$$

would complete the extension of the Beardwood-Halton-Hammersley Theorem to our more general setting (with dynamic constraints), though it is also possible that such an extension might only be possible for Π satisfying stronger regularity conditions.

Furthermore, the techniques developed here (hierarchical cell tiling strategies and Cost-Balanced Orienteering) may yield results on the following extensions of or problems related to the DSTSP or Stochastic Orienteering:

- As discussed in Appendix C, our techniques can yield strong upper bounds for the Euclidean TSP where the targets are distributed over a subset of fractal dimension; in particular, the Symmetric Hierarchical Cell Structure is suited to such cases. However, this leaves open the question of fractal distribution of targets with dynamically-constrained tours, as well as matching lower bounds.

- Another interesting extension may be to have targets be sets within the workspace

rather than points. While having the targets be balls of fixed size is generally trivial in the limit (a fixed-length trajectory can 'sweep' the space and visit all possible ε-balls), possibilities such as having the targets be random k-dimensional affine subspaces embedded in a d-dimensional workspace may be solvable with our techniques.

- The Orienteering problem can also be thought of as trying to pass a curve of limited length obeying constraints through as many targets as possible, which has applications in image processing [11], and an important extension of this is to pass a higher-dimensional region of limited size obeying dynamic constraints through targets.

- Another problem related to Orienteering is the problem of trying to fit a function satisfying local constraints (for instance, Lipschitz continuity) to target points with a loss function (as opposed to trying to exactly pass through as many targets as possible).

- Finally, the TSP is related to a number of other problems in which mobile agents must visit unordered targets, such as the Traveling Repairperson Problem [10] in which targets pop up stochastically (for instance, according to a Poisson process) over time and must be visited with a minimum delay per target.

Finally, it remains open whether a guaranteed approximation algorithm exists for the Dynamic TSP in general. Theorem 1 shows that our Hierarchical Collection Problem algorithms achieve a constant factor approximation of the shortest tour with very high probability (the difference between the symmetric and nonsymmetric cases primarily being that the approximation factor can depend arbitrarily on the density f of the target point distribution); this in turn shows that our algorithms also achieve a constant factor approximation of the expected length of the shortest tour (see Corollary 1 and Appendix A.1.1). Finally, Theorem 2 shows that our Symmetric HCP algorithm achieves a constant factor approximation of the shortest tour through adversarially-distributed target points. However, our analysis leaves open the possibility that the targets may be cleverly placed so as to permit a very short tour

while the algorithms only find loner ones (though when the targets are random this becomes vanishingly unlikely), and it is interesting to consider whether an algorithm exists which prevents this. See Appendix A.2 for a full discussion.

Bibliography

[1] Richard M. Karp, *Reducibility among Combinatorial Problems*, pp. 85–103, Springer US, Boston, MA, 1972.

[2] Christos H. Papadimitriou, "The euclidean travelling salesman problem is np-complete," *Theoretical Computer Science*, vol. 4, no. 3, pp. 237–244, 1977.

[3] S. M. LaValle, *Planning Algorithms*, Cambridge University Press, Cambridge, U.K., 2006, Available at http://planning.cs.uiuc.edu/.

[4] Jerome Ny, Eric Feron, and Emilio Frazzoli, "On the dubins traveling salesman problem," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 265–270, 2012.

[5] A. M. Frieze and J. E. Yukich, *Probabilistic Analysis of the TSP*, pp. 257–307, Springer US, Boston, MA, 2007.

[6] Richard M. Karp, "Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane," *Mathematics of Operations Research*, vol. 2, no. 3, pp. 209–224, 1977.

[7] Dimitris Bertsimas, Philippe Chervi, and Michael Peterson, "Computational approaches to stochastic vehicle routing problems," *Transportation Science*, vol. 29, no. 4, pp. 342–352, 1995.

[8] Ludger Rüschendorf Birgit Anthes, "On the weighted euclidean matching problem in r^d," *Applicationes Mathematicae*, vol. 28, no. 2, pp. 181–190, 2001.

[9] Jillian Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 55, no. 4, pp. 299–327, 1959.

[10] Sleiman Itani, *Dynamic systems and subadditive functionals*, Ph.D. thesis, Massachusetts Institute of Technology, 2012.

[11] Ery Arias-Castro, David L. Donoho, Xiaoming Huo, and Craig A. Tovey, "Connect the dots: How many random points can a regular curve pass through?," *Advances in Applied Probability*, vol. 37, no. 3, pp. 571–603, 2005.

[12] John J. Enright and Emilio Frazzoli, "The traveling salesman problem for the reeds-shepp car and the differential drive robot," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 3058–3064.

[13] K. Savla, E. Frazzoli, and F. Bullo, "On the point-to-point and traveling salesperson problems for dubins' vehicle," in *Proceedings of the 2005, American Control Conference, 2005.*, 2005, pp. 786–791 vol. 2.

[14] Ketan Savla, Francesco Bullo, and Emilio Frazzoli, "On traveling salesperson problems for a double integrator," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 5305–5310.

[15] Ketan Savla, Emilio Frazzoli, and Francesco Bullo, "On the dubins traveling salesperson problems: Novel approximation algorithms," in *Robotics: Science and Systems*, 2006.

[16] Ketan Savla, Emilio Frazzoli, and Francesco Bullo, "Traveling salesperson problems for the dubins vehicle," *IEEE Transactions on Automatic Control*, vol. 53, no. 6, pp. 1378–1391, 2008.

[17] Sleiman Itani, Emilio Frazzoli, and Munther A Dahleh, "Travelling salesperson problem for dynamic systems," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 13318–13323, 2008, 17th IFAC World Congress.

[18] Ery Arias-Castro, David L Donoho, and Xiaoming Huo, "Adaptive multiscale detection of filamentary structures in a background of uniform random points," *The Annals of Statistics*, pp. 326–349, 2006.

[19] Frédéric Jean, *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning*, SpringerBriefs in Mathematics. Springer International Publishing, 2014.

[20] Aviv Adler and Sertac Karaman, "The stochastic traveling salesman problem and orienteering for kinodynamic vehicles," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 2788–2795.

[21] Patrick Rebeschini, "Algorithmic foundations of learning, lecture 7 notes," September 2020.

[22] Richard Durrett, *Essentials of stochastic processes*, Springer texts in statistics. Springer, Cham, 3rd ed. edition, 2016.

[23] Michael R Garey and David S Johnson, "Computers and intractability," *A Guide to the*, 1979.

[24] Nicos Christofides, "Worst-case analysis of a new heuristic for the travelling salesman problem," Tech. Rep., Carnegie-Mellon Univ Pittsburgh Pa Management Sciences Research Group, 1976.

Appendix A

Probability, Optimization, and Complexity

In this chapter we discuss and define important notions in probability, optimization, and algorithms, as well as prove a few important lemmas. Aside from Appendix A.1.1, all notions discussed here are well-known and standard; and while (to our knowledge) the notion of "with very high probability" is novel, the results and proofs contained in Appendix A.1.1 are elementary and analogous to known results about the related notion of "with high probability".

A.1 Probability and Random Variables

Because our results primarily deal with a stochastic version of the Traveling Salesman Problem (TSP), in which the targets are randomly placed into a space, the length of the shortest TSP tour is a random variable, and we deal with various notions from probability theory which we will briefly outline here. When we discuss *probability events*, measurable subsets of an elementary outcome space Ω of the type

$$A = \{\omega \in \Omega : [\text{some condition}]\} \tag{A.1}$$

we will drop the ω and just equate the event with the condition. For instance, if we have a random variable Z, we denote the event that $Z \geqslant a$ as $A = \{Z \geqslant a\}$ (rather than the more complete $A = \{\omega \in \Omega : Z(\omega) \geqslant a\}$).

We will also deal heavily with *stochastic domination*:

Definition 30. A random variable Z *stochastically dominates* a random variable Z' if for any $z \in \mathbb{R}$,

$$\mathbb{P}[Z \leqslant z] \leqslant \mathbb{P}[Z' \leqslant z] \tag{A.2}$$

We denote this relationship $Z \succeq Z'$.

This is equivalent to saying the cumulative distribution function (CDF) of Z is a lower bound to that of Z'; it is also equivalent to say that there is a coupling (of the distributions of Z, Z') such that $Z \geqslant Z'$ almost surely.

A.1.1 With Very High Probability

We now take a closer look at the notion of 'with very high probability'. Recall that an infinite sequence of events A_n parameterized by integer n happens *with very high probability* (Definition 7) if there are constants $c_1, c_2, c_3 > 0$ such that, for all sufficiently large n,

$$\mathbb{P}[A_n] \geqslant 1 - c_1 e^{-c_2 n^{c_3}} \tag{A.3}$$

and recall that this means it converges to 1 faster than any inverse polynomial as $n \to \infty$. It also gives us some additional information for which "with high probability" is not sufficient:

Lemma 37. *If $\{A_n\}$ happens with very high probability, then (where $\neg$ means 'not')*

$$\mathbb{P}[(\max n : \neg A_n) < \infty] = 1 \tag{A.4}$$

i.e. almost surely there is some finite n^* s.t. A_n happens for all $n > n^*$.

Proof. We show this by using the Borel-Cantelli Lemma, which states that if a sequence of events $\{E_n\}$ has $\sum_{n=1}^{\infty} \mathbb{P}[E_n] < \infty$, then

$$\mathbb{P}[(\max n : E_n) < \infty] = 1 \tag{A.5}$$

But letting $E_n = \neg A_n$, we have

$$\mathbb{P}[E_n] \leqslant c_1 e^{-c_2 n^{c_3}} \tag{A.6}$$

for sufficiently large n (say, $n > n'$). Of course, for $n \leqslant n'$, we have the trivial bound $\mathbb{P}[E_n] \leqslant 1$. Thus,

$$\sum_{n=1}^{\infty} \mathbb{P}[E_n] \leqslant n' + \sum_{n>n'} c_1 e^{-c_2 n^{c_3}} < \infty \tag{A.7}$$

since $c_1 e^{-c_2 n^{c_3}}$ (which decreases faster than n^{-k} for any k) is a convergent series.

Therefore, $\mathbb{P}[(\max n : \neg A_n) < \infty] = 1$, as desired. □

We then also define a related notion of a random variable growing "with very high probability":

Definition 31. If $\{Z_n\}$ is a sequence of random variables and $\alpha(n) : \mathbb{Z}_{>0} \to \mathbb{R}_{\geqslant 0}$ is any function then we say "$Z_n \overset{wvhp}{=} O(\alpha(n))$" if there exists a constant $c > 0$ such that the sequence of events $A_n = \{Z_n \leqslant c\alpha(n)\}$ happens with very high probability.

We similarly say $Z_n \overset{wvhp}{=} \Omega(\alpha(n))$ if there exists a constant $c > 0$ such that $A_n = \{Z_n \leqslant c\alpha(n)\}$ happens with very high probability, and $Z_n \overset{wvhp}{=} \Theta(\alpha(n))$ if there exist constants $c_2 > c_1 > 0$ such that $A_n = \{c_1\alpha(n) \leqslant Z_n \leqslant c_2\alpha(n)\}$ happens with very high probability.

Finally, we say that "$Z_n \overset{wvhp}{\to} \alpha(n)$" (multiplicatively) if there exists a nonneg-

ative sequence δ_n satisfying $\lim_{n\to\infty} \delta_n = 0$ such that

$$B_n = \{1 - \delta_n \leqslant Z_n/\alpha(n) \leqslant 1 + \delta_n\} \tag{A.8}$$

happens with very high probability.

The following then hold:

Lemma 38. *If $Z_1, Z_2, \ldots$ and $Z'_1, Z'_2, \ldots$ are sequences of random variables parameterized by n such that $Z_n \succeq Z'_n$ for all n, then*

$$Z_n \overset{wvhp}{=} O(\alpha(n)) \implies Z'_n \overset{wvhp}{=} O(\alpha(n)) \tag{A.9}$$

$$\textit{and} \quad Z'_n \overset{wvhp}{=} \Omega(\alpha(n)) \implies Z_n \overset{wvhp}{=} \Omega(\alpha(n)) \tag{A.10}$$

Proof. This follows from the definitions: if $Z_n \overset{wvhp}{=} O(\alpha(n))$ then we have some $c > 0$ and $c_1, c_2, c_3 > 0$ such that

$$\mathbb{P}[Z_n > c\alpha(n)] \leqslant c_1 e^{-c_2 n^{c_3}} \tag{A.11}$$

Since $Z_n \succeq Z'_n$ implies $\mathbb{P}[Z'_n > c\alpha(n)] \leqslant \mathbb{P}[Z_n > c\alpha(n)]$, we are done with Eq. (A.9). The proof of Eq. (A.10) is analogous. □

Lemma 39. *If $Z_n \overset{wvhp}{\to} \alpha(n)$, then $Z_n/\alpha(n) \to 1$ almost surely.*

Proof. By Lemma 37, with probability 1 there is some finite n^* such that B_n happens for all $n > n^*$. Thus, for $n > n^*$, we have

$$1 - \delta_n \leqslant Z_n/\alpha(n) \leqslant 1 + \delta_n \tag{A.12}$$

and since $\delta_n \to 0$ as $n \to \infty$, the result follows. □

The next lemma shows that our main results on the very-high-probability behavior of the DSTSP translate directly to expected value.

Proposition 21. If $\{Z_n\}$ is a sequence of nonnegative random variables such that $Z_n = \Theta(n^k)$ with very high probability and $Z_n = O(n^\ell)$ (non-probabilistic upper bound), where $0 < k \leqslant \ell$, then $\mathbb{E}[Z_n] = \Theta(n^k)$.

Proof. We give upper and lower bounds for $\mathbb{E}[Z_n]$ by splitting it into three parts based on the following events: let $A_n^{(1)} = \{Z_n < a_1 n^k\}$; $A_n^{(2)} = \{a_1 n^k \leqslant Z_n \leqslant a_2 n^k\}$ (where $0 < a_1 < a_2$ and a_2 is the constant factor in the wvhp upper bound as in Definition 31); and $A_n^{(3)} = \{a_2 n^k < Z_n\}$. Then:

$$\mathbb{E}[Z_n] = \sum_{i=1}^{3} \mathbb{P}[A_n^{(i)}]\, \mathbb{E}[Z_n \,|\, A_n^{(i)}] \tag{A.13}$$

We can then upper bound and lower bound these since: $\mathbb{P}[A_n^{(2)}] \geqslant 1 - c_1 e^{-c_2 n^{c_3}}$ (for $c_1, c_2, c_3 > 0$ as in Definition 7) we know that $\mathbb{P}[A_n^{(1)}], \mathbb{P}[A_n^{(3)}] \leqslant c_1 e^{-c_2 n^{c_3}}$; and $\mathbb{E}[Z_n \,|\, A_n^{(i)}] = O(n^\ell)$ (and are nonnegative), and $\mathbb{E}[Z_n \,|\, A_n^{(2)}] = \Theta(n^k)$ (by definition).

Then since $c_1 e^{-c_2 n^{c_3}} \to 0$ faster than $a_2^{-1} n^{-\ell}$, terms $i = 1$ and 3 vanish as $n \to \infty$ while term $i = 2$ goes to $\Theta(n^k)$ and we are done. □

This shows that Corollary 1 follows directly from Theorem 1 and Proposition 1.

A.1.2 IID and Poisson Targets

A big part of our analysis depends on the technique of *Poissonization*, replacing n iid random targets with targets generated via an appropriate (possibly non-uniform-intensity) Poisson process. This small tweak means that the position and number of targets present in disjoint regions are independent (whereas with iid targets, more targets in one region means less, on average, for the others), greatly simplifying a lot of the analysis. Furthermore, one can more easily bound Poisson processes above and below because of the remarkable property that the sum of two independent Poisson processes is itself a Poisson process, even if their intensity functions have no relation. However, it needs to be shown that Poissonization doesn't alter the limiting behavior of the minimum-length tour (as $n \to \infty$).

To describe this, we need to clearly establish the meaning of iid (independently and identically distributed) targets in $\mathcal{X}$ and of targets distributed according to a Poisson process, and the connection between the two types of random distribution. We will not develop a complete description of these, but rather provide basic definitions and crucial lemmas. For a more complete treatment, see [22].

We also note that we will be dealing only with *continuous* distributions and Poisson processes with full-dimensional intensities, as our theorems are limited to these cases. It is an interesting open question what happens when targets have lower-dimensional distributions (especially for symmetric vehicles).

Definition 32. A *continuous probability distribution* F over $\mathcal{X}$ is a probability distribution taking $\mathcal{X}$ as its space of elementary outcomes and described by a measurable function $f : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ such that

$$\int_{\mathcal{X}} f(\boldsymbol{x}) d\boldsymbol{x} = 1 \tag{A.14}$$

in which, for any measurable subset $\mathcal{A} \subseteq \mathcal{X}$, if $X \sim F$,

$$\mathbb{P}[X \in \mathcal{A}] = \int_{\mathcal{A}} f(\boldsymbol{x}) d\boldsymbol{x} \tag{A.15}$$

f is referred to as the *probability density function* (PDF) of F.

For instance, in $\mathbb{R}^3$, the uniform distribution over the cube $[0,1]^3$ (with PDF $f = \mathbf{1}_{\boldsymbol{x} \in [0,1]^3}$, where $\mathbf{1}_{\mathcal{A}}$ is the indicator function which takes value 1 on $\mathcal{A}$ and 0 everywhere else) is continuous, whereas in $\mathbb{R}^3$ the uniform distribution over the line segment $[0,0,0]$ to $[1,0,0]$ is *not* considered "continuous" because no density function (integrated in 3 dimensions) can express it (since a line segment has measure 0 in $\mathbb{R}^3$ all functions with support only on the line segment integrate to 0).

$X_1, \ldots, X_n \overset{iid}{\sim} F$ refers to the case where $X_1, \ldots, X_n$ are all distributed according to F independently. If we want to ignore the indices and consider the targets as an

un-ordered n-element set, we write

$$\{X_1, \ldots, X_n\} \overset{iid}{\sim} F \tag{A.16}$$

Definition 33. A (full-dimensional) *Poisson process* over $\mathcal{X}$ is a random distribution described by a function $\lambda : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ such that $\bar{\lambda} = \int_{\mathcal{X}} \lambda(\boldsymbol{x}) d\boldsymbol{x} < \infty$ which does the following:

- pick $N \sim \text{Pois}(\bar{\lambda})$, where $\text{Pois}(\bar{\lambda})$ is the Poisson distribution with rate $\bar{\lambda}$;
- pick $X_1, \ldots, X_N \overset{iid}{\sim} F_\lambda$, where F_λ is the continuous distribution with probability density function $f_\lambda : \lambda/\bar{\lambda}$.

λ is referred to as the *intensity function* of the Poisson process, and we write $(X_1, \ldots, X_N) \sim \text{PoisProc}(\lambda)$ to denote that $(X_1, \ldots, X_N)$ is generated this way (noting that N is a random variable too). To denote the unordered set generated (i.e. remove the indices) we write $\{X_1, \ldots, X_N\} \sim \text{PoisProc}(\lambda)$.

The value $\bar{\lambda}$ denotes the expected number of points generated by $\text{PoisProc}(\lambda)$. Note that the Poisson process does not generate a fixed number of target points since N is also random.

This is admittedly not the most fundamental definition of the Poisson process with intensity λ, which can be considered as a limiting distribution of more easily defined discrete probability distributions (the definition we use is then a theorem about the more fundamental definition). However, it is sufficient for our purposes.

Remark 27. *It's worth noting the distinction between generating an ordered list of points or an unordered set. Generating n targets iid naturally applies indices, whereas Poisson processes are most fundamentally understood as being unordered point sets; since we want to link the two, we have notation for distinguishing both ordered and unordered versions, though we will more often be concerned with the unordered version.*

We now state some key properties of the Poisson process:

Proposition 22. Suppose $(X_1, \ldots, X_N) \sim \text{PoisProc}(\lambda)$ over $\mathcal{X}$:

1. For disjoint $\mathcal{A}, \mathcal{A}' \subseteq \mathcal{X}$, the random sets $\{X_1, \ldots, X_N\} \cap \mathcal{A}$ and $\{X_1, \ldots, X_N\} \cap \mathcal{A}'$ are independent (note that we use the set of targets rather than the indexed sequence). In particular, this also means $|\{X_1, \ldots, X_N\} \cap \mathcal{A}|$ and $|\{X_1, \ldots, X_N\} \cap \mathcal{A}'|$ (considered as random variables) are independent.

2. For any $\mathcal{A} \subseteq \mathcal{X}$,

$$\mathbb{E}\big[|\{X_1, \ldots, X_N\} \cap \mathcal{A}|\big] = \int_{\mathcal{A}} \lambda(\boldsymbol{x}) d\boldsymbol{x} \tag{A.17}$$

   As a special case, note that when $\bar{\lambda} = \int_{\mathcal{X}} \lambda(\boldsymbol{x}) d\boldsymbol{x}$ as in Definition 33:

$$\mathbb{E}[N] = \mathbb{E}\big[|\{X_1, \ldots, X_N\}|\big] = \mathbb{E}\big[|\{X_1, \ldots, X_N\} \cap \mathcal{X}|\big] = \bar{\lambda} \tag{A.18}$$

3. If $\{X_1, \ldots, X_N\} \sim \text{PoisProc}(\lambda)$ and $\{X'_1, \ldots, X'_{N'}\} \sim \text{PoisProc}(\lambda')$ independently, then

$$\{X_1, \ldots, X_N, X'_1, \ldots, X'_{N'}\} \sim \text{PoisProc}(\lambda + \lambda') \tag{A.19}$$

Point (2) indicates that if we have a distribution F with PDF f, and we want to emulate $X_1, \ldots, X_n \overset{iid}{\sim} F$ as closely as possible with a Poisson process, it is natural to consider PoisProc(nf): if $\lambda = nf$ then $f_\lambda = f$ and so $F_\lambda = F$; and $\mathbb{E}[N] = n$. Thus, in expectation n targets will be generated, and they will be distributed iid according to F. Of course this is not the same distribution since N is still random, but it is reasonable to expect that the two are fairly close, especially when n is large and concentration bounds on Pois(n) kick in. This observation is the basis of the Poissonization technique from the analysis of the Nonsymmetric Hierarchical Collection Problem in Section 5.2.

A.1.3 Simplification through Poissonization

These notions also allow us to greatly simplify the problem when we are only concerned about the order-of-growth because we can WLOG assume that the targets fall uniformly on a rectangular region. Formally, to define a region on a curved manifold we have to take an *adapted frame* (see Section 2.2 and Appendix B); however, we can show a more general result allowing simplification.

Proposition 23. For Π satisfying the assumptions in Section 2.1, if there is a finite set of bounded $\mathcal{X}_1, \ldots, \mathcal{X}_m \subseteq \mathcal{X}$ such that $\mathcal{X}_f \subseteq \bigcup_{j=1}^{m} \mathcal{X}_j$, and probability density functions $f_1, \ldots, f_m$ where each f_j has support $\mathcal{X}_j$ and there is some $c > 0$ such that $f_j(x) > c$ for all $x \in \mathcal{X}_j$, then if for all j,

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \overset{wvhp}{=} O(n^{1-\frac{1}{\gamma}}) \text{ when } X_1, \ldots, X_n \overset{iid}{\sim} f_j \tag{A.20}$$

we can conclude that for any f satisfying Assumption 5,

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \overset{wvhp}{=} O(n^{1-\frac{1}{\gamma}}) \text{ when } X_1, \ldots, X_n \overset{iid}{\sim} f\,. \tag{A.21}$$

Note that having a single $c > 0$ as the lower bound for the densities f_j is equivalent to having a separate $c_j > 0$ for each since we can use $c = \min_j c_j$.

In order to prove this, we consider the *Poissonization* of the target point distribution $X_1, \ldots, X_n \overset{iid}{\sim} f$, which is the distribution $\{X_1, \ldots, X_N\} \sim \mathrm{PoisProc}(n \cdot f)$ of target points; it replaces the n iid targets according to f with a Poisson process according to intensity $n \cdot f$, or, equivalently, it generates some $N \sim \mathrm{Pois}(n)$ and then distributes N targets iid according to f. We now show that the length of the shortest TSP tour keeps its order-of-growth when the target distribution is Poissonized:

Lemma 40. *If Π satisfies the assumptions in Section 2.1, then*

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \overset{wvhp}{=} \Theta(n^{1-\frac{1}{\gamma}}) \textit{ when } X_1, \ldots, X_n \overset{iid}{\sim} f \tag{A.22}$$

$$\iff \mathrm{TSP}_\Pi(X_1, \ldots, X_N) \overset{wvhp}{=} \Theta(n^{1-\frac{1}{\gamma}}) \textit{ when } \{X_1, \ldots, X_N\} \sim \mathrm{PoisProc}(n \cdot f) \tag{A.23}$$

Proof. This follows directly from the fact that if $N \sim \text{Pois}(n)$, then $N \leqslant 2n$ and $N \geqslant n/2$ with very high probability, and the fact that adding target points can only make the shortest DTSP tour longer. For clarity, we will let $X_1, X_2, \ldots \overset{iid}{\sim} f$ (an infinite sequence of random targets), and we will let the iid targets be the first n from this list, while the Poissonized targets are the first N where $N \sim \text{Pois}(n)$.

We then show both directions:

Forward: Suppose there are constants $0 < a < A$ and $0 < c_1, c_2, c_3$ such that

$$\mathbb{P}\big[\text{TSP}_\Pi(X_1, \ldots, X_n) \geqslant an^{1-\frac{1}{\gamma}}\big] \geqslant 1 - c_1 e^{-c_2 n^{c_3}} \tag{A.24}$$

$$\text{and } \mathbb{P}\big[\text{TSP}_\Pi(X_1, \ldots, X_n) \leqslant An^{1-\frac{1}{\gamma}}\big] \geqslant 1 - c_1 e^{-c_2 n^{c_3}} \tag{A.25}$$

(while technically we can have different c_1, c_2, c_3 for the upper and lower bounds, it's equivalent to assume they're the same since we can always choose the larger c_1 and the smaller c_2, c_3). We can also without loss of generality assume that c_1, c_2, c_3 also imply that for sufficiently large n,

$$\mathbb{P}[N \leqslant 2n] \geqslant 1 - c_1 e^{-c_2 (2n)^{c_3}} \text{ and } \mathbb{P}[N \geqslant n/2] \geqslant 1 - c_1 e^{-c_2 (n/2)^{c_3}} \tag{A.26}$$

Then we consider $\text{TSP}_\Pi(X_1, \ldots, X_N)$; then we conclude

$$\begin{aligned} &\text{TSP}_\Pi(X_1, \ldots, X_{\lceil n/2 \rceil}) \geqslant a(n/2)^{1-\frac{1}{\gamma}} \text{ and } N \geqslant n/2 \\ \implies &\text{TSP}_\Pi(X_1, \ldots, X_N) \geqslant a(n/2)^{1-\frac{1}{\gamma}} \geqslant (a/2)n^{1-\frac{1}{\gamma}} \end{aligned} \tag{A.27}$$

which in turn implies by the union bound

$$\mathbb{P}[\text{TSP}_\Pi(X_1, \ldots, X_N) \geqslant (a/2)n^{1-\frac{1}{\gamma}}] \tag{A.28}$$

$$\geqslant \mathbb{P}[\text{TSP}_\Pi(X_1, \ldots, X_{\lceil n/2 \rceil}) \geqslant a(n/2)^{1-\frac{1}{\gamma}} \text{ and } N \geqslant n/2] \tag{A.29}$$

$$\geqslant 1 - 2c_1 e^{-c_2 (n/2)^{c_3}} \tag{A.30}$$

yielding a very high probability lower bound with order-of-growth $n^{1-\frac{1}{\gamma}}$. To get the

upper bound, we use a similar trick:

$$\begin{aligned} &\mathrm{TSP}_\Pi(X_1,\ldots,X_{2n}) \leqslant A(2n)^{1-\frac{1}{\gamma}} \text{ and } N \leqslant 2n \\ \implies &\mathrm{TSP}_\Pi(X_1,\ldots,X_N) \leqslant A(2n)^{1-\frac{1}{\gamma}} \leqslant (2A)n^{1-\frac{1}{\gamma}} \end{aligned} \tag{A.31}$$

which in turn implies by the union bound

$$\mathbb{P}[\mathrm{TSP}_\Pi(X_1,\ldots,X_N) \leqslant (2A)n^{1-\frac{1}{\gamma}}] \tag{A.32}$$

$$\geqslant \mathbb{P}[\mathrm{TSP}_\Pi(X_1,\ldots,X_{2n}) \leqslant A(2n)^{1-\frac{1}{\gamma}} \text{ and } N \leqslant 2n] \tag{A.33}$$

$$\geqslant 1 - 2c_1 e^{-c_2(2n)^{c_3}} \tag{A.34}$$

yielding a very high probability upper bound with order-of-growth $n^{1-\frac{1}{\gamma}}$. Combining the two yields the very high probability order of growth.

Backward: For any $0 < a < A$, define

$$p_1(n;a) := 1 - \mathbb{P}\big[\mathrm{TSP}_\Pi(X_1,\ldots,X_n) \geqslant an^{1-\frac{1}{\gamma}}\big] \tag{A.35}$$

$$\text{and } p_2(n;A) := 1 - \mathbb{P}\big[\mathrm{TSP}_\Pi(X_1,\ldots,X_n) \leqslant An^{1-\frac{1}{\gamma}}\big] \tag{A.36}$$

Then $\mathrm{TSP}_\Pi(X_1,\ldots,X_n)$ is not $\Theta(n^{1-\frac{1}{\gamma}})$ if (and only if) either of the following:

i. for all $a > 0$, $p_1(n;a) \to 0$ as $n \to \infty$ slower than $c_1' e^{-c_2' n^{c_3'}}$ for any $c_1', c_2', c_3' > 0$ (or doesn't approach 0 at all).

ii. for all $A > 0$, $p_2(n;A) \to 0$ as $n \to \infty$ slower than $c_1' e^{-c_2' n^{c_3'}}$ for any $c_1', c_2', c_3' > 0$ (or doesn't approach 0 at all).

Fix $a > 0$; then if $p_1(n;a) \to 0$ as $n \to \infty$ slower than $c_1' e^{-c_2' n^{c_3'}}$ for any $c_1', c_2', c_3' > 0$ (or doesn't approach 0 at all), we note that by (A.27) that

$$\mathbb{P}[\mathrm{TSP}_\Pi(X_1,\ldots,X_N) < (a/2)n^{1-\frac{1}{\gamma}}] \geqslant p_1(n;a) - c_1 e^{-c_2(n/2)^{c_3}} \tag{A.37}$$

because if $N \geqslant n/2$ and $\mathrm{TSP}_\Pi(X_1, \ldots, X_{\lfloor n/2 \rfloor}) \geqslant a(n/2)^{1-\frac{1}{\gamma}}$ then

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_N) \geqslant a(n/2)^{1-\frac{1}{\gamma}} \geqslant (a/2)n^{1-\frac{1}{\gamma}} \tag{A.38}$$

and $N \geqslant n/2$ with probability at least $1 - c_1 e^{c_2 (n/2)^{c_3}}$. However, if $p_1(n; a) \to 0$ as $n \to \infty$ slower than $c_1' e^{-c_2' n^{c_3'}}$ for any $c_1', c_2', c_3' > 0$ (or doesn't approach 0 at all) then so does $p_1(n; a) - c_1 e^{-c_2 (n/2)^{c_3}}$. Thus, if (i) holds then $\mathrm{TSP}_\Pi(X_1, \ldots, X_N) \geqslant a' n^{1-\frac{1}{\gamma}}$ does not hold with very high probability for any a' (as we can use $a = 2a'$), and hence $\mathrm{TSP}_\Pi(X_1, \ldots, X_N)$ does not have order-of-growth $n^{1-\frac{1}{\gamma}}$ with very high probability.

For possibility (ii) we use an analogous argument based on (A.31) and the very high probability that $N \leqslant 2n$.

Thus, we have shown that if $\mathrm{TSP}_\Pi(X_1, \ldots, X_n)$ does not have order-of-growth $n^{1-\frac{1}{\gamma}}$ with very high probability, neither does $\mathrm{TSP}_\Pi(X_1, \ldots, X_N)$, so we have shown both the forward and backward implications and we are done. □

We can now show Proposition 23.

Proof of Proposition 23. Fix some f satisfying Assumption 5; in particular, there is some $f_{\max}$ such that $f(x) \leqslant f_{\max}$ for all $x \in \mathcal{X}_f$.

We now consider each f_j over $\mathcal{X}_j$. Since $f_j(x) \geqslant c$ for all $x \in \mathcal{X}_j$, letting $a = \lceil f_{\max}/c \rceil$ (by definition an integer), we have that $a\, f_j(x) \geqslant f(x)$ for all $x \in \mathcal{X}_f$ (which holds for all j). Then by assumption in Proposition 23 we have for all j: when $X_1^{(j)}, \ldots, X_n^{(j)} \overset{iid}{\sim} f_j$,

$$\mathrm{TSP}_\Pi(X_1^{(j)}, \ldots, X_n^{(j)}) \overset{wvhp}{=} O(n^{1-\frac{1}{\gamma}}) \tag{A.39}$$

Then by definition when $X_1^{(j)}, \ldots, X_{an}^{(j)} \overset{iid}{\sim} f_j$, since a is a constant,

$$\mathrm{TSP}_\Pi(X_1^{(j)}, \ldots, X_{an}^{(j)}) \overset{wvhp}{=} O((an)^{1-\frac{1}{\gamma}}) = O(n^{1-\frac{1}{\gamma}}) \tag{A.40}$$

By Lemma 40 this implies that if $\{X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}\} \sim \text{PoisProc}(na \cdot f_j)$,

$$\text{TSP}_\Pi(X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}) \overset{wvhp}{=} O(n^{1-\frac{1}{\gamma}}) \tag{A.41}$$

Since there are a fixed, finite number of $\mathcal{X}_j$ and f_j, and by Assumption 4 the time needed to travel between any two configurations in $[\mathcal{X}_f]^{\mathcal{Q}}$ is at most C, generating $\{X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}\} \sim \text{PoisProc}(na \cdot f_j)$ independently, we have

$$\text{TSP}_\Pi\left(\bigcup_{j=1}^{m} \{X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}\}\right) \leqslant \sum_{j=1}^{m} \text{TSP}_\Pi(X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}) + mC \tag{A.42}$$

$$= O(n^{1-\frac{1}{\gamma}}) \tag{A.43}$$

Finally, since $a \cdot f_j(x) \geqslant f(x)$ for all $x \in \mathcal{X}_j$, and $\mathcal{X}_f \subseteq \bigcup_{j=1}^{m} \mathcal{X}_j$, we get that

$$f(x) \leqslant \sum_{j=1}^{m} a \cdot f_j(x) \text{ for all } x \in \mathcal{X}_f\,. \tag{A.44}$$

Then let $f'(x) = \sum_{j=1}^{m} a \cdot f_j(x) - f(x)$ for all x, and let $\{X_1, \ldots, X_N\} \sim \text{PoisProc}(n \cdot f)$ and $\{X'_1, \ldots, X'_{N'}\} \sim \text{PoisProc}(n \cdot f')$. Then $\{X_1, \ldots, X_n, X'_1, \ldots, X'_{N'}\}$ has the same distribution as $\bigcup_{j=1}^{m} \{X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}\}$ by the properties of the Poisson process. But this means that

$$\text{TSP}_\Pi\left(\bigcup_{j=1}^{m} \{X_1^{(j)}, \ldots, X_{N^{(j)}}^{(j)}\}\right) = O(n^{1-\frac{1}{\gamma}}) \tag{A.45}$$

$$\text{TSP}_\Pi(X_1, \ldots, X_N) = O(n^{1-\frac{1}{\gamma}}) \tag{A.46}$$

which, by Lemma 40 then implies

$$\text{TSP}_\Pi(X_1, \ldots, X_n) = O(n^{1-\frac{1}{\gamma}})\,. \tag{A.47}$$

Since this holds for all f satisfying Assumption 5, we are done. □

A.2 The TSP, Complexity, and Approximation

As already discussed in the introduction, the TSP is an extremely important problem in a number of fields, with a rich history and enormous number of applications. In the original formulation, a traveling salesman needs to visit a number of cities, in any order; given the distances (or costs) of traveling between pairs of cities, modeled as a weighted graph, we want to compute how quickly (or cheaply) he can accomplish this. The question of computation complexity is: how quickly can this be computed? If it is too difficult to compute, can an approximate value be computed in a reasonable time instead?

To start, we must first clearly define the notions of complexity and approximation we are working with. While we assume the reader is familiar with the basic complexity theory notions discussed here, in particular the Nondeterministic Polynomial (NP) class of problems, we will briefly outline the relevant notions.

A.2.1 Complexity Basics

Here we cover some basic notions needed to discuss complexity and approximation.

Problems and Instances A computational problem is modeled as a general question which admits different *instances*. For example, the classic TSP is to find the shortest tour through all the nodes of a (connected) undirected graph, and any specific connected graph G describes an instance of the problem. The *complexity* of a problem is the relationship between the length of description (typically in bits) needed to specify the instance and the computation time and/or space required to determine the answer. We will generally denote an instance of a problem with the letter A.

Decision vs Optimization Problems A problem is a *decision problem* if the objective is to determine the truth or falsity of some proposition about any given instance A (a yes or no question): for example, in the classic SAT problem you are given a boolean formula (the instance) and asked to determine whether some setting of the variables would make the formula true. We contrast that with an *optimization*

problem, where the objective is to set problem-defined variables to minimize some objective function under constraints defined by A: for example, in the Linear Programming problem (standard form) the instance is a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and vectors $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^n$ (so $A = (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c})$), and you are asked to find the minimum value $\boldsymbol{c}^\top \boldsymbol{x}$ can take while $\boldsymbol{x} \in \mathbb{R}^n$ (the variables) satisfy $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ (if no $\boldsymbol{x}$ satisfies this then the value of the problem is regarded as infinite).

For a generic optimization problem, we denote the value of instance A as $\mathrm{Opt}(A)$. Note that the TSP is an optimization problem (where the "variables" are just a description of the tour, the constraint is to visit each target, and the objective function is the length of the tour); however, since its our problem of interest, we will denote the shortest tour length on instance A as $\mathrm{TSP}(A) = \mathrm{Opt}(A)$. We note that we only ask for the optimal value here, but often in practice it is also necessary to show how that optimal value is achieve (for instance, what to set the variables to).

Nonoptimal Algorithms for Optimization Problems Furthermore, in optimization problems one also has the notion of a *feasible* solution which may or may not be optimal, where the variables are set so the constraints are satisfied but the objective may not take its minimum possible value – e.g. in the TSP, any tour that visits all the targets is a feasible solution, even though it may be much longer than necessary – and we may have algorithms for an optimization problem which don't necessarily return the optimal value. For a generic algorithm ALG and instance A, we denote the value returned as $\mathrm{ALG}(A)$. Note that by definition, $\mathrm{ALG}(A) \geqslant \mathrm{Opt}(A)$; we say that ALG *provides an exact solution* to the problem if, for all instances A, it is guaranteed that $\mathrm{ALG}(A) = \mathrm{Opt}(A)$.

The reason we want to study algorithms which do *not* provide exact solutions to an optimization problem is that it might be the case that the best value $\mathrm{Opt}(A)$ is very time-consuming to compute, whereas $\mathrm{ALG}(A)$ for a well-constructed algorithm might be both a good approximation of $\mathrm{Opt}(A)$ and significantly easier to compute.

Remark 28. *Optimization problems can also have* infeasible *instances, where the constraints can never be satisfied (e.g. in the TSP on a disconnected graph G the agent*

cannot visit all the targets). In this work we assume all instances of our problems are feasible (for the DSTSP, Assumption 4 guarantees this).

Decision Versions of Optimization Problems While decision and optimization problems are fundamentally different, we can construct a decision version of an optimization problem as follows: a threshold τ can be added to an optimization problem to yield the decision problem "given (A, τ), is $\mathrm{Opt}(A) \leqslant \tau$?"; an algorithm for this in turn can (usually) be used to solve the original optimization problem via binary search, which is typically very efficient.

A.2.2 NP and Reductions

We now consider the well-known complexity class NP and the notion of reduction. In this (sub)section we deal only with decision problems, i.e. an instance A evaluates either to `true` or `false`, and it is our job to determine which it is. Given a problem L, we denote the truth-value of instance A under L as $\mathrm{VAL}_L(A)$, and we denote the size of instance A as $n(A)$ (usually determined by the number of bits needed to fully describe A).

A computational problem L is usually considered "tractable" or "easy" if $\mathrm{VAL}_L(A)$ can be evaluated in time polynomial in $n(A)$ ('time' being synonymous with the number of elementary computations needed); if an algorithm solves the problem in polynomial time we say it is *efficient*. The class of such problems is denoted P (for 'polynomial').

Remark 29. *'Tractable' problems may be exceptionally costly to compute for realistic instances – for instance, if* $\mathrm{VAL}_L(A)$ *can be computed in* $10^{1000}n(A)^{1000}$ *time, it is "tractable" even though for essentially any instance and any computer it would take eons to finish the computation.*

A computational problem L can be reduced to problem L' (in polynomial time) if there is a polynomial-time algorithm which, for any instance A of L produces an

instance A' of L' such that

$$\mathrm{VAL}_{L'}(A') = \mathrm{VAL}_L(A) \tag{A.48}$$

Note that this means if we have a black-box way of solving L', we can use it to solve instances of L as well. In general, a polynomial-time reduction from L to L' shows that solving L cannot be harder than solving L' (however, L' can still be harder to solve than L). Reductions can be chained together, i.e. combining a reduction from L to L' with a reduction from L' to L'' produces a reduction from L to L''.

The notion of (polynomial-time) reduction can then be used to define and analyze broader 'complexity classes' (of which P is an example). A problem L is *Hard* for a given class if all problems in that class can be reduced to L; it is *Complete* in that class if it both belongs in the class and is Hard.

The crucial class of *Nondeterministic Polynomial* (NP) problems consists of problems L which admit an efficiently-verifiable *certificate* for all A such that $\mathrm{VAL}_L(A) =$ `true`; that is, if $\mathrm{VAL}_L(A) =$ `true`, there is some object θ (for instance, if A is a SAT instance, θ may be the settings of the variables which satisfy A) such that given A and θ, it only takes polynomial time to verify that $\mathrm{VAL}_L(A) =$ `true`. Note that actually finding θ may be hard, and hence evaluating $\mathrm{VAL}_L(A)$ without being given θ may still be hard, and that if $\mathrm{VAL}_L(A) =$ `false` there may be no way to efficiently verify even with extra knowledge.

Problems that belong in NP commonly ask for the existence of some object, for instance the existence of a solution to a boolean formula: given a viable solution it is easy to check that it works, but there may be no easy way to check that an unsolvable boolean formula does not have a solution. The TSP decision problem is also in NP: if there exists a tour of length $\leqslant \tau$ on instance A, then it can be easily verified. It is also known that most variants of the TSP (such as the classic TSP on undirected graphs, or the Euclidean TSP limited to integer targets in 2 dimensions) and their decision variants are NP-Hard. In general we will mildly abuse the terminology and, given a variant of the TSP, write that it is "NP-Complete" or "NP-Hard" to mean its

decision variant is NP-Complete or NP-Hard.

A.2.3 Different Types of Approximation

Because the standard TSP is known to be NP-complete [23], even in extremely constrained settings (such as targets only on $\mathbb{Z}^2$ with Euclidean distances), works on TSP algorithms generally focus on approximation: given a TSP instance, the objective is to compute a path through all the targets which, while not necessarily the exact shortest path, approximates its length. However, there are a few different notions of approximation depending on the exact problem being discussed.

Definition 34. We define our notions of approximation as follows:

1. **Guaranteed Approximation:** The algorithm ALG achieves a *guaranteed approximation to factor* α if, for any instance A, it is guaranteed that $\mathrm{ALG}(A) \leqslant \alpha\,\mathrm{Opt}(A)$. α is usually a constant but can also be a function of $n(A)$, e.g. a log-factor approximation algorithm satisfies $\mathrm{ALG}(A) \leqslant c\log(n(A))\mathrm{Opt}(A)$ for some constant $c > 0$.

2. **Approximation in Expectation:** For some random procedure of generating A (e.g. uniform random A of complexity $n(A) = n$), the algorithm ALG achieves *approximation in expectation to factor* α if

$$\mathbb{E}[\mathrm{ALG}(A)] \leqslant \alpha\mathbb{E}[\mathrm{Opt}(A)] \tag{A.49}$$

3. **Approximation with Very High Probability:** For some random procedure of generating A, the algorithm ALG achieves *approximation with very high probability to factor* α if there are $c_1, c_2, c_3 > 0$ such that

$$\mathbb{P}[\mathrm{ALG}(A) \leqslant \alpha\,\mathrm{Opt}(A)] \geqslant 1 - c_1 e^{-c_2 n(A)^{c_3}} \tag{A.50}$$

4. **Adversarial Approximation:** The algorithm ALG achieves *adversarial*

approximation to factor α if

$$\sup_{A \in \mathcal{A}_n} \mathrm{ALG}(A) \leqslant \alpha \sup_{A \in \mathcal{A}_n} \mathrm{Opt}(A) \tag{A.51}$$

where $\mathcal{A}_n = \{A : n(A) \leqslant n\}$.

Here $\alpha \geqslant 1$ by definition (or else ALG would be better than Opt, which is by definition impossible).

Guaranteed approximation is the most common notion of approximation algorithm (it is sometimes referred to as *worst-case approximation* to distinguish it from *average-case*, but we will not use that term here to avoid confusion with adversarial approximation), and it is the most powerful in the sense that if it holds, it implies the other three. Guaranteed approximation algorithms for the classic (discrete, non-dynamic) TSP are well-studied, in particular the question of the smallest factor α achievable: the well-known Christofides approximation algorithm [24] achieves $\alpha = 3/2$, and more advanced algorithms can reduce it further. However, as noted in Section 1.1 these algorithms do not apply to the Dynamic TSP due to the lack of a well-defined notion of 'distance' between target points.

Our main DSTSP results (Theorem 1) show that our algorithms achieve approximation with very high probability (type 3 from Definition 34 above), which in turn shows that they also achieve approximation in expectation (type 2 from Definition 34 above, and see Corollary 1). A significant open problem remains over whether there is an algorithm capable of achieving guaranteed approximation (type 1 from Definition 34 above) for the general Dynamic TSP, as there is for the TSP on graphs. Finally, Theorem 2 shows that in the symmetric case our Symmetric Hierarchical Collection Problem algorithm achieves adversarial approximation (type 4 from Definition 34 above, see Chapter 6).

We remark that we do not originate these notions: in particular, previous works on the Dynamic TSP or Euclidean TSP have used notions 2-4 [13, 14, 16], while 1 is the standard notion of approximation algorithms (and is generally the one used when discussing the TSP non-dynamic settings).

Appendix B

Nonsymmetric Ball-Box Conjecture

In this appendix we study the fundamental differences between symmetric (or driftless) and nonsymmetric (or drifting) control-affine dynamics. In particular, we consider how to formulate and prove an analogue to the Uniform Ball-Box Theorem ([19], Theorem 2.4) which holds for nonsymmetric dynamics. The Uniform Ball-Box Theorem shows how to inscribe a box in and circumscribe a box around the reachable set (or *ball*) of a given set of dynamics at any sufficiently small scale ε; it is often used to show that even at very small scales the vehicle can move in any direction through clever use of commutators of its control vector fields. While this property doesn't hold for nonsymmetric vehicles (typically the drift will move the vehicle forward faster than it can produce many other motions), for our purposees we want to use it to do two things:

- Show that the reachable set has a certain volume, which governs the maneuverability of the vehicle at small scales, and consequently yields the lower bound as proved in Chapter 3.

- Allow the inscribed boxes to be used to produce recursive tilings (the Hierarchical Cell Structures used in Chapter 4 and Chapter 5).

The goal is to extend this to nonsymmetric dynamics.

Recall that our dynamics are control-affine (Definition 2) if there are smooth vector

fields $h_0, h_1, \ldots, h_m$ on $\mathcal{Q}$ such that the control law can be expressed as

$$\dot{\boldsymbol{q}} = h_0(\boldsymbol{q}) + \sum_{j=1}^{m} h_j(\boldsymbol{q}) u_j \tag{B.1}$$

for control $\boldsymbol{u} = (u_1, \ldots, j_m)$ in some $\mathcal{U} \subseteq \mathbb{R}^m$, and that $\mathcal{U}$ satisfies the following properties:

- $\mathcal{U}$ contains $\mathbf{0} = (0, \ldots, 0)$ in its interior (and is thus full-dimensional in $\mathbb{R}^m$);
- $\mathcal{U}$ is convex;
- $\mathcal{U}$ is bounded and closed;
- $\mathcal{U}$ is radially symmetric, i.e. $\boldsymbol{u} \in \mathcal{U} \iff -\boldsymbol{u} \in \mathcal{U}$.

We note that the last of these conditions implies that if h_0 is zero everywhere (i.e. the dynamics are *driftless*) then any motion can be reversed by negating the inputs, making the dynamics symmetric. Thus, in this chapter, we use *driftless* as a synonym for symmetric, and *drifting* as a synonym for nonsymmetric.

Remark 30. *As the Nonsymmetric Uniform Ball-Box Conjecture remains unproven, this section is intended as an exploration and discussion of the topic. Some results are stated and proved as potential partial steps to a future proof.*

B.1 Notation

Because this section deals in detail with a topic which is largely absorbed into the assumptions or only dealt with at a high level in the rest of this work, and because we don't require certain concepts (notably the target distribution f or the number of targets n) we use different notation intended to clarify the distinction between vector fields and functions over manifolds. Specifically:

- we let n denote the dimension of $\mathcal{Q}$;

- we use capital letters such as V or W to represent vector fields, generally of the control vector fields (denoted h_j elsewhere) or their Lie brackets;

- f to denote functions over the manifold (either to the real numbers or other spaces).

A control-affine system Π is specified as a sequence of control vector fields followed by control set, where the first vector field is the drift i.e. $\Pi = (V_0, V_1, \ldots, V_m; \mathcal{U})$. Note that if we want to symmetrize a control-affine system $\Pi = (V_0, V_1, \ldots, V_m; \mathcal{U})$, we do so by adding a control to the drift term, which in this notation means adding a zero vector field as the new drift and adding a dimension to the control set, which we define as

$$\Pi = (V_0, V_1, \ldots, V_m; \mathcal{U}) \implies \overleftrightarrow{\Pi} := (\mathbf{0}, V_0, V_1, \ldots, V_m; [-1, 1] \times \mathcal{U}) \,. \tag{B.2}$$

We also refer to the reachable sets as $R_\Pi(\boldsymbol{q}, \varepsilon)$ where Π are the dynamics, $\boldsymbol{q}$ is the center, and ε is the radius.

All notation in this section is specified within.

B.2 Lie Brackets

We begin with a description of Lie brackets and some of their important properties. This section deals with definitions from driftless systems – even though the drift vector field V_0 is present it will not be treated as distinct.

Let V be a smooth vector field. Then for any f (either a function to real numbers or another vector field or a flow, etc), we can differentiate f along V. We denote this as Vf (or $f \to Vf$ being the transformation).

Given a vector field V, we denote its *flow* (from $\boldsymbol{q} \in \mathcal{Q}$) by $\varphi_V^t(\boldsymbol{q}) : \mathbb{R} \to \mathcal{Q}$; this denotes the point in $\mathcal{Q}$ reached by integrating along V for t time, i.e. it satisfies

$$\frac{d}{dt}\varphi_V^t(\boldsymbol{q}) = V(\varphi_V^t(\boldsymbol{q})) \tag{B.3}$$

at all t and $\boldsymbol{q}$, and $\varphi_V^0(\boldsymbol{q}) = \boldsymbol{q}$ for all $\boldsymbol{q} \in \mathcal{Q}$.

Definition 35. For smooth vector fields V, W on $\mathcal{Q}$, their *Lie bracket* $[V, W]$ is a vector field on $\mathcal{Q}$ given by either of the following (equivalent) definitions:

(a) $[V, W]$ is the unique vector field satisfying $[V, W]f = VWf - WVf$ for all f.

(b) For any $\boldsymbol{q} \in \mathcal{Q}$,

$$[V, W](\boldsymbol{q}) = \frac{d}{dt}\Big|_{t=0_+} \left(\varphi_V^{\sqrt{t}} \circ \varphi_W^{\sqrt{t}} \circ \varphi_V^{-\sqrt{t}} \circ \varphi_W^{-\sqrt{t}}(\boldsymbol{q})\right) \tag{B.4}$$

Proposition 24. Lie brackets satisfy the following properties:

(a) Linearity in both arguments: for (finite) sets of vector fields $\{V_i\}$, $\{W_j\}$ and coefficients $\{a_i\}$, $\{b_j\}$,

$$\Big[\sum_i a_i V_i, \sum_j b_j W_j\Big] = \sum_{i,j} a_i b_j [V_i, W_j] \tag{B.5}$$

(b) Skew symmetry: $[V, W] = -[W, V]$ for all vector fields V, W.

(c) Jacobi identity: $[V_1, [V_2, V_3]] + [V_2, [V_3, V_1]] + [V_3, [V_1, V_2]] = 0$

Since Lie brackets are themselves vector fields, we can take Lie brackets of Lie brackets, and so forth:

Definition 36. Given a set of vector fields $\mathcal{V}^1 = \{V_0, V_1, \ldots, V_m\}$ the set of order-s iterated Lie brackets is defined inductively by

$$\mathcal{V}^{s+1} := \mathcal{V}^s \cup \{[W, V_i] : W \in \mathcal{V}^s, V_i \in \mathcal{V}^1\} \tag{B.6}$$

Since $\mathcal{V}^s$ at $\boldsymbol{q}$ is a set of vectors in $T_{\boldsymbol{q}}\mathcal{Q}$, we can define $\triangle^s(\boldsymbol{q}) = \mathrm{span}(\mathcal{V}^s(\boldsymbol{q})) \subseteq T_{\boldsymbol{q}}\mathcal{Q}$. *Chow's condition*[3] (also known as the *Hörmander condition*) holds if for all

$\boldsymbol{q} \in \mathcal{Q}$, the *degree of nonholonomy at* $\boldsymbol{q}$

$$r(\boldsymbol{q}) := \min\{s : \triangle^s(\boldsymbol{q}) = T_{\boldsymbol{q}}\mathcal{Q}\} < \infty \tag{B.7}$$

i.e. if the iterated Lie brackets span the whole tangent space.[a] The *Lie algebra* of $\{V_0, V_1, \ldots, V_m\}$ is

$$\text{Lie}(V_0, V_1, \ldots, V_m) := \bigcup_{j \geqslant 1} \triangle^s \tag{B.8}$$

and is smallest linear subspace of $VF(\mathcal{Q})$ (the space of smooth vector fields on $\mathcal{Q}$) containing $V_0, V_1, \ldots, V_m$ and invariant under Lie bracketing.

We also define $n_s(\boldsymbol{q}) = \dim(\triangle^s)$ (where $\dim(\triangle^0(\boldsymbol{q})) = 0$ at all $\boldsymbol{q}$ by convention), and the vector $\boldsymbol{n}(\boldsymbol{q}) = (n_1(\boldsymbol{q}), \ldots, n_{r(\boldsymbol{q})}(\boldsymbol{q}))$ is called the *growth vector at* $\boldsymbol{q}$. $\boldsymbol{q}$ is *regular* if it has a neighborhood in which $\boldsymbol{n}$ is constant, and control system Π is *equiregular* if $\mathcal{Q}$ is connected and all $\boldsymbol{q}$ are regular. This implies that $\boldsymbol{n}$ is constant over $\mathcal{Q}$, and in particular that the degree of nonholonomy $r := r(\boldsymbol{q})$ is the same everywhere.

[a] Given that V_0 represents the drift term, the *strong Hörmander condition* holds if the iterated Lie brackets of $\{V_1, \ldots, V_m\}$ span $T_{\boldsymbol{q}}\mathcal{Q}$. While used in various works concerning drifting systems[**?**], many important systems (for instance the Dubins car) do not satisfy it.

While all the definitions above depend on $\boldsymbol{q}$, for simplicity (and because for equiregular systems each n_s is a constant) when a $\boldsymbol{q}$ is fixed we will usually drop it and write e.g. $\triangle^s := \triangle^s(\boldsymbol{q})$.

Each iterated Lie bracket can be written out in similar fashion to Definition 35(b): the derivative at $t = 0$ of a sequence of movements (either forward or backward) along the component vector fields. We will need this representation in order to discuss the behavior of Lie brackets under drift.

B.3 Privileged Coordinates and Ball-Box Theorem

While the Ball-Box Theorem can be intuitively stated as relating to the inclusion of sub-Riemannian balls inside boxes and vice versa, formalizing it is tricky. One immediate problem is defining a rigorous notion of a 'box' in a manifold. This is done, within a neighborhood of a point of interest $\boldsymbol{q} \in \mathcal{Q}$, by setting up a coordinate map: a 'box' is then the pre-image of a Euclidean box under the coordinate map. However, not every coordinate map will have the needed properties. Instead, *privileged coordinates* are generally used:

Definition 37. Given a driftless equiregular control system[a]

$$\Pi^* = (\mathbf{0}, V_0, V_1, \dots V_m; [-1,1]^{m+1}) \tag{B.9}$$

define the weights w_j for $j = 1, 2, \dots, n$ as

$$w_j = s \iff n_{s-1} < j \leqslant n_s \tag{B.10}$$

i.e. the number of weights of value s is the number of dimensions added to the span when iterating Lie brackets from $\mathcal{V}^{s-1}$ to $\mathcal{V}^s$. We also define the *order* of a function $g : \mathcal{Q} \to \mathbb{R}$ at $\boldsymbol{q} \in \mathcal{Q}$ (under Π) as

$$\operatorname{ord}_{\boldsymbol{q}}(g) := \sup\left(\theta \in \mathbb{R} : g(\boldsymbol{q}') = O(d_{\Pi^*}(\boldsymbol{q}, \boldsymbol{q}')^\theta)\right) \tag{B.11}$$

Given $\boldsymbol{q} \in \mathcal{Q}$, a *system of privileged coordinates* at $\boldsymbol{q}$ is a local coordinate map $\boldsymbol{z} : \mathcal{Q} \to \mathbb{R}^n$ (where $\boldsymbol{z}(\boldsymbol{q}') = (z_1(\boldsymbol{q}'), \dots, z_n(\boldsymbol{q}'))$) in a neighborhood of $\boldsymbol{q}$ such that

$$\operatorname{ord}_{\boldsymbol{q}}(z_j) = w_j \text{ for all } j \tag{B.12}$$

We also define the *scale-ε box* at $\boldsymbol{q}$ to be

$$P^{\varepsilon}_{\boldsymbol{q}} := \boldsymbol{z}^{-1}\Big(\prod_{j=1}^{n}[-\varepsilon,\varepsilon]\Big) \subseteq \mathcal{Q} \tag{B.13}$$

To be precise, $\prod_{j=1}^{n}[-\varepsilon,\varepsilon] \subseteq \mathbb{R}^n$ is the Euclidean box parallel to the privileged coordinates, centered at $\mathbf{0}$ and of width 2ε, and $P^{\varepsilon}_{\boldsymbol{q}}$ is the preimage of this box. Because of the nature of privileged coordinates, the preimage will have size $\Theta(\varepsilon^{w_j})$ in the direction of z_j.

[a] We use this 'symmetrized' version of Π so we can employ $V_0, V_1, \ldots, V_m$ in the usual manner, and so $d_{\Pi *}$ is a metric.

Since $\mathcal{Q}$ is a manifold, as $\varepsilon \to 0$, the volume of $cP^{\varepsilon}_{\boldsymbol{q}}$ in $\mathcal{Q}$ approaches $(2\varepsilon)^{\gamma}$ where $\gamma := \sum_{j=1}^{n} w_j$, i.e. the volume of the Euclidean box representing it. The privileged coordinates obey an important rule:

Proposition 25. For a system of privileged coordinates $(z_1, \ldots, z_n)$ at $\boldsymbol{q}$,

$$dz_j(\triangle^{w_j}(\boldsymbol{q})) \neq 0 \;\text{ and }\; dz_j(\triangle^{w_j-1}(\boldsymbol{q})) = 0 \;\text{ for } j = 1, \ldots, n \tag{B.14}$$

This is referred to as the coordinates being *linearly adapted*.

We now re-state the Ball-Box Theorem (for driftless systems) in the notation of this appendix:

Uniform Ball-Box Theorem. Let $\Pi = \Pi(\mathbf{0}, V_0, V_1, \ldots, V_m; \mathcal{U})$ be a driftless control-affine system satisfying Meta-Assumption 1, where $\{V_0, V_1, \ldots, V_m\}$ are equiregular. Then is a set of privileged coordinates and constants $0 < c < C$ and $\varepsilon^* > 0$ such that for all $0 < \varepsilon < \varepsilon^*$,

$$P^{c\varepsilon}_{\boldsymbol{q}} \subseteq R_{\Pi}(\boldsymbol{q}, \varepsilon) \subseteq P^{C\varepsilon}_{\boldsymbol{q}} \tag{B.15}$$

The change from $[-1,1]^{m+1}$ to $\mathcal{U}$ is not consequential, since $\delta[-1,1]^{m+1} \subseteq \mathcal{U} \subseteq \Delta[-1,1]^{m+1}$ for some constants $0 < \delta < \Delta$; therefore, d_{Π} and $d_{\Pi *}$ (from Definition 37)

differ by no more than a Δ/δ constant factor, so the growth vector is unaffected and the constants c, C change by at most a Δ/δ factor.

As will be discussed later in greater depth, it is important to note that the proper driftless 'equivalent' to a drifting control system $\Pi = (V_0, V_1, \ldots, V_m; \mathcal{U})$ (i.e. the driftless system with which it shares important properties like size of the reachable set) is not $(\mathbf{0}, V_1, \ldots, V_m; \mathcal{U})$ but rather

$$\overleftrightarrow{\Pi} := (\mathbf{0}, V_0, V_1, \ldots, V_m; [-1, 1] \times \mathcal{U}) \tag{B.16}$$

Instead of removing the drift term (which would prevent a Dubins car from even moving, for example), we add a control to it.

Finally, we state a slightly more general version of the Nonsymmetric Uniform Ball-Box Conjecture in the notation of this appendix:

Conjecture 1. Let $\Pi = \Pi(V_0, V_1, \ldots, V_m; \mathcal{U})$ be a control-affine system where $V_0, \ldots, V_m$ satisfy Meta-Assumption 1. Then there is set of privileged coordinates such that for any $0 < \lambda < 1$ there are constants $0 < c_\lambda < C_\lambda$ such that for every sufficiently small $\varepsilon > 0$,

$$P^{c_\lambda \varepsilon}_{\varphi_0^{\lambda\varepsilon}(\boldsymbol{q})} \subseteq R_\Pi(\boldsymbol{q}, \varepsilon) \subseteq P^{C_\lambda \varepsilon}_{\varphi_0^{\lambda\varepsilon}(\boldsymbol{q})} \tag{B.17}$$

This differs from the original because it permits the center to be at any $\lambda\varepsilon$ down the flow of V_0 from $\boldsymbol{q}$

B.4 Lie Brackets Under Drift

In order to extend the Unigotm Ball-Box Theorem to control-affine dynamics with drift, we extend the notion of Lie brackets; this extension is intended as an exploration of how the Nonsymmetric Uniform Ball-Box Conjecture may be proved in the future. There are two main challenges to doing so: (i) Lie brackets are commonly thought of as the limit to a commutator of V_i, V_j, and therefore are vectors in $T_{\boldsymbol{q}}\mathcal{Q}$, but under the influence of drift the end point will be more the result of the drift along V_0 than

the motions of the commutator, which must then be somehow subtracted out; (ii) since many systems we are interested in require the drift vector field V_0 in order for the Lie brackets to span the space, we want to include V_0 in the Lie brackets, despite the system's inability to directly control movement along it.

Definition 38. Under the coordinate chart as defined above, given smooth vector fields V_i, V_j, the *drifting Lie bracket* (which drifts along V_0) is

$$[V_i, V_j]_* = \frac{d}{dt}\Big|_{t=0}\left(\varphi_{0+i}^{\sqrt{t}} \circ \varphi_{0+j}^{\sqrt{t}} \circ \varphi_{0-i}^{\sqrt{t}} \circ \varphi_{0-j}^{\sqrt{t}} \circ \varphi_{-0}^{4\sqrt{t}}\right) \tag{B.18}$$

This corresponds to the usual definition of the Lie bracket, but each movement is accompanied by drift (as it must be); to compensate, we pull the starting point back by the corresponding amount along the drift, so that it gives a comparison of the motion with a comparison of doing nothing and following the drift alone.

Remark 31. *As mentioned, we want to be able to use V_0 itself as a term in the drifting Lie brackets, and to do this we must be able to write Definition 38 as a series of valid movements (except for the initial pullback term). Note that every term is of the form φ_{0+i}^{τ} or φ_{0-i}^{τ}, for an appropriate i, τ. If $i = 0$, then φ_{0+0}^{τ} (moving twice as fast along V_0) and φ_{0-0}^{τ} (staying stationary) are not valid movements – but we can rewrite $\varphi_{0+0}^{\tau} = \varphi_0^{2\tau}$, and φ_{0-0}^{τ} can be removed entirely since it is the identity function. Thus, even with V_0 as a term, the drifting Lie bracket corresponds to a sequence of valid movements.*

We now consider how the drifting Lie bracket compares to normal Lie brackets:

Lemma 41. $[V_i, V_j]_* = [V_i, V_j] + 2[V_i, V_0] - 2[V_0, V_j]$ *for all* i, j.

Proof. We prove this using Definition 35(a), i.e. that for every smooth f and V, W,

$$[V, W]f = VWf - WVf \tag{B.19}$$

We define a function

$$\beta(t) := \varphi_{-0}^{4t} \circ \varphi_{0+i}^{t} \circ \varphi_{0+j}^{t} \circ \varphi_{0-i}^{t} \circ \varphi_{0-j}^{t}(\boldsymbol{q}) \tag{B.20}$$

We then define $\alpha(t) = \beta(\sqrt{t})$, so that

$$[V_i, V_j]_* f(\boldsymbol{q}) = \lim_{t \to 0} \frac{1}{t}\big(f(\alpha(t)) - f(\boldsymbol{q})\big) \tag{B.21}$$

Now we let $t = s^2$, so we can re-write this limit as $s \to 0$:

$$[V_i, V_j]_* f(\boldsymbol{q}) = \lim_{s \to 0} \frac{1}{s^2}\big(f(\beta(s)) - f(\boldsymbol{q})\big) \tag{B.22}$$

We now apply L'Hopital's rule, taking the first derivative of both $f(\beta(s)) - f(\boldsymbol{q})$ on top and s^2 on the bottom (w.r.t. s), getting

$$[V_i, V_j]_* f(\boldsymbol{q}) = \lim_{s \to 0} \frac{1}{2s} \frac{d}{d\tau}\Big|_{\tau = s} f(\beta(\tau)) = \lim_{s \to 0} \frac{(f \circ \beta)'(s)}{2s} \tag{B.23}$$

(switching to Newton notation for derivatives, for clarity).

We now want to show that $\lim_{s \to 0}(f \circ \beta)'(s)$ goes to $\mathbf{0}$ (the zero vector in the coordinate chart) thus allowing us to apply L'Hopital's rule again. This happens because of the chain rule, where

$$\lim_{s \to 0}(f \circ \beta)'(s) = \lim_{s \to 0} \beta'(s) f = \beta'(0) f \tag{B.24}$$

(the derivative of f along $\beta'(0)$). $\beta'(0)$ can also be analyzed with the chain rule. To simplify, we define

$$H(x_1, x_2, x_3, x_4, x_5) = \varphi_{-0}^{x_1} \circ \varphi_{0+i}^{x_2} \circ \varphi_{0+j}^{x_3} \circ \varphi_{0-i}^{x_4} \circ \varphi_{0-j}^{x_5}(\boldsymbol{q}) \tag{B.25}$$

Then we can re-write (where $x_1 = 4t$ and $x_2 = x_3 = x_4 = x_5 = t$)

$$\beta'(0) = \frac{d}{dt}\Big|_{t=0} H(x_1, x_2, x_3, x_4, x_5) \tag{B.26}$$

$$= \sum_{k=1}^{5} \frac{\partial H}{\partial x_k} \frac{dx_k}{dt} \tag{B.27}$$

$$= -4V_0 + (V_0 + V_i) + (V_0 + V_j) + (V_0 - V_i) + (V_0 - V_j) = \mathbf{0} \tag{B.28}$$

Thus we can apply L'Hopital again to get

$$[V_i, V_j]_* f(\boldsymbol{q}) = \lim_{s \to 0} \frac{(f \circ \beta)'(s)}{2s} = \frac{(f \circ \beta)''(0)}{2} \tag{B.29}$$

Using the function H from above, we define $G(x_1, x_2, x_3, x_4, x_5) = (f \circ H)(x_1, x_2, x_3, x_4, x_5)$, and let $g(t) = G(4t, t, t, t, t)$ so that $(f \circ \beta)''(t) = g''(t)$. Then:

$$g''(0) = \sum_{k,\ell=1}^{5} \frac{\partial x_k}{\partial t}\Big|_{t=0} \frac{\partial x_\ell}{\partial t}\Big|_{t=0} \frac{\partial^2 G}{\partial x_k \partial x_\ell}\Big|_{x_1,x_2,x_3,x_4,x_5=0} \tag{B.30}$$

Note that at $x_1 = x_2 = x_3 = x_4 = x_5 = 0$, the function G is just $f(\boldsymbol{q})$ (and all the flow components are the identity function). Thus, taking a partial derivative of G at this point is the same as taking the derivative of the corresponding flow term, and so

$$\sum_{k,\ell=1}^{5} \frac{\partial x_k}{\partial t}\Big|_{t=0} \frac{\partial x_\ell}{\partial t}\Big|_{t=0} \frac{\partial^2 G}{\partial x_k \partial x_\ell}\Big|_{x_1,x_2,x_3,x_4,x_5=0} \tag{B.31}$$

$$= 16(-V_0)(-V_0)f + 4(-V_0)(V_0 + V_i)f + 4(-V_0)(V_0 + V_j)f \tag{B.32}$$

$$+ 4(-V_0)(V_0 - V_i)f + 4(-V_0)(V_0 - V_j)f \tag{B.33}$$

$$+ 4(-V_0)(V_0 + V_i)f + (V_0 + V_i)(V_0 + V_i)f + (V_0 + V_i)(V_0 + V_j)f \tag{B.34}$$

$$+ (V_0 + V_i)(V_0 - V_i)f + (V_0 + V_i)(V_0 - V_j)f \tag{B.35}$$

$$+ 4(-V_0)(V_0 + V_j)f + (V_0 + V_i)(V_0 + V_j)f + (V_0 + V_j)(V_0 + V_j)f \tag{B.36}$$

$$+ (V_0 + V_j)(V_0 - V_i)f + (V_0 + V_j)(V_0 - V_j)f \tag{B.37}$$

$$+ 4(-V_0)(V_0 - V_i)f + (V_0 + V_i)(V_0 - V_i)f + (V_0 + V_j)(V_0 - V_i)f \tag{B.38}$$

$$+ (V_0 - V_i)(V_0 - V_i)f + (V_0 - V_i)(V_0 - V_j)f \tag{B.39}$$

$$+ 4(-V_0)(V_0 - V_j)f + (V_0 + V_i)(V_0 - V_j)f + (V_0 + V_j)(V_0 - V_j)f \tag{B.40}$$

$$+ (V_0 - V_i)(V_0 - V_j)f + (V_0 - V_j)(V_0 - V_j)f \tag{B.41}$$

$$= -4V_0V_if + 4V_iV_0f - 4V_0V_jf + 4V_0V_jf + 2V_iV_jf - 2V_jV_if \tag{B.42}$$

The above happens because while deriving along vector fields is not commutative, it is distributive. The derivation is long and complicated but one can visualize it as

follows: consider the sequence

$$-4V_0,\ (V_0+V_i),\ (V_0+V_j),\ (V_0-V_i),\ (V_0-V_j). \tag{B.43}$$

Call these terms (in order) W_1, W_2, W_3, W_4, W_5; then

$$\frac{\partial x_k}{\partial t}\Big|_{t=0}\frac{\partial x_\ell}{\partial t}\Big|_{t=0}\frac{\partial^2 G}{\partial x_k \partial x_\ell}\Big|_{x_1,x_2,x_3,x_4,x_5=0} = \begin{cases} W_k W_\ell f & \text{if } k \leqslant \ell \\ W_\ell W_k f & \text{if } \ell < k \end{cases} \tag{B.44}$$

Note that k, ℓ are interchangeable (as we would expect) and that the the order that W_k, W_ℓ appear in is determined by their order in the sequence. We will consider the nine possible terms YZf for $Y, Z \in \{V_0, V_i, V_j\}$ separately:

- V_0V_0f has 16 from W_1W_1; another 16 from W_kW_ℓ and $W_\ell W_k$ (depending on whether $k \leqslant \ell$) for $k, \ell \in \{2,3,4,5\}$; and -32 from W_1W_k and W_1W_ℓ (both apply) for $k, \ell \in \{2,3,4,5\}$. Thus, the V_0V_0f term cancels completely.

- V_iV_if has 2 from $k = \ell = 2$ and $k = \ell = 4$, but has -2 from $k = 2, \ell = 4$ and $k = 4, \ell = 2$, and therefore also cancels completely.

- V_jV_jf is analogous to V_iV_if, except for $k, \ell \in \{3,5\}$ instead of $\{2,4\}$, and cancels completely.

- V_0V_if: by symmetry we can look at the result for $k < \ell$, double it (to account for $\ell < k$) and add the result for $k = \ell$. For $k < \ell$, when $k = 1$, the terms for $\ell = 2, 4$ cancel each other; for $k = 4, 5$, there is no $\ell > k$ which produces any V_0V_if terms; and for $k = 2, 3$, only $\ell = 4$ produces the correct term, and we get $-V_0V_if$ both times. Thus (doubling to account for $k > \ell$ as well) we get $-4V_0V_if$. For $k = \ell$, only $k = \ell = 2$ and $k = \ell = 4$ contain V_0V_if, and they cancel out (we get V_0V_if from $k = \ell = 2$, and $-V_0V_if$ from $k = \ell = 4$). Thus the overall result is $-4V_0V_if$.

- V_0V_jf: this is analogous to V_0V_if (but with $k, \ell \in \{3,5\}$ being the terms of interest rather than $k, \ell = \{2,4\}$) and we get $-4V_0V_jf$.

- V_iV_0f: this is analogous to V_0V_if, but with $k = 2$ and $\ell = 3, 4$ (and vice versa) being the outstanding terms, giving $4V_iV_0f$.

- V_jV_0f: this is analogous to V_iV_0f, but with $k, \ell \in \{3, 5\}$ being the terms of interest (rather than $k, \ell \in \{2, 4\}$), giving $4V_jV_0f$.

- V_iV_jf: we split it up again into $k < \ell$ (and double to also capture $k > \ell$) and $k = \ell$. Since V_i, V_j never appear in the same W_k, the $k = \ell$ case doesn't produce any V_iV_jf terms. For $k < \ell$: when $k = 1, 3, 5$, there is no ℓ which produces a V_iV_jf term; when $k = 2$, the term cancel between $\ell = 4, 5$; and when $k = 4$, we get $(-V_i)(-V_j)f = V_iV_jf$. Thus (doubling to account for $\ell < k$) we get $2V_iV_jf$.

- V_jV_if: this is analogous to V_iV_jf except that the term left over is $k = 3, \ell = 4$ and we get $V_j(-V_i)f = -V_jV_if$; doubling (for $k = 4, \ell = 3$) we get a final result of $-2V_jV_if$.

Thus, we know that

$$[V_i, V_j]_*f(\boldsymbol{q}) = g''(0)/2 \tag{B.45}$$

$$= \big(- 2V_0V_if + 2V_iV_0f - 2V_0V_jf + 2V_0V_jf + V_iV_jf - V_jV_if\big)(\boldsymbol{q}) \tag{B.46}$$

Now we employ Definition 35(a); we know that $[V_i, V_0]f = V_iV_0f - V_0V_if$ and $[V_0, V_j]f = V_0V_jf - V_jV_0f$ for all f. Thus, we have

$$[V_i, V_j]_*f = \big([V_i, V_j] + 2[V_i, V_0] - 2[V_0, V_j]\big)f \tag{B.47}$$

for all f, and we are done. □

Note the corollary when $V_i = V_0$ and/or $V_j = V_0$ (since $[V_0, V_0] = 0$):

Corollary 3. For any V_i, we have

$$[V_0, V_i]_* = -[V_0, V_i] \quad \text{and} \quad [V_i, V_0]_* = 3[V_i, V_0] \tag{B.48}$$

We note also that Lemma 41 means that

$$\operatorname{Span}(\{V_i\}_i \cup \{[V_i, V_j]\}_{i,j}) = \operatorname{Span}(\{V_i\}_i \cup \{[V_i, V_j]_*\}_{i,j}) \tag{B.49}$$

i.e. the set of all V_i plus all length-2 Lie brackets spans the same space (at any point) as the set of all V_i plus all length-2 drifting Lie brackets, which in other words is that the degree of nonholonomy is $\leqslant 2$. In that case, we have our result:

Proposition 26. Let $\Pi = \Pi(V_0, V_1, \ldots, V_m; \mathcal{U})$ be a control-affine system where $V_0, \ldots, V_m$ satisfy Meta-Assumption 1 and such that the degree of nonholonomy is $\leqslant 2$. Then there is set of privileged coordinates such that for any $0 < \lambda < 1$ there are constants $0 < c_\lambda < C_\lambda$ such that for every sufficiently small $\varepsilon > 0$,

$$P^{c_\lambda \varepsilon}_{\varphi_0^{\lambda\varepsilon}(\boldsymbol{q})} \subseteq R_\varepsilon(\boldsymbol{q}) \subseteq P^{C_\lambda \varepsilon}_{\varphi_0^{\lambda\varepsilon}(\boldsymbol{q})} \tag{B.50}$$

Extending this to the full Nonsymmetric Uniform Ball-Box Conjecture might be done by analyzing iterated drifting Lie brackets and showing that they result in the same spans as the Lie brackets (though there is a combinatorial explosion in the different terms which need to be canceled out).

Appendix C

Fractal Targets

While γ is an integer for control-affine systems and continuous target point distributions, it is interesting theoretically to consider the case of noninteger γ. This can occur when the targets are distributed over a set of fractal dimension. Since this appendix is meant to illustrate how our techniques (particularly the Symmetric Hierarchical Cell Structures) can be applied in different settings, we will make some simplifying assumptions.

First, we limit ourselves to targets distributed in strictly self-similar fractal sets in $\mathbb{R}^d$ and a single scaling factor (such as the Sierpinski triangle, Menger sponge, or Koch curve) and the Euclidean TSP with varying top speed. Note that in this case $\mathcal{X} = \mathcal{Q} = \mathbb{R}^d$, so there is no distinction between $\boldsymbol{q} \in \mathcal{Q}$ and $x \in \mathcal{X}$.[1] We will therefore refer to configurations as $x \in \mathbb{R}^d$, and the reachable sets $R_\varepsilon(x)$ are the same for both workspace and configuration space.

Formally, we assume that $\mathcal{X}_f$ (the set on which the targets are distributed) is composed of $\hat{b}$ copies of itself at s^{-1} scale (translated and/or rotated and/or reflected), which makes imposing a Hierarchical Cell Structure with Euclidean dynamics trivial. Using γ as revealed through the HCS, we find that $\gamma = \log_s \hat{b}$ is the Hausdorff dimension of $\mathcal{X}_f$; we assume that $\gamma \geqslant 1$ (otherwise the $O(1)$ time to travel between HCS's in the HCS cover will overwhelm the $O(n^{1-\frac{1}{\gamma}})$ tour time from the hierarchical collec-

[1] This means that standard TSP approximation algorithms such as Christofides can be applied, but our analysis is still needed to determine the tour length as $n \to \infty$.

tion problem). Additionally, choosing a random $X \in \mathcal{X}_f$ can be done by iteratively choosing sub-copies.

We also limit ourselves to considering the Euclidean TSP over such distributions, with a top speed function h, i.e. our control law is

$$\dot{x} = h(x)\boldsymbol{u} \tag{C.1}$$

where the control set is $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\|_2 \leqslant 1\}$. We assume h is Lipschitz-continuous and $h_{\min} := \inf_x h(x) > 0$ and $h_{\max} := \sup_x h(x) < \infty$. Formally, the Lipschitz continuity of h allows us to get a ζ-approximate HCS cover for any $\zeta > 0$ (locally h is approximately constant so we can just use nesting cells at a small scale ε_0).

Instead of 'density f' and 'agility g', we let

$$\phi(x) = \lim_{\varepsilon \to 0} \frac{\mathbb{P}_{X \sim f}[X \in \bar{R}_\varepsilon(x)]}{\varepsilon^\gamma} \tag{C.2}$$

(where $X \sim f$ means the given distribution over the fractal, even though f is no longer a probability density function). Then we replace the integral $\int f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx$ with its equivalent $\mathbb{E}_{X \sim f}[\phi(x)^{-\frac{1}{\gamma}}]$.

Our upper bound can still tell us about this case:

Proposition 27. For noninteger γ, we get the following bounds: if $\gamma \geqslant 3$ is noninteger, then bounds (1.35), (1.36), and (1.38) from Proposition 3 still hold. If $\gamma \in (2, 3)$, then for all sufficiently large n,

$$\mathbb{P}\big[\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big) n^{1-\frac{1}{\gamma}} \mathbb{E}_{X \sim f}[\phi(x)^{-\frac{1}{\gamma}}]\big] \tag{C.3}$$

$$\leqslant 1 - e^{-\frac{(\gamma-2)p_1 n (\sum_{j=1}^m p_j^{1-1/\gamma})^2}{13\gamma}} \tag{C.4}$$

And, finally, if $\gamma \in (1,2)$, then for all sufficiently large n,

$$\mathbb{P}\big[\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\mathbb{E}_{X\sim f}[\phi(x)^{-\frac{1}{\gamma}}]\big] \tag{C.5}$$

$$\leqslant 1 - e^{-\frac{p_1 n^{2(1-1/\gamma)}(\sum_{j=1}^m p_j^{1-1/\gamma})^2}{\frac{560}{3}\log(\gamma/(\gamma-1))+4+18(1-1/\gamma)^2\frac{\gamma}{\gamma-2}(\frac{280}{3}\log(\gamma/(\gamma-1)))^{1-2/\gamma}}} \tag{C.6}$$

Proof. This is proved in the same manner as the upper bounds in Chapter 4 using the HCS cover and the collection problem. □

Appendix D

Good Target, Bad Target

In this appendix we show that even though our Symmetric HCS cover (see Chapter 4) might not be entirely overlap-free, since it can be have arbitrarily small overlap parameter ρ (see Lemma 19), we can derive the main results in the same way as in Chapter 4. Recall that we showed that the probability that $X \sim f$ falls into multiple HCS's can be made arbitrarily small, i.e. less than any $\rho > 0$. Additionally, in Lemma 28 we made a distinction of cells in a HCS cover between ρ_1-*good* and *not* ρ_1-*good* cells, and showed that for any $\rho_1, \rho_2 > 0$ we could find a sufficiently fine scale so that the probability of a target point not falling into a ρ_1-good cell is at most ρ_2.

This allows us to define two kinds of target points X_i: those that fall into a single ρ_1-good cell, and those that don't (either by falling into a not-ρ_1-good cell, or by falling into multiple cells) which we respectively call *good* and *bad* target points. Each target point has at most a $\rho^* \leqslant \rho + \rho_2$ chance of being bad, where ρ^* can be made arbitrarily small, and they are independent. We now show that we can simplify to ignore the bad ones. We first fix some HCS cover of $\mathcal{X}_f$ at scale $\varepsilon^{(\text{bad})}$ with $m^{(\text{bad})}$ different HCS's, as in Lemma 18; this is allowed to overlap to any amount, but note that it is fixed independent of ρ^*.

Let $n^{(\text{bad})}$ be the number of bad target points, and $n_j^{(\text{bad})}$ be the number of bad target points in the jth cell of the fixed HCS cover with $m^{(\text{bad})}$ cells at scale $\varepsilon^{(\text{bad})}$.

Note by Proposition 13 and the concavity of $x^{1-\frac{1}{\gamma}}$ that this means that

$$\sum_{j=1}^{m^{(\text{bad})}} \text{HCP}^*(n^{(\text{bad})}; \hat{b}, s) \leqslant \sum_{j=1}^{m^{(\text{bad})}} 6s(n^{(\text{bad})}/m^{(\text{bad})})^{1-\frac{1}{\gamma}} = 6s(m^{(\text{bad})})^{\frac{1}{\gamma}}(n^{(\text{bad})})^{1-\frac{1}{\gamma}} \quad \text{(D.1)}$$

Thus, letting $\boldsymbol{q}_1^{(\text{bad})}, \ldots, \boldsymbol{q}_{m^{(\text{bad})}}^{(\text{bad})}$ be the roots of this HCS cover and letting

$$C^{(\text{bad})} := \text{TSP}_\Pi(\boldsymbol{q}_1^{(\text{bad})}, \ldots, \boldsymbol{q}_{m^{(\text{bad})}}^{(\text{bad})}) \,. \quad \text{(D.2)}$$

As before we don't necessarily need to know the exact value of this; what's important is that it is fixed with respect to n.

Thus, we know that

$$\text{TSP}_\Pi(\{X_i : X_i \text{ is bad}\}) \leqslant C^{(\text{bad})} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}}\varepsilon^{(\text{bad})}(n^{(\text{bad})})^{1-\frac{1}{\gamma}} \,. \quad \text{(D.3)}$$

This is a guaranteed bound, not a probabilistic one. We can then note that $n^{(\text{bad})}$ is itself a binomial random variable of n draws of $\leqslant \rho^*$ probability of success each. Therefore, with very high probability (for any fixed $\rho^* > 0$) we have $n^{(\text{bad})} \leqslant 2\rho^* n$, which implies

$$\text{TSP}_\Pi(\{X_i : X_i \text{ is bad}\}) \leqslant C^{(\text{bad})} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}}\varepsilon^{(\text{bad})}2^{1-\frac{1}{\gamma}}(\rho^*)^{1-\frac{1}{\gamma}}n^{1-\frac{1}{\gamma}} \quad \text{(D.4)}$$

By Lemma 28, for any $\delta^* > 0$, we can then fix ρ^* sufficiently small so that for sufficiently large n,

$$C^{(\text{bad})}/n^{1-\frac{1}{\gamma}} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}}\varepsilon^{(\text{bad})}2^{1-\frac{1}{\gamma}}(\rho^*)^{1-\frac{1}{\gamma}} \leqslant \delta^* \quad \text{(D.5)}$$

i.e. while the above scales according to $n^{1-\frac{1}{\gamma}}$ we can make the constant term as small as we need to so that it is negligible compared to the overall constant as derived in the main text.

Appendix E

Counterexamples

We now delve into some non-theorems – things that sound like they might be true, but in fact are not. The most basic point of describing these (other than the fact that some of them are really quite interesting) is to show why certain assumptions we make are necessary – without them, the theorems break down.

E.1 When $\mathcal{Q}_f$ is too expansive and the agent too fast

A key assumption for our lower-bound proof for the TSP (or, more accurately, for the upper-bound proof of the Orienteering problem) is Assumption 9, that $\mathcal{Q}_f$ can be covered by a polynomial number of reachable sets. Another key assumption is Assumption 3: that the agent cannot move arbitrarily quickly through the workspace. We ask: are these assumptions necessary? Our technique seems to rely mainly on the volume of the reachable sets, which can have the properties we want (for instance, bounded agility g at all configurations) even when $\mathcal{Q}_f$ allows for arbitrarily large speeds (as the following example shows).

We show a counterexample for the case where Assumption 4 and Assumption 3 don't hold but all the other examples still hold:

Example 2. *Consider the space $\mathcal{X} = \mathcal{T}^2$, where $\mathcal{T}^2$ is the 2-dimensional torus. We represent this as $[0, 1)^2$ with opposing edges joined together, with two axes being (as usual) the x-axis and y-axis. We introduce a vehicle traveling through $\mathcal{X}$ whose x-axis*

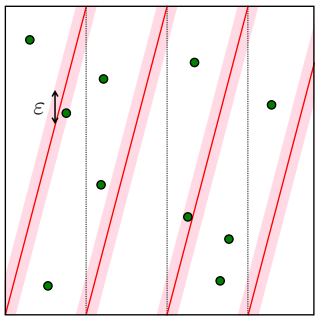

Figure E-1: Illustration of Example 2, with a steep 'track' from which we can deviate and visit targets. Because we can select from an infinite range of slopes, we can find one that comes arbitrarily close to every target point, allowing us to collect them all in a fixed amount of time (not dependent on n).

position always increases at speed 1 *(except for wrapping around), while its y-axis position is controlled through double integration, i.e. it controls its acceleration on the y-axis.*

Thus, its current configuration $\boldsymbol{q}(t)$ *at time* t *includes an x-axis position* $x(t)$*, a y-axis position* $y(t)$*, and a y-axis velocity* $v(t)$*; note that since its x-velocity is always* 1 *we have no need to record it. Therefore,* $\mathcal{Q} = \mathcal{T}^2 \times \mathbb{R}$*. The dynamics it obeys are:*

$$\dot{\boldsymbol{q}} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} 1 \\ v \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u_1 \quad \textit{with control } u_1 \in [-1, 1] \tag{E.1}$$

The vehicle is allowed to start at any $\boldsymbol{q}_0 \in \mathcal{Q}$*.*

Unlike the double integrator which is generally studied, we do not place any limits on the velocity. For the following, we use $\langle z \rangle := z - \lfloor z \rfloor$, i.e. $\langle z \rangle$ is the fractional part of z.

First, we need the following lemma, to show that γ and g satisfy Assumption 6:

Lemma 42. *For this vehicle,* $\gamma = 3$*, and* $g(\boldsymbol{q}) = 1/3$ *for all* $\boldsymbol{q} \in \mathcal{Q}$*.*

Proof. We consider the shape of the small-time reachable set. For simplicity we 'unroll' the workspace $\mathcal{T}^2$ into $\mathbb{R}^2$ where the mapping from $(x^*, y^*) \in \mathbb{R}^2$ to $(x, y) \in \mathcal{T}^2$

is $(x, y) = (\langle x^* \rangle, \langle y^* \rangle)$; as we will be primarily concerned with the height of intervals along the y-axis, working in $\mathbb{R}^2$ presents no problems unless the height is larger than 1, which we will prevent by working on short time scales. Because the dynamics are translation-invariant over the (toroidal) workspace, we may also assume WLOG that we are at $(x(0), y(0), v(0)) = (0, 0, v_0) \in \mathcal{Q}$ for some $v_0 \in \mathbb{R}$. In that case, at time $t = \delta$ (with the assumption that $\delta < 1/10$) we must be at $(\delta, y(\delta), v(\delta))$: what are the minimum and maximum values of $y(\delta)$ achievable? The extremes of this interval are if we input $u(t) = 1$ or $u(t) = -1$ for all $t \in [0, \delta]$. Let us consider $u(t) = 1$. Then, at time $t = t^* + \delta$:

$$y = v_0\delta + \delta^2/2 \tag{E.2}$$

and similarly, if we have input -1 at all times, $y = v_0\delta - \delta^2/2$. It is obvious that all points in between the extremes are reachable as well (just by setting $u(t)$ to an appropriate value in $[-1, 1]$), so the interval has size δ^2. So now we want to calculate the size of the set reachable from configuration $q = (x^*, y^*, m)$ in time *at most* ε. This can then be computed as:

$$\mathrm{Vol}_{\mathcal{T}^2}(R_\varepsilon(q)) = \int_{\delta=0}^{\varepsilon} \delta^2 d\delta = \frac{\varepsilon^3}{3} \tag{E.3}$$

which by definition means that $\gamma = 3$ and $g((0, 0, v_0)) = 1/3$ for all v_0 (and, by translation invariance, for all $\boldsymbol{q} \in \mathcal{Q}$). □

Thus, the unbounded velocity does not cause e.g. agility g to become unbounded. The reason that this example violates the lower bound that our results would have provided is that, because Assumption 9 is violated, the number of discretizations of valid paths is infinite. Not only does this break the proof given in Chapter 3, we can show that it causes there to always be TSP tour of length < 1:

Proposition 28. For this vehicle, if the target points X_i are distributed unifomly

at random over $[0,1]^2$:

$$\mathrm{TSP}_{\Pi}(X_1,\ldots,X_n) \leqslant 1 \text{ almost surely (i.e. with probability 1)} \tag{E.4}$$

The proof of this relies on the following lemma:

Lemma 43. *If* $(x_1,y_1),\ldots,(x_n,y_n) \overset{iid}{\sim} \mathrm{Unif}_{[0,1]^2}$ *and* $\varepsilon > 0$*, then with probability* 1 *there exists* $v_0 \in \mathbb{R}$ *such that*

$$\langle\, |y_i - v_0 x_i| \,\rangle \leqslant \varepsilon \text{ for all } i \in [n] \tag{E.5}$$

Proof. We first define, for all $v \in \mathbb{R}$ and $\varepsilon > 0$, the set

$$S_{v,\varepsilon} := \{(x,y) : \langle\, |y - vx| \,\rangle \leqslant \varepsilon\} \tag{E.6}$$

Note that our lemma states that no matter how small $\varepsilon > 0$ is, there is (with probability 1) some v such that $(x_i,y_i) \in S_{v,\varepsilon}$ for all i. For what follows, since ε is fixed, we will denote $S_v := S_{v,\varepsilon}$.

To prove this, we need a particular claim. In what follows, $\mathrm{Vol}(\cdot)$ refers to the 2-dimensional area:

Claim: For any union of finitely many polygons $U \subseteq [0,1]^2$ with sides parallel to the axes,

$$\lim_{v\to\infty} \mathrm{Vol}(S_v \cap U) = 2\varepsilon\,\mathrm{Vol}(U) \tag{E.7}$$

We first prove the claim for the special case that U is a rectangle with sides parallel to the axes. Let U be bounded by $x_{\min}, x_{\max}$ on the x-axis and by $y_{\min}, y_{\max}$ on the y-axis. First, we consider the set $S_v(y') := S_v \cap \{(x,y') : x \in [x_{\min}, x_{\max}]\}$, i.e. the intersection of S_v with a horizontal line at $y = y'$. We note that this intersection consists of regularly-spaced intervals of width $2\varepsilon/v$, spaced (center-to-center) $1/v$ apart; we let $a_v(y')$ be the (1-dimensional) length of $S_v(y')$. We note that since the centers of the line segments are spaced $1/v$ apart, accounting for rounding at the edges of U, there should be between $(x_{\max} - x_{\min})v - 2$ and $(x_{\max} - x_{\min})v + 2$ centers

of the intervals passing through the segment $[x_{\min}, x_{\max}] \times y'$; thus

$$2\varepsilon(x_{\max} - x_{\min}) - 2\varepsilon/v \leqslant a_v(y') \leqslant 2\varepsilon(x_{\max} - x_{\min}) + 2\varepsilon/v$$

and as $v \to \infty$ this means that $a_v(y') \to 2\varepsilon(x_{\max} - x_{\min})$ uniformly over all y.

Since S_v and U are both measurable in area (both are unions of finitely many polygons), we get that

$$\lim_{v\to\infty} \operatorname{Vol}(S_v \cap U) = \lim_{v\to\infty} \int_{y_{\min}}^{y_{\max}} a_v(y')dy' = 2\varepsilon(x_{\max} - x_{\min})(y_{\max} - y_{\min}) = 2\varepsilon \operatorname{Vol}(U)$$

For U which is not a rectangle with sides parallel to the axes, we simply decompose U into (finitely many, disjoint) polygons, then approximate each polygon by a finite number of rectangles parallel to the axes (which can be done to arbitrary position by increasing the number of rectangles), and apply Claim 1; the limit applies uniformly over all the rectangles simply because there are a finite number of them.

Now that we have established the claim, we can prove the lemma. We now generate a sequence $v_1, v_2, \cdots \in \mathbb{R}$ with a special property. For any fixed sequence $\boldsymbol{v} = (v_1, v_2, \dots)$ let $\boldsymbol{v}^{(k)} := (v_1, \dots, v_k)$, and for any length-$k$ bitstring $\boldsymbol{b} \in \{0,1\}^k$ define the following:

$$S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b}) := \bigcap_{j=1}^{k} S_{v_j}(b_j)$$

where we define $S_v(0) := S_v^c$ (the complement) and $S_v(1) := S_v$; essentially, $\boldsymbol{b}$ is a bitstring telling us whether we need to be in S_{v_j} (if $b_j = 1$) or in its complement (if $b_j = 0$). We note that the set $\{S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b})\}_{\boldsymbol{b}\in\{0,1\}^k}$ of 2^k subsets of $[0,1]^2$ is a partition of $[0,1]^2$, and we note that no matter what $v_1, \dots, v_k$ are, $S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b})$ is a union of finitely many polygons for any $\boldsymbol{b}$.

We now consider the following procedure:

1. $v_1 = 1$;

2. for any $k \geqslant 1$, we iteratively choose v_{k+1} sufficiently large so that

$$\mathrm{Vol}(S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b}) \cap S_{v_{k+1}}) \geqslant \varepsilon \,\mathrm{Vol}(S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b})) \tag{E.8}$$

for all $\boldsymbol{b} \in \{0,1\}^k$.

By the claim we showed before, since there are finitely many $\boldsymbol{b} \in \{0,1\}^k$, the property is satisfied by any sufficiently large v_{k+1} (since the limiting volume is twice the volume we require).

Now we consider a random target $(x, y) \sim \mathrm{Unif}([0,1]^2)$; for any $\boldsymbol{b} \in \{0,1\}^k$ and any k,

$$\mathbb{P}[(x,y) \in S_{v_{k+1}} \,|\, (x,y) \in S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b})] = \frac{\mathrm{Vol}(S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b}) \cap S_{v_{k+1}})}{\mathrm{Vol}(S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b}))} \geqslant \varepsilon \tag{E.9}$$

(we don't need to worry about zero-probability issues since there is always a positive probability, and in any case we need the above only for $\boldsymbol{b}$ where $\mathrm{Vol}(\mathrm{Vol}(S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b})) > 0$). Considering n independent targets $(x_1, y_1), \ldots, (x_n, y_n) \overset{iid}{\sim} \mathrm{Unif}([0,1]^2)$, we get that

$$\mathbb{P}[(x_i, y_i) \in S_{v_{k+1}} \text{ for all } i \,|\, (x_i, y_i) \in S_{\boldsymbol{v}^{(k)}}(\boldsymbol{b}_i)] \geqslant \varepsilon^n \tag{E.10}$$

for any set of length-k bitstrings $\boldsymbol{b}_i$. Therefore, taking the inverse of the above statement (the probability that at least one (x_i, y_i) is not in S_{v_k}) and considering all k from 1 to ℓ, we get

$$\mathbb{P}[\forall\, k \in [\ell],\, \exists\, i \in [n] \text{ such that } (x_i, y_i) \notin S_{v_k}] \leqslant (1 - \varepsilon^n)^\ell \tag{E.11}$$

But since $1 - \varepsilon^n < 1$, we get

$$\begin{aligned}
&\mathbb{P}[\forall\, k \in \mathbb{Z}_{\geqslant 0},\, \exists\, i \in [n] \text{ such that } (x_i, y_i) \notin S_{v_k}] && \text{(E.12)}\\
&\qquad = \lim_{\ell \to \infty} \mathbb{P}[\forall\, k \in [\ell],\, \exists\, i \in [n] \text{ such that } (x_i, y_i) \notin S_{v_k}] && \text{(E.13)}\\
&\qquad \leqslant \lim_{\ell \to \infty} (1 - \varepsilon^n)^\ell = 0 && \text{(E.14)}
\end{aligned}$$

Thus, we know that the probability that the set $\{(x_i, y_i)\}_{i=1}^n$ of n random target points escaping *all* S_{v_k} is 0, thus showing our lemma. □

Armed with Lemma 43 we can now prove Proposition 28.

Proof. We denote the coordinates of target X_i as (x_i, y_i). Assume, without loss of generality, that the the points X_i are ordered by increasing x-values; with probability 1, all of them have unique (and nonzero) x-values. We then let $\delta := \min_i(x_i - x_{i-1})$ (where x_0 is defined to be 0) for all $i = 1, 2, \ldots, n$, and consider a control policy which does the following over $t \in [0, \delta]$, for some $a \in [-1, 1]$:

$$u(t) = \begin{cases} a & \text{if } t \in [0, (1/4)\delta) \\ -a & \text{if } t \in [(1/4)\delta, (3/4)\delta) \\ a & \text{if } t \in [(3/4)\delta, \delta] \end{cases} \tag{E.15}$$

Letting $q(0) = (x_0, y_0, v_0))$, and we will use the trick from before of 'unrolling' $\mathcal{T}^2$, so our trajectory will be considered on $\mathbb{R}^2$. We are particularly interested in $\boldsymbol{q}((1/2)\delta)$ and $\boldsymbol{q}(\delta)$. We note that $v(t)$ is the following (easy to calculate since $\dot{v}(t) = u(t)$ and $v(0) = v_0$):

$$v(t) = \begin{cases} v_0 + at & \text{if } t \in [0, (1/4)\delta) \\ v_0 + a(\delta/2) - at & \text{if } t \in [(1/4)\delta, (3/4)\delta) \\ v_0 - a\delta + at & \text{if } t \in [(3/4)\delta, \delta] \end{cases} \tag{E.16}$$

There are several things of note: (i) in all cases, the v_0 appears by itself in $v(t)$; (ii) $v((1/4)\delta + s) = v((1/4)\delta - s)$ when $s \leqslant (1/4)\delta$; and (iii) when $t \leqslant (1/2)\delta$, then $v(t + (1/2)\delta) = -v(t)$.

Since $y(t) = y_0 + \int_0^t v(s)ds$, we see from (i) that we can separate out the v_0 term, i.e. if we define $w(t) := v(t) - v_0$ we find that $w(t)$ has no dependence at all on v_0 so

we can write

$$y(t) = y_0 + v_0 t + \int_0^t w(s)\, ds \tag{E.17}$$

Furthermore, by (ii) we know that

$$\int_0^{(1/2)\delta} w(s)\, ds = 2\int_0^{(1/4)\delta} w(s)\, ds = 2\int_0^{(1/4)\delta} as\, ds = a(1/16)(\delta^2) \tag{E.18}$$

Finally, we get by (iii) that $\int_0^\delta w(s)ds = 0$ since the parts on $[0, (1/2)\delta)$ and $[(1/2)\delta, \delta]$ cancel out, and that $w(\delta) = 0$ (from looking at $v(\delta)$). Therefore:

$$y((1/2)\delta) = y_0 + (\delta/2)v_0 + a(\delta^2/16) \text{ and } y(\delta) = y_0 + \delta v_0 \text{ and } v(\delta) = v_0 \tag{E.19}$$

Note that its configuration at $t = \delta$, i.e. $(x_0 + \delta, y_0 + \delta v_0, v_0)$, is exactly the configuration it would have had with a zero control, i.e. $u(t) = 0$ for $t \in [0, \delta]$; but its y-position at $t = (1/2)\delta$ deviated by $a(\delta^2/16)$ (where $a \in [-1, 1]$).

Using this knowledge, we let $\varepsilon = \delta^2/16$ and find $v^{(\varepsilon)} \in \mathbb{R}$ as given in Lemma 43. For each $i \in [n]$, let $x_i' = x_i - \delta/2$ and $x_i'' = x_i + \delta/2$ (except for x_n'', which we don't need). Since $\delta = \min_i(x_i - x_{i-1})$ where x_0 is defined as 0, we know that

$$0 \leqslant x_1' \leqslant x_1 \leqslant x_1'' \leqslant x_2' \leqslant x_2 \leqslant x_2'' \leqslant \cdots \leqslant x_n' \leqslant x_n \leqslant 1 \tag{E.20}$$

(since we aren't using x_n''). We again want to work in the "unrolled" space, which essentially is $\mathbb{R}^2$ tiled with 1×1 squares, each with its own copies of the targets. To do this, we need to define

$$m^* = \arg\min_{m\in\mathbb{Z}}(|y_i + m^* - v^{(\varepsilon)x_i}|) \text{ and } y_i^* := y_i + m^* \tag{E.21}$$

for all $i \in [n]$: this basically identifies the "copy" of y_i in the unrolled space closest to the line $v^{(\varepsilon)}x_i$. Finally, we define $a_i := (y_i^* - v^{(\varepsilon)}x_i)/\varepsilon$, and note that by the definition of $v^{(\varepsilon)}$, we have $|a_i| \leqslant 1$. We now build a trajectory starting at $\boldsymbol{q}(0) = (0, 0, v^{(\varepsilon)})$

with the following control (recalling that $x(t) = t$ is guaranteed since $x(0) = 0$) for $t \in [0, 1]$:

$$u_1(t) = \begin{cases} u_i^*(t - x_i') & \text{if } t \in [x_i', x_i''] \\ 0 & \text{otherwise} \end{cases} \tag{E.22}$$

where $u_1^{(i)} : [0, \delta] \to \mathbb{R}$ is

$$u_1^{(i)}(t) = \begin{cases} a_i & \text{if } t \in [0, (1/4)\delta) \\ -a_i & \text{if } t \in [(1/4)\delta, (3/4)\delta) \\ a_i & \text{if } t \in [(3/4)\delta, \delta] \end{cases} \tag{E.23}$$

i.e. the control analyzed earlier, shifted to the δ-length interval $[x_i', x_i'']$ (in the case of $i = n$, we cut the trajectory off at 1 if it happens that $1 < x_n''$). What does this control do?

We already saw that from $t = 0$, sub-control $u_1^{(i)}(t)$ produces the same state at $t = \delta$ as a zero control would, but that at $t = \delta/2$ it deviates from the zero-control path by $a_i(\delta^2/16) = a_i\varepsilon = y_i^* - v^{(\varepsilon)}x_i$. Thus (since outside of the sub-control areas it just sets $u_1^{(i)} = 0$) the trajectory satisfies $\boldsymbol{q}(t) = (t, v^{(\varepsilon)}t, v^{\varepsilon})$ (in the unrolled space) for $t = x_i'$ and x_i'' for all i, but at $t = x_i$ (the midpoint of $[x_i', x_i'']$) it satisfies

$$\boldsymbol{q}(x_i) = \begin{bmatrix} x_i \\ v^{(\varepsilon)}x_i + y_i^* - v^{(\varepsilon)}x_i \\ v^{(\varepsilon)} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i^* \\ v^{(\varepsilon)} \end{bmatrix} \tag{E.24}$$

However, we note that $\langle y_i^* \rangle = y_i$, i.e. being at (x_i, y_i^*) in the unrolled space equates to being at (x_i, y_i) in $\mathcal{T}^2$, so the trajectory successfully collects target X_i. But we now have a trajectory that collects all n targets within $0 \leqslant t \leqslant 1$, thus showing the result. □

Hence, even though $\gamma = 3$ and there is no issue of unbounded agility, we have

still managed to show that the law of $\mathrm{TSP}(X_1, \ldots, X_n) = n^{1-\frac{1}{\gamma}}$ does *not* hold in this case, since $\mathrm{TSP}(X_1, \ldots, X_n) \leqslant 1$ (almost surely) regardless of how large n gets. The reason this happens is because the vehicle is allowed to select from an unbounded set of starting configurations: hence our ORNT upper bound proof fails because we cannot cover our starting configuration set $\mathcal{Q}_f$ with polynomially many (or any finite number for that matter) reachable sets; the counterexample shows that the boundedness of $\mathcal{Q}_f$ is not only necessary for the proof but for the result.

However, if we bound $\mathcal{Q}_f$ by insisting that the starting configuration have a slope $v \in [-M, M]$, for fixed M, this construction no longer works because the appropriate slope $v^{(\varepsilon)}$ will soon fall far outside of the allowed interval, and if we wanted to achieve such a slope, we would have to spend a lot of time just accelerating. In such a scenario, the algorithms given in this work are much more efficient than accelerating up to the "perfect" constant-length trajectory, because the acceleration process itself will be too protracted.