Agility and Target Distribution in the Dynamic Stochastic Traveling Salesman Problem

Aviv Adler[1], Oren Gal[2], and Sertac Karaman[3]

[1,3]*Laboratory for Information & Decision Systems, MIT*

[2]*Technion, Israel Institute of Technology*

[1,3]*{adlera,sertac}@mit.edu,* [2]*orengal@alumni.technion.ac.il*

Abstract

An important variant of the classic Traveling Salesman Problem (TSP) is the Dynamic TSP, in which a system with dynamic constraints is tasked with visiting a set of n target locations (in any order) in the shortest amount of time. Such tasks arise naturally in many robotic motion planning problems, particularly in exploration, surveillance and reconnaissance, and classical TSP algorithms on graphs are typically inapplicable in this setting. An important question about such problems is: if the target points are random, what is the length of the tour (either in expectation or as a concentration bound) as n grows? This problem is the Dynamic Stochastic TSP (DSTSP), and has been studied both for specific important vehicle models and for general dynamic systems; however, in general only the order of growth is known. In this work, we explore the connection between the distribution from which the targets are drawn and the dynamics of the system, yielding a more precise lower bound on tour length as well as a matching upper bound for the case of symmetric (or *driftless*) systems. We then extend the symmetric dynamics results to the case when the points are selected by a (non-random) adversary whose goal is to maximize the length, thus showing worst-case bounds on the tour length.

I. INTRODUCTION AND MOTIVATION

A. Introduction, motivation, and previous work

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 are subject to non-trivial differential constraints which have a substantial impact on the optimal tour. Furthermore, the addition of dynamic constraints means there is no fixed notion of 'distance' between 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]. Furthermore, 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 this tour length grows as the number n of targets is increased. We refer to this as the *Stochastic TSP*; when combined with dynamic constraints, it is the *Dynamic Stochastic TSP* (DSTSP).

The importance of the Stochastic TSP has not gone unnoticed [5]. The general asymptotic behavior of the Euclidean Stochastic TSP was characterized by Beardwood et al. in 1959 [6]; later, it was studied extensively from an algorithmic perspective, in particular on discovering algorithms which on average achieve high-quality approximations in polynomial time [7]. In addition, it was found to have connections to a variety of practical and theoretical problems, such as vehicle routing [8] and matching on the Euclidean plane [9]. In the second half of the

2000's, tight (up to a constant factor) probabilistic asymptotic bounds on the stochastic TSP 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) [10], [11], [12], [13], [14]. The stochastic TSP for a symmetric dynamical system in 2 dimensions was considered by Itani et al. [15], using a technique based on subadditive functionals, which was then expanded in Itani's Ph.D thesis [16] to include a wide class of translation-invariant dynamics in any Euclidean space $\mathbb{R}^d$. Finally, in 2016, we used an alternative elementary approach based on a discretization technique developed by Arias-Castro et al. [17] for the related Stochastic Orienteering problem to find the asymptotic order-of-growth as $n \to \infty$ for a wide class of dynamics on manifolds of any dimension [18]. In all of these works, algorithms were given which achieved (with high probability as the number of targets grows) a tour length within a constant factor of the optimal with respect to n.

B. Contribution

While previous work on this problem strongly characterizes the asymptotic behavior of the length of the DSTSP tour as the number n of targets increases - in particular, the dynamics induce a parameter, called the *small-time constraint factor* [16], [18] and denoted in this work as γ, such that the length of the tour is with high probability $\Theta(n^{1-\frac{1}{\gamma}})$ – the constant factor associated with this growth was generally not studied closely, and its relationship to the target point distribution was not generally known.

Furthermore, previous work generally assumed that the distribution of target points is uniform over a unit square or rectilinear region. While this does not affect the growth rate, as any full-dimensional continuous probability distribution with bounded support can be divided into finitely many approximately square regions, which can then be handled in sequence, changing the target point distribution may greatly affect the constant factor attached to the $\Theta(n^{1-\frac{1}{\gamma}})$ DSTSP tour length growth rate.

In this work, we give the first high-probability DSTSP bounds with explicit constants that hold over a wide class of dynamic constraints and target point probability distributions. Specifically, we give an explicit high-probability lower bound for a broad class of dynamics, and a matching upper bound (to a factor constant in both the number n and density function f of the target points) for the special case of *symmetric* (or *driftless*) dynamics, which admit a simplified algorithm. To get

these bounds, for any dynamics we define the *agility function* it induces over the workspace, and show how the agility function and the probability density function of the target point distribution affect the constant factor of the DSTSP tour length.

C. Paper organization

The rest of this work is organized as follows. In Section II we formally define the problem and the terms used in the results. Because our results require several assumptions we state the results first, in Section III, before we give the assumptions in Section IV; many of the assumptions are also shown to hold in general when the dynamic constraints are *control-affine*. We then show our probabilistic lower bound for the TSP tour length in Section V, where we first set up a version of the Stochastic Orienteering problem and show an important preliminary result which we then use to show the TSP lower bound, and our probabilistic upper bound in Section VI. Finally, we extend these results to the case where the targets are placed to maximize the tour length in Section VII.

II. PRELIMINARIES AND PROBLEM DEFINITION

In this work we study the Dynamic Stochastic TSP, in which a dynamic system 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 dynamic systems 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).

In what follows, 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'.

A. The dynamic system

Although we are primarily concerned about the system'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 system 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 $[q]_{\mathcal{X}} = x \in \mathcal{X}$; [1] we denote the preimage of $x \in \mathcal{X}$ as $[x]^{\mathcal{Q}} = \{q \in \mathcal{Q} : [q]_{\mathcal{X}} = x\}$. 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 q, x such that $[q]_{\mathcal{X}} = x$, there are some neighborhoods around 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 IV-F and Appendix D) but we assume it for what follows.

Given a position $q(t) \in \mathcal{Q}$ at time t and a control input u from some control set $\mathcal{U}$, the system evolves according to a *control law* $\dot{q}(t) = h^{\text{cont}}(q(t), u)$, where h^{cont} takes a position in $\mathcal{Q}$ and a control and returns a tangent vector indicating the direction in which the system moves [3]. We call a trajectory $\pi : [0, T] \to \mathcal{Q}$ *valid* if there is some control function $u : [0, T] \to \mathcal{U}$ such that $\dot{\pi}(t) = h^{\text{cont}}(\pi(t), u(t))$ 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 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:

$$\bar{\pi}(t) = [\pi(t)]_{\mathcal{X}} \tag{1}$$

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 "$q \in \pi$" if there exists t such that $q = \pi(t)$.

---

[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.

B. Symmetric dynamics

We say that dynamics Π defined by control law h^{cont} and control set $\mathcal{U}$ are *symmetric* if for every $\boldsymbol{q} \in \mathcal{Q}, \boldsymbol{u} \in \mathcal{U}$ there is a $\boldsymbol{u}' \in \mathcal{U}$ such that

$$h^{\mathrm{cont}}(\boldsymbol{q}, \boldsymbol{u}') = -h^{\mathrm{cont}}(\boldsymbol{q}, \boldsymbol{u}) \tag{2}$$

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

$$\overleftarrow{\pi}(t) := \pi(\ell(\pi) - t) \tag{3}$$

is also valid. This allows the system to maneuver indefinitely in arbitrarily small regions without having to leave and come back, and the DSTSP algorithm we discuss in this work depends on this ability; for an efficient nonsymmetric DSTSP algorithm, see [18].

C. Control-affine dynamics

A special type of control system is the *control-affine system* [3] (also called *affine-in-control*) in which the control set is $\mathcal{U} \subseteq \mathbb{R}^m$ (so controls are m-dimensional real vectors) and the control function satisfies

$$h^{\mathrm{cont}}(\boldsymbol{q}, \boldsymbol{u}) = h_0^{\mathrm{cont}}(\boldsymbol{q}) + \sum_{i=1}^{m} h_i^{\mathrm{cont}}(\boldsymbol{q}) u_i \tag{4}$$

where $h_0^{\mathrm{cont}}, h_1^{\mathrm{cont}}, \ldots, h_m^{\mathrm{cont}}$ are smooth vector fields on $\mathcal{Q}$. The uncontrolled vector field h_0^{cont} is called the *drift*; if $h_0^{\mathrm{cont}}(\boldsymbol{q}) = \boldsymbol{0}$ (the trivial tangent vector at $\boldsymbol{q}$) for all $\boldsymbol{q} \in \mathcal{Q}$, the system is symmetric, which for control-affine systems is also called *driftless*.

In some control-affine systems, $h_0^{\mathrm{cont}}, \ldots, h_m^{\mathrm{cont}}$ span the tangent space at $\boldsymbol{q}$, allowing linear combinations to move in any direction (subject to drift); however, the systems 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 d_Π around $\boldsymbol{q}$ will contain $\boldsymbol{q}$ in its interior.

D. Metrics and distances

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 $\boldsymbol{q} \in \mathcal{Q}$. When Π is symmetric, this is a metric, 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.

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 – these are not necessarily equivalent because different control inputs to the system might cause motions of different speeds relative to these metrics. We introduce these to aid in the analysis and in particular so that the assumption of a 'speed limit' of the system (see Assumption 4) is well-defined.

E. Reachability sets, small-time constraint factor, and agility function

We first define the distance function induced by dynamics Π:

$$d_{\Pi}(\boldsymbol{q}, \boldsymbol{q}') = \inf_{\pi \in \Pi} (\ell(\pi) : \pi(0) = \boldsymbol{q} \text{ and } \pi(\ell(\pi)) = \boldsymbol{q}') \tag{5}$$

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 $x' \to x''$.

Definition 1. *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*

$$R_{\varepsilon}(\boldsymbol{q}) := \{\boldsymbol{q}' \in \mathcal{Q} : d_{\Pi}(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon\} \quad and \quad \bar{R}_{\varepsilon}(\boldsymbol{q}) := [R_{\varepsilon}(\boldsymbol{q})]_{\mathcal{X}}. \tag{6}$$

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 dynamic system is. In general, this volume will scale polynomially with ε, i.e. $\text{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon}(\boldsymbol{q})) \propto \varepsilon^{\gamma}$ for some appropriate γ, which we call the *small-time constraint factor* (following [16]), 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 systems) 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

$$\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q})) \approx g(\boldsymbol{q})\varepsilon^\gamma \tag{7}$$

Formally we can define these as follows:

Definition 2. *The* small-time constraint factor *of* Π *is* γ *such that at all* $\boldsymbol{q}$

$$\gamma := \lim_{\varepsilon \to 0} \frac{\log(\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q})))}{\log(\varepsilon)} \tag{8}$$

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

$$g(\boldsymbol{q}) := \lim_{\varepsilon \to 0} \frac{\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} . \tag{9}$$

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 system can have when at a configuration projecting to $x \in \mathcal{X}$. This is important because if the system 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 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 3. *The* workspace agility function *is* $g : \mathcal{X} \to \mathbb{R}_{>0}$ *defined as*

$$g(x) := \sup_{\boldsymbol{q} \in [x]^\mathcal{Q}} g(\boldsymbol{q}) \tag{10}$$

F. The Dynamic TSP and Dynamic Stochastic TSP

The Dynamic TSP (DTSP) is then the following problem: given control system Π on $\mathcal{Q}$, and given n targets $x_1, \ldots, x_n \in \mathcal{X}$, find

$$\mathrm{TSP}_\Pi(\{x_i\}) := \mathrm{TSP}_\Pi(x_1, \ldots, x_n) := \inf_{\pi \in \Pi}(\ell(\pi) : x_i \in \bar{\pi} \text{ for all } i) \tag{11}$$

that is, for the shortest valid path that passes through all the targets. We refer to any trajectory which visits all the target points $x_1, \ldots, x_n$ as a *TSP trajectory*.

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 $X_i \overset{iid}{\sim} f$. 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

$$\mathcal{X}_f := \{x \in \mathcal{X} : f(x) > 0\} \tag{12}$$

which we will assume is bounded (Assumption 2).

III. MAIN RESULTS

While our main results depend on a number of conditions, for clarity we will state the results first. The conditions are given in Section IV.

In the DSTSP, the length of the optimal TSP trajectory is a random variable dependent on the placement of the targets $X_i \overset{iid}{\sim} f$, where f is a probability density function on the workspace. Thus, our results take the form of bounds on $\mathrm{TSP}_\Pi(X_1, \ldots, X_n)$ which hold with high probability as $n \to \infty$. We not only want to show that the probability that these bounds fail goes to 0 as $n \to \infty$, we want to show that it approaches 0 rapidly:

Definition 4. *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 *if there are constants* $c_1, c_2, c_3 > 0$ *such that, for all sufficiently large* n,

$$\mathbb{P}[A_n] \geqslant 1 - c_1 e^{-c_2 n^{c_3}} \tag{13}$$

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.

A. 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 IV; 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 2), satisfying the rough relation

$$\sup_{\boldsymbol{q}\in[x]^\mathcal{Q}} \mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q})) \approx g(x)\varepsilon^\gamma \tag{14}$$

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 5. *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:*

$$\beta = (1+\xi)r^\gamma \quad \textit{where} \quad \xi = \begin{cases} 3(\log(b)/r^\gamma) & \textit{if } \log(b) > r^\gamma \\ 3\sqrt{\log(b)/r^\gamma} & \textit{if } \log(b) \leqslant r^\gamma \end{cases} \tag{15}$$

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

For the upper bound, we need to consider *hierarchical cell structures* (HCS) (see Definition 7), which we sketch here. A hierarchical cell structure is a recursively-defined structure of nested cells, having an integer *scaling parameter* s and an *efficiency parameter* $\alpha \leqslant 1$. An HCS 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 the formal definition), which can covered by s^γ HCS's at scale ε/s (the next 'level' down). The scaling parameter s thus denotes how much smaller the HCS'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, HCS's can be constructed with scaling factor $s = 2$; for instance, control-affine systems satisfying common regularity conditions have HCS's with $s = 2$ (see Lemma 6 and Appendix B). However, we make a more general definition because it can be extended to the Euclidean TSP with target points distributed on a set of fractal dimension, which in certain cases

(such as the Menger Sponge) are naturally scaled using $s \neq 2$; see Appendix F for the relevant results.

B. Main concentration bounds

We first state a trivial, non-probabilistic bound which follows from Assumption 5:

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

$$\mathrm{TSP}_\Pi(x_1, \ldots, x_n) \leqslant Cn. \tag{16}$$

Proof. For each x_i, we select (arbitrarily) some $\boldsymbol{q}_i$ such that $[\boldsymbol{q}_i]_{\mathcal{X}} = x_i$; then by Assumption 5, 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 $\leqslant C(n-1)$. $\square$

This will be useful for showing that our probabilistic bounds from Theorem 1 imply equivalent bounds in expectation (Corollary 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 IV hold, then for any $\delta > 0$, the following holds for sufficiently large n:*

$$\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \quad wvhp \tag{17}$$

where β is the lower constant (see Definition 5).

Furthermore, if Π is symmetric and $\mathcal{X}_f$ can be covered with hierarchical cell structures with scaling parameter $s \geqslant 2$ and efficiency parameter $\alpha \leqslant 1$ (see Definition 7), then for any $\delta > 0$ the following holds for sufficiently large n:

$$\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\left(12s\alpha^{-\frac{1}{\gamma}}\right)n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \quad wvhp \tag{18}$$

These results extend to the expected value (see Appendix A for the proof):

Corollary 1. *If the assumptions in Section IV hold, then for any $\delta > 0$,*

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[\mathrm{TSP}_\Pi(\{X_i\})] \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{19}$$

for all sufficiently large n. Furthermore, if Π is symmetric and $\mathcal{X}_f$ can be covered with hierarchical cell structures with scaling parameter $s \geqslant 2$ and efficiency parameter α (see Definition 7), then for any $\delta > 0$,

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[\mathrm{TSP}_\Pi(\{X_i\})] \leqslant (1+\delta)\left(12s\alpha^{-\frac{1}{\gamma}}\right) n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{20}$$

for all sufficiently large n.

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

$$\mathrm{TSP}_\Pi(\{X_i\}) = \Theta\left(n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx\right) \tag{21}$$

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

$$c_{\mathrm{gap}} := c_{\mathrm{gap}}(\Pi) = 12s\alpha^{-\frac{1}{\gamma}}\beta \tag{22}$$

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, (22) 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 IV). 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 system governed by Π can visit many targets distributed iid by f.

Remark 1. *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 F.*

Remark 2. *When Π is nonsymmetric, an upper bound of growth rate $\Theta(n^{1-\frac{1}{\gamma}})$ holds with very high probability; see [18]. However, the dependence of the constant factor on the probability density function f cannot be cleanly characterized in the same way as in (21).*

C. 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 IV hold, and let*

$$v_{f,g} = \mathrm{Var}_{f \sim X}\big[f(X)^{-\frac{1}{\gamma}}g(X)^{-\frac{1}{\gamma}}\big] \tag{23}$$

$$= \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}}g(x)^{-\frac{2}{\gamma}}\,dx - \left(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right)^2 \tag{24}$$

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

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{25}$$

$$\geqslant 1 - \exp\left(-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}\right) \tag{26}$$

$$\quad - \exp\left(-\frac{n(1-\delta/2)\big(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\big)^2\delta^2/32}{(1+\delta/2)v_{f,g} + (1-\delta/2)\big(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\big)^2\delta/12}\right) \tag{27}$$

Proposition 3. *If Π is symmetric with hierarchical cell structure with scaling parameter $s \geqslant 2$ and efficiency parameter α, for any $\delta > 0$: if $\gamma \geqslant 3$,*

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{28}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1-2/\gamma}} \tag{29}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1/3}} \tag{30}$$

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 hierarchical cell structure. Then

$$\mathbb{P}\left[\mathrm{TSP}_{\Pi}(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{31}$$

$$\geqslant 1 - e^{-\frac{1}{2}n^{1-2/\gamma}(\sum_{j=1}^m p_j^{1-1/\gamma})^2} \tag{32}$$

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 hierarchical structure 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,

$$\mathbb{P}\left[\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\left(12s\alpha^{-\frac{1}{\gamma}}\right)n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx\right] \tag{33}$$

$$\geqslant 1 - e^{-\frac{(2/9)p_1 n (\sum_{j=1}^m p_j^{1/2})^2}{127 - \log(1/p_1) + \log(n)}} \tag{34}$$

Remark 3. *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 and can make it converge slower until n is 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 extremely big.*

D. 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 dynamic constraints satisfying the assumptions in Section IV and $\mathcal{X}^* \subseteq \mathcal{X}$ be bounded. Then for any $\delta > 0$, for any sufficiently large n,*

$$\sup_{X_1,\ldots,X_n \in \mathcal{X}^*} \left(\mathrm{TSP}_\Pi(\{X_i\})\right) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}^*} g(x)^{-1}\, dx\right)^{\frac{1}{\gamma}} \tag{35}$$

If Π is symmetric with a hierarchical cell structure with scaling parameter $s \geqslant 2$ and efficiency parameter α, then for any $\delta > 0$,

$$\sup_{X_1,\ldots,X_n \in \mathcal{X}^*} \left(\mathrm{TSP}_\Pi(\{X_i\})\right) \leqslant (1+\delta)\left(6s\alpha^{-\frac{1}{\gamma}}\right)n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}^*} g(x)^{-1}\, dx\right)^{\frac{1}{\gamma}} \tag{36}$$

for any sufficiently large n.

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

$$(1-\delta)\beta^{-1} \leqslant \frac{\sup_{X_1,\ldots,X_n \in \mathcal{X}*} \mathrm{TSP}_{\Pi}(X_1,\ldots,X_n)}{n^{1-\frac{1}{\gamma}}\left(\int_{\mathcal{X}*} g(x)^{-1}\, dx\right)^{\frac{1}{\gamma}}} \leqslant (1+\delta)6s\alpha^{-\frac{1}{\gamma}} \tag{37}$$

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,

$$f_g(x) = \frac{g(x)^{-1}}{\int_{\mathcal{X}*} g(y)^{-1}\, dy}\,. \tag{38}$$

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 derives from our algorithm which covers the space using Hierarchical Cell Structures.

Remark 4. *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 5. *One odd feature of this theorem is that the adversarial upper bound (equation (36)) has a better absolute constant than the equivalent stochastic bound (equation (20)), i.e. 6 as opposed to 12. 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 which was not needed in the adversarial case.*

IV. ASSUMPTIONS

Our main results follow from a few basic assumptions about the properties of the control system Π at very small scales. Our assumptions largely follow from a key meta-assumption:

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

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 $q \in \mathcal{Q}$ always contain a rectilinear neighborhood of a certain size around q. We will note which assumptions follow from the meta-assumption 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); for more details see Appendix B.

Remark 6. *Our results follow if the assumptions in this section are met, even if the meta-assumption itself does not hold.*

A. 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 6. *For any function h and $\zeta > 0$ on domain $\mathcal{Y}$ with metric $d_{\mathcal{Y}}$, let*

$$\widehat{\mathcal{F}}(h, \zeta) := \{h^* : h^*(y) \geqslant \max(h(y), \zeta), \text{ and } h^* \text{ is } (1/\zeta)\text{-Lipschitz continuous}\} \qquad (39)$$

$$\text{and} \quad \check{\mathcal{F}}(h, \zeta) := \{h^* : h^*(y) \leqslant \min(h(y), 1/\zeta), \text{ and } h^* \text{ is } (1/\zeta)\text{-Lipschitz continuous}\}. \qquad (40)$$

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

$$\hat{h}^{(\zeta)}(y) := \inf_{h^* \in \widehat{\mathcal{F}}(h, \zeta)} (h^*(y)) \qquad (41)$$

$$\text{and} \quad \check{h}^{(\zeta)}(y) := \sup_{h^* \in \check{\mathcal{F}}(h, \zeta)} (h^*(y)). \qquad (42)$$

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 $\mathcal{A}$,*

$$\lim_{\zeta \to 0} \int_{\mathcal{A}} \hat{h}^{(\zeta)}(x)\,dx = \int_{\mathcal{A}} h(x)\,dx \tag{43}$$

 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

$$|\hat{h}^{(\zeta)}(y_2) - \hat{h}^{(\zeta)}(y_1)| > (1/\zeta)d_{\mathcal{X}}(y_1, y_2). \tag{44}$$

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

$$h^*(y_1) - \hat{h}^{(\zeta)}(y_1) < |\hat{h}^{(\zeta)}(y_1) - \hat{h}^{(\zeta)}(y_2)| - (1/\zeta)d_{\mathcal{Y}}(y_1, y_2) \tag{45}$$

$$\implies h^*(y_2) - h^*(y_1) \geqslant \hat{h}^{(\zeta)}(y_2) - h^*(y_1) \tag{46}$$

$$> (\hat{h}^{(\zeta)}(y_2) - \hat{h}^{(\zeta)}(y_1)) - \left(|\hat{h}^{(\zeta)}(y_1) - \hat{h}^{(\zeta)}(y_2)| - (1/\zeta)d_{\mathcal{Y}}(y_1, y_2)\right) \tag{47}$$

$$= (1/\zeta)d_{\mathcal{Y}}(y_1, y_2) \tag{48}$$

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 $\hat{\mathcal{F}}(h, \zeta)$. 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. $\qquad\square$

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 system 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.

B. Assumptions about the target distribution

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

Assumption 2. *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 $\int_{\mathcal{X}} f(x)\, dx = 1$. Then for any subset $\mathcal{A} \subseteq \mathcal{X}$,*

$$\mathbb{P}_{X \sim f}[X \in \mathcal{A}] = \int_{\mathcal{A}} f(x)\, dx\,. \tag{49}$$

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.

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 $n > 2$.

C. 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 system 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 3. *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 4. *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 $\|h^{\mathrm{cont}}(\boldsymbol{q}, \boldsymbol{u})\|_{\mathcal{X}} \leqslant c_{\Pi}$.*

Note that Assumption 3 and Assumption 4 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

$$\bar{R}_{\varepsilon}(\boldsymbol{q}) \subseteq \mathcal{B}_{c_{\Pi}\varepsilon}([\boldsymbol{q}]_{\mathcal{X}}) \tag{50}$$

$$\implies \mathrm{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon}(\boldsymbol{q})) \leqslant \mathrm{Vol}_{\mathcal{X}}(\mathcal{B}_{c_{\Pi}\varepsilon}([\boldsymbol{q}]_{\mathcal{X}})) \tag{51}$$

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 $\log(\varepsilon) < 0$,

$$\gamma = \lim_{\varepsilon \to 0} \frac{\log(\mathrm{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon}(\boldsymbol{q})))}{\log(\varepsilon)} \geqslant \frac{\log(\mathrm{Vol}_{\mathcal{X}}(\mathcal{B}_{c_{\Pi}\varepsilon}([\boldsymbol{q}]_{\mathcal{X}})))}{\log(\varepsilon)} = \dim(\mathcal{X}) \geqslant 2 \,. \tag{52}$$

$\square$

We also assume that any two configurations are connected by a finite path $\pi \in \Pi$:

Assumption 5. *For any $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$, there is a valid $\pi \in \Pi$ with finite $\ell(\pi)$ such that $\pi(0) = \boldsymbol{q}$ and $\pi(\ell(\pi)) = \boldsymbol{q}'$.*

In other words, $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') < \infty$. For symmetric equiregular control-affine dynamics satisfying Chow's condition [19], 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).[1]

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 $g : \mathcal{Q} \to \mathbb{R}_{\geq 0}$ *and a* small-time reachability factor γ *such that for any* $\rho > 0$*, there is some* $\varepsilon_\rho^* > 0$ *such that for all* $0 < \varepsilon \leq \varepsilon_\rho^*$ *and all* $\boldsymbol{q} \in \mathcal{Q}$,

$$(1 - \rho)g(\boldsymbol{q}) \leq \frac{\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leq (1 + \rho)g(\boldsymbol{q}) \tag{53}$$

Note that this assumption 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* $0 < \varepsilon \leq \varepsilon_\rho^*$ *and all* $x \in \mathcal{X}$,

$$(1 - \rho)g(x) \leq \sup_{\boldsymbol{q}:[\boldsymbol{q}]_\mathcal{X}=x} \frac{\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leq (1 + \rho)g(x) \tag{54}$$

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

$$\frac{\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \leq (1 + \rho)g(\boldsymbol{q}) \leq (1 + \rho)g(x) \tag{55}$$

Additionally, we have

$$\sup_{\boldsymbol{q}:[\boldsymbol{q}]_\mathcal{X}=x} \frac{\mathrm{Vol}_\mathcal{X}(\bar{R}_\varepsilon(\boldsymbol{q}))}{\varepsilon^\gamma} \geq \sup_{\boldsymbol{q}:[\boldsymbol{q}]_\mathcal{X}=x} (1 - \rho)g(\boldsymbol{q}) = (1 - \rho)g(x) \tag{56}$$

and we have shown both sides of the inequality. $\square$

We also assume bounds on g over the workspace:

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

$$0 < g_{\min} \leq g_{\max} < \infty. \tag{57}$$

---

[1]To apply Theorem 1 for nonsymmetric systems, Assumption 5 must be shown to hold explicitly.

D. 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 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*

$$R_{2\varepsilon}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_{\rho,\varepsilon}(\boldsymbol{q})} R_{(1+\rho)\varepsilon}(\boldsymbol{q}') \ and \ |Q_{\rho,\varepsilon}(\boldsymbol{q})| \leqslant b \tag{58}$$

Lemma 4. *If Π is symmetric and satisfies Meta-Assumption 1, Assumption 8 holds.*

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*

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{\boldsymbol{q} \in Q_\varepsilon^0} R_\varepsilon(\boldsymbol{q}) \ and \ |Q_\varepsilon^0| \leqslant P(1/\varepsilon) \tag{59}$$

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 2) 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 is symmetric and satisfies Meta-Assumption 1, Assumption 9 holds.*

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 Appendix B for the complete proof.

E. Hierarchical Cell Structures

There are some additional conditions which are needed to make our algorithm work (through which we prove our upper bounds on the TSP length), as we need to know that we can range over the whole space in an efficient manner. In particular, we assume the existence of (and ability to cover the whole space with) *hierarchical cell structures*. The hierarchical cell structure 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. We define:

Definition 7. *A ζ-regular 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 $\alpha \leqslant 1$, 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*

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \ and \ \mathrm{Vol}_{\mathcal{X}}(S(\boldsymbol{q}_0, \varepsilon_0)) \geqslant \alpha \check{g}^{(\zeta)}([\boldsymbol{q}_0]_{\mathcal{X}})\varepsilon_0^{\gamma} \tag{60}$$

  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*

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bigcup_{j=1}^{s^{\gamma}} S(\boldsymbol{q}_j, \varepsilon_0/s) \quad and \quad \boldsymbol{q}_j \in \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \ for \ all \ j \tag{61}$$

  We call these the sub-HCS*'s of the original hierarchical cell structure.*

We will sometimes refer to a set of HCS'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 Ball-Box Theorem (see Lemma 6).

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 8.

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 HCS with radius ε_0, efficiency α, and scaling parameter s containing x in its interior.*

We can show this holds for our (well-behaved) control-affine dynamics:

Lemma 6. *If Π is symmetric and satisfies Meta-Assumption 1 then Assumption 10 holds with $s = 2$.*

See Appendix B for proof. Note that this lemma doesn't show anything about how large α is. However, the existence of an α tells us a lot since α does not depend on f (since the definition of the HCS makes no reference to the density function).

We now discuss covering $\mathcal{X}_f$ with Hierarchical Cell Structures, which is the basis of our DTSP algorithm for symmetric dynamics:

Definition 8. *For $\rho \geqslant 0$, a ρ-accurate Hierarchical Cell Structure cover (HCS cover) of $\mathcal{X}_f$ is a set $S_1, \ldots, S_m$ of HCS's such that:*

- *All HCS's have the same parameters (ε_0, α, ζ, s).*
- *$\mathcal{X}_f \subseteq \bigcup_{j=1}^m S_j$.*
- *$\mathbb{P}_{X \sim f}[X$ is in more than one $S_j] \leqslant \rho$.*

Any HCS with overlap parameter α necessarily satisfies the condition for any overlap parameter $\alpha' \leqslant \alpha$; thus, if the HCS's have different values of α, we can use the minimum.

Assumption 10 implies the existence of HCS covers of arbitrarily good accuracy; for proofs

of these results, see Appendix C. First, we have the existence of a fixed-scale finite HCS cover (possibly with bad accuracy ρ):

Lemma 7. *There exist α, s such that for all $\zeta > 0$, there is some ε_0 such that there exists a finite HCS cover (of some accuracy parameter ρ) of $\mathcal{X}_f$.*

Then, given a finite HCS covering, for any $\rho > 0$ we can produce a ρ-accurate covering at arbitrarily small scales by taking the sub-HCS's recursively until we reach the desired scale, and deleting any redundant ones.

Lemma 8. *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 HCS 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.*

In our upper bound proofs in Section VI and Section VII, we will assume that the HCS cover we base our algorithm on is 0-accurate, i.e. there is no overlap at all between different HCS's. See Appendix E for a proof that this does not affect the main results (even without it the results hold) as long as Lemma 8 holds.

Remark 7. *For regular symmetric control-affine dynamics (on which we can invoke the Ball-Box Theorem) and continuous density functions f we can find HCS'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.*

F. 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 region), 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 $\mathcal{X}^* \subseteq \mathcal{X}$ or $\mathcal{Q}^* \subseteq \mathcal{Q}$.

Definition 9. *An assumption holds* approximately everywhere *on a set A in a space $\mathcal{A}$ with metric $d_{\mathcal{A}}$ and volume $\text{Vol}_{\mathcal{A}}$ if, for every $\eta > 0$, there is some A_η such that:*

- *The assumption holds on A_η.*
- *$\text{Vol}_{\mathcal{A}}(A \backslash A'_\eta) \leqslant \eta$ 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 8. *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 8 shows why we need Definition 9 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 4. *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 $\mathcal{X}_f$.*

For the proof, see Appendix D.

V. THE DSTSP LOWER BOUND

We now show a wvhp lower bound for the DSTSP 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 system, and the small-time constraint factor γ. Following the techniques of [18] and [17], we wish to do this using the

connection between Orienteering and the TSP; however, the need to be more precise regarding f and g causes a difficulty. An important difference between the Dynamic Stochastic TSP and Orienteering (on random targets) is 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 dynamic system has a larger range of motion. Thus, the TSP depends on f and g over all of $\mathcal{X}_f$ while 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.

A. *Lucrativity and the Adjusted Cost-Balancing Function*

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 $\bar{R}_\varepsilon(\boldsymbol{q})$:

$$\mathbb{P}_{X \sim f}[X \in \bar{R}_\varepsilon(\boldsymbol{q})] \approx f([\boldsymbol{q}]_{\mathcal{X}})g(\boldsymbol{q})\varepsilon^\gamma \tag{62}$$

$$\leqslant f([\boldsymbol{q}]_{\mathcal{X}})g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma \tag{63}$$

Thus, the expected number of targets that fall within a given radius-ε reachable set anchored at $\boldsymbol{q} \in [x]^{\mathcal{Q}}$ can be bounded by

$$\mathbb{E}_{X_i \overset{iid}{\sim} f}[|\{X_1, \ldots, X_n\} \cap \bar{R}_\varepsilon(\boldsymbol{q})|] \leqslant nf(x)g(x)\varepsilon^\gamma . \tag{64}$$

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

$$\mathcal{c}^*(x) := \mathcal{c}^*_{f,g}(x) := \big(f(x)g(x)\big)^{\frac{1}{\gamma}} \tag{65}$$

which is proportional to the rate at which the system 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 system is 'charged' $\mathcal{c}^*(x)$ 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 9. *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, (63) 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 $\mathglyph{c}^*$ too powerful to provide a useful lower bound to the TSP.*

To deal with the issue raised in Remark 9, we modify the lucrativity function to obtain a cost function for our dynamics:

Definition 10. *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 $\mathglyph{c} : \mathcal{X} \to \mathbb{R}$ is*

$$\mathglyph{c}_\zeta(x) := (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \tag{66}$$

i.e. $\mathglyph{c}_\zeta(x)$ 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 6).

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

$$\ell^{\mathglyph{c}}(\pi) := \ell^{\mathglyph{c}}_\zeta(\pi) := \int_0^{\ell(\pi)} \mathglyph{c}\big(\pi(t)\big)dt. \tag{67}$$

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

$$t^{\mathglyph{c}} := t^{\mathglyph{c}}_\pi := \int_0^t \mathglyph{c}_\zeta(\pi(t')) \, dt' \tag{68}$$

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

$$\pi^{\mathglyph{c}}(t^{\mathglyph{c}}) = \pi(t) \,. \tag{69}$$

We will use the '$\mathglyph{c}$' symbol in general to denote cost-denominated versions of definitions from the previous section, e.g. $\bar{R}^{\mathglyph{c}}_\varepsilon(\boldsymbol{q})$ 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

$$\alpha = \frac{1}{\gamma}\zeta^{\frac{2}{\gamma}-3}(f_{\max} + g_{\max}) \tag{70}$$

  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

$$\mathfrak{c}_\zeta(x) = (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \geqslant (f(x)g(x))^{\frac{1}{\gamma}} = \mathfrak{c}^*(x). \tag{71}$$

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

$$\mathfrak{c}_\zeta(x) = (\hat{f}^{(\zeta)}(x)\hat{g}^{(\zeta)}(x))^{\frac{1}{\gamma}} \geqslant \zeta^{\frac{2}{\gamma}}. \tag{72}$$

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

$$\alpha_1 = \begin{cases} \frac{1}{\gamma}\zeta^{2(\frac{1}{\gamma}-1)} & \text{if } \gamma \geqslant 1 \\ \frac{1}{\gamma}(f_{\max}g_{\max})^{\frac{1}{\gamma}-1} & \text{if } \gamma < 1 \end{cases} \tag{73}$$

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

$$\alpha_2 = (1/\zeta)(f_{\max} + g_{\max}). \tag{74}$$

Hence $\cent_\zeta(x)$ 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

$$\alpha = \frac{1}{\gamma}\zeta^{\frac{2}{\gamma}-3}(f_{\max} + g_{\max}). \tag{75}$$

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). $\qquad\square$

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

Definition 11. *Given $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$, the cost-distance between them is*

$$d_\Pi^\cent(\boldsymbol{q}, \boldsymbol{q}') = \inf(\ell^\cent(\pi) : \boldsymbol{q} \xrightarrow{\pi} \boldsymbol{q}', \pi \in \Pi) \tag{76}$$

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

$$R_\varepsilon^\cent(\boldsymbol{q}) := \{\boldsymbol{q}' \in \mathcal{Q} : d_\Pi^\cent(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon\} \quad \text{and} \quad \bar{R}_\varepsilon^\cent(\boldsymbol{q}) = [R_\varepsilon^\cent(\boldsymbol{q})]_\mathcal{X} \tag{77}$$

Although intuitively on small scales the cost-reachable sets will resemble our original length-reachable sets (except scaled by the cost function $\cent$), 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 $P^\cent$ (which depends through the cost function $\cent := \cent_\zeta$ 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 $Q_\varepsilon^\cent(\boldsymbol{q})$ such that $|Q_\varepsilon^\cent(\boldsymbol{q})| \leqslant b$ and

$$R_{2\varepsilon}^\cent(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_\varepsilon^\cent(\boldsymbol{q})} R_{(1+\rho_2(\zeta))\varepsilon}^\cent(\boldsymbol{q}'). \tag{78}$$

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

$$\mathcal{Q}_f \subseteq \bigcup_{\boldsymbol{q} \in Q_\varepsilon^{\mathcal{C},0}} R_\varepsilon^{\mathcal{C}}(\boldsymbol{q}) \tag{79}$$

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

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

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 $\mathcal{C}_\zeta(x)$, 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

$$\overleftarrow{R}_\varepsilon(\boldsymbol{q}) \text{ and } \overleftrightarrow{R}_\varepsilon^{\mathcal{C}}(\boldsymbol{q}) \tag{80}$$

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 $R_\varepsilon(\boldsymbol{q}) = \overleftarrow{R}_\varepsilon(\boldsymbol{q})$ and $R_\varepsilon^{\mathcal{C}}(\boldsymbol{q}) = \overleftrightarrow{R}_\varepsilon^{\mathcal{C}}(\boldsymbol{q})$).

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

$$\mathcal{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) - (5c_\Pi(\alpha_2/\alpha_1))\varepsilon \leqslant \mathcal{C}_\zeta([\boldsymbol{q}']_{\mathcal{X}}) \leqslant \mathcal{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) + (5c_\Pi(\alpha_2/\alpha_1))\varepsilon \tag{81}$$

$$\text{for any } \boldsymbol{q} \in \mathcal{Q} \text{ and } \boldsymbol{q}' \in \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{82}$$

since $\mathcal{C}$ 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 $\mathcal{C}$ by Lipschitz continuity. Then since $\mathcal{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \geqslant \alpha_1$, we can turn this into multiplicative bounds:

$$(1 - (5c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathcal{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \leqslant \mathcal{C}_\zeta([\boldsymbol{q}']_{\mathcal{X}}) \leqslant (1 + (5c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\,\mathcal{C}_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \tag{83}$$

$$\text{for any } \boldsymbol{q} \in \mathcal{Q} \text{ and } \boldsymbol{q}' \in \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{84}$$

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 $\mathcal{C}_\zeta(x) \geqslant \alpha_1$ for all x, we know that

$$\overleftrightarrow{R}_{5\varepsilon}^{\mathcal{C}}(\boldsymbol{q}) \subseteq \overleftrightarrow{R}_{(5/\alpha_1)\varepsilon}(\boldsymbol{q}) \tag{85}$$

since any trajectory (allowing backwards movement) of cost $\leqslant 5\varepsilon$ must have length $\leqslant (5/\alpha_1)\varepsilon$. Thus, our bounds (83) hold for all $\boldsymbol{q}' \in \overleftrightarrow{R}^{\mathcal{C}}_{5\varepsilon}(\boldsymbol{q})$ as well.

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

$$R_{\varepsilon'/((1+\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq R^{\mathcal{C}}_{\varepsilon'}(\boldsymbol{q}') \subseteq R_{\varepsilon'/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{86}$$

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

$$\overrightarrow{R}_{\varepsilon'/((1+\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq \overrightarrow{R}^{\mathcal{C}}_{\varepsilon'}(\boldsymbol{q}') \subseteq \overrightarrow{R}_{\varepsilon'/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{87}$$

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

We now prove part (i). Equation (86) gives us the following:

$$R^{\mathcal{C}}_{2\varepsilon}(\boldsymbol{q}) \subseteq R_{2\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}) \tag{88}$$

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

$$R_{2\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q})} R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{89}$$

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

$$Q_{\varepsilon'}(\boldsymbol{q}') \subset \overleftrightarrow{R}_{3\varepsilon'}(\boldsymbol{q}') \tag{90}$$

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

$$R_{\varepsilon'}(\boldsymbol{q}'') \cap R_{2\varepsilon'}(\boldsymbol{q}') \neq \varnothing \tag{91}$$

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

$$R_{\varepsilon'}(\boldsymbol{q}'') \subseteq \overleftrightarrow{R}_{4\varepsilon'}(\boldsymbol{q}') \tag{92}$$

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

$$\varepsilon' = \varepsilon/((1 - \rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X})) \text{ and } \boldsymbol{q}' = \boldsymbol{q} \tag{93}$$

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

$$R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subset \overleftrightarrow{R}_{4\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathscr{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}) \tag{94}$$

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

$$\rho_\zeta^*(\varepsilon) \leqslant 1/9 \tag{95}$$

which in turn means

$$4(1 + \rho_\zeta^*(\varepsilon))/(1 - \rho_\zeta^*(\varepsilon)) \leqslant 4(10/9)/(8/9) = 5 \tag{96}$$

and hence by (87) we have

$$\overrightarrow{R}_{4\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}) \subseteq \overleftrightarrow{R}_{5\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}) \tag{97}$$

This then finally yields by (86) that for all $\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q})$,

$$R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq R_{((1+\rho_\zeta^*(\varepsilon))/(1-\rho_\zeta^*(\varepsilon)))\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}') \tag{98}$$

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

$$\rho_1(\zeta) = \zeta\alpha_1^2/(45\alpha_2 c_\Pi) \text{ and } \rho_2(\zeta) = \frac{1 + \zeta/9}{1 - \zeta/9} - 1 \tag{99}$$

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

$$\frac{1 + \rho_\zeta^*(\varepsilon)}{1 - \rho_\zeta^*(\varepsilon)} \leqslant \frac{1 + \rho_\zeta^*(\rho_1(\zeta))}{1 - \rho_\zeta^*(\rho_1(\zeta))} \leqslant \frac{1 + \zeta/9}{1 - \zeta/9} = 1 + \rho_2(\zeta) \tag{100}$$

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

$$R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \subseteq R_{(1+\rho_2(\zeta))\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}') \tag{101}$$

Putting this together with the above, we get

$$R_{2\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}) \subseteq R_{2\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}) \tag{102}$$

$$\subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q})} R_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}') \tag{103}$$

$$\subseteq \bigcup_{\boldsymbol{q}' \in Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q})} R_{(1+\rho_2(\zeta))\varepsilon}^{\mathfrak{c}}(\boldsymbol{q}') \tag{104}$$

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

$$Q_\varepsilon^{\mathfrak{c}}(\boldsymbol{q}) = Q_{\varepsilon/((1-\rho_\zeta^*(\varepsilon))\,\mathfrak{c}_\zeta([\boldsymbol{q}]_\mathcal{X}))}(\boldsymbol{q}). \tag{105}$$

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

Part (iii) follows because when $\varepsilon \leqslant \rho_1(\zeta)$ and $\boldsymbol{q}' \in R_\varepsilon^{\wp}(\boldsymbol{q})$,

$$(1 - (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\, \wp_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \leqslant \wp_\zeta([\boldsymbol{q}']_{\mathcal{X}}) \leqslant (1 + (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\, \wp_\zeta([\boldsymbol{q}]_{\mathcal{X}}) \tag{106}$$

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

$$R_\varepsilon^{\wp}(\boldsymbol{q}) \subseteq R_{\varepsilon/((1-(c_\Pi(\alpha_2/\alpha_1^2))\varepsilon)\, \wp_\zeta([\boldsymbol{q}]_{\mathcal{X}}))}(\boldsymbol{q}) \tag{107}$$

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

$$1 - (c_\Pi(\alpha_2/\alpha_1^2))\varepsilon \geqslant 1 - \zeta/45 \tag{108}$$

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

$$\bar{R}_\varepsilon^{\wp}(\boldsymbol{q}) \subseteq \bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\wp_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q}) \tag{109}$$

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 $\varepsilon \leqslant \rho_{1*}(\zeta)$,

$$\mathrm{Vol}(\bar{R}_\varepsilon(\boldsymbol{q})) \leqslant (1 + \rho_{3*}(\zeta))g(\boldsymbol{q})\varepsilon^\gamma \leqslant (1 + \rho_{3*}(\zeta))g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma \tag{110}$$

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

$$\mathrm{Vol}(\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\wp_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})) \leqslant (1 + \rho_{3*}(\zeta))(1 - \zeta/45)^{-\gamma}\,\wp_\zeta([\boldsymbol{q}]_{\mathcal{X}})^{-\gamma}g([\boldsymbol{q}]_{\mathcal{X}})\varepsilon^\gamma \tag{111}$$

$$= (1 + \rho_{3*}(\zeta))(1 - \zeta/45)^{-\gamma}\hat{f}^{(\zeta)}([\boldsymbol{q}]_{\mathcal{X}})^{-1}\varepsilon^\gamma \tag{112}$$

Finally we note that the entire reachable set is (by definition) within a distance of at most $\varepsilon(1 - \zeta/45)^{-1}/\wp_\zeta([\boldsymbol{q}]_{\mathcal{X}})$ 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 $\varepsilon < \rho_{1**}(\zeta)$,

$$\hat{f}^{(\zeta)}(\boldsymbol{q}') \leqslant (1 + \rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q}) \text{ for all } \boldsymbol{q}' \text{ within distance } \varepsilon \text{ of } \boldsymbol{q} \tag{113}$$

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

$$\mathbb{E}_{X \sim f}[X \in \bar{R}_\varepsilon^{\mathfrak{c}}(\boldsymbol{q})] \leqslant \mathbb{E}_{X \sim f}[X \in \bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})] \tag{114}$$

$$= \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} f(x)\,dx \tag{115}$$

$$\leqslant \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} \hat{f}^{(\zeta)}(x)\,dx \tag{116}$$

$$\leqslant \int_{\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})} (1 + \rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})\,dx \tag{117}$$

$$= (1 + \rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})\mathrm{Vol}(\bar{R}_{\varepsilon(1-\zeta/45)^{-1}/\mathfrak{c}_\zeta([\boldsymbol{q}]_{\mathcal{X}})}(\boldsymbol{q})) \tag{118}$$

$$\leqslant (1 + \rho_{3**}(\zeta))\hat{f}^{(\zeta)}(\boldsymbol{q})(1 + \rho_{3*}(\zeta))(1 - \zeta/45)^{-\gamma}\hat{f}^{(\zeta)}([\boldsymbol{q}]_{\mathcal{X}})^{-1}\varepsilon^\gamma \tag{119}$$

$$= (1 + \rho_{3**}(\zeta))(1 + \rho_{3*}(\zeta))(1 - \zeta/45)^{-\gamma}\varepsilon^\gamma \tag{120}$$

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

$$\rho_3(\zeta) = (1 + \rho_{3**}(\zeta))(1 + \rho_{3*}(\zeta))(1 - \zeta/45)^{-\gamma} - 1 \tag{121}$$

and the above will hold for all ε such that

$$\varepsilon \leqslant \rho_{1*}(\zeta)(1 - \zeta/45)\zeta^{\frac{2}{\gamma}} \text{ and } \rho_{1**}(\zeta)(1 - \zeta/45)\zeta^{\frac{2}{\gamma}} \tag{122}$$

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

$$\mathbb{E}_{X \sim f}[(f(X)g(X))^{-\frac{1}{\gamma}}] = \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx < \infty \tag{123}$$

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

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).

B. Cost-Balanced Orienteering

Definition 12. *The* cost-bounded trajectory set *of dynamic system Π is denoted*

$$\Pi_\lambda^\mathfrak{c} := \{\pi \in \Pi : \ell^\mathfrak{c}(\pi) \leqslant \lambda\} \tag{124}$$

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

$$\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda) := \max_{\pi \in \Pi_\lambda^\mathfrak{c}}(|\bar{\pi} \cap \{X_i\}_{i=1}^n|) \tag{125}$$

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 want to 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 5 (CBO upper bound)**.** *Let Π be a symmetric dynamic system and f be a probability density function satisfying the assumptions in Section IV, 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*

$$\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \leqslant \beta \lambda \, n^{\frac{1}{\gamma}} \tag{126}$$

with very high probability, where β is a constant that 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 $\mathfrak{c}_\zeta$ to preserve the balance.

We make the constants and probability bound of Proposition 5 more explicit:

Proposition 6. *If the assumptions in Section IV hold, then for any $\delta > 0$,*

$$\mathbb{P}[\mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \leqslant (1 + \delta)\beta \lambda \, n^{\frac{1}{\gamma}}] \geqslant 1 - e^{-\frac{4}{5}(1+\delta)\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{127}$$

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

Definition 13. *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*

$$\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_{\lceil \lambda/((1-\rho)\varepsilon) \rceil}) \quad where \quad \psi_k \in \mathcal{Q} \text{ for all } k \tag{128}$$

$$such \ that \quad \psi_0 \in Q^{\varrho,0}_{(1-\rho)\varepsilon} \text{ and } \psi_k \in Q^{\varrho}_{(1-\rho/2)\varepsilon}(\psi_{k-1}) \text{ for all } k \geqslant 1 \tag{129}$$

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

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

$$\pi^{\varrho}(k(1-\rho)\varepsilon) \in R^{\varrho}_{\varepsilon}(\psi_k) \text{ for all } k \in \{0, 1, \ldots, n\} \tag{130}$$

We denote the set of all representation sequences (with parameters $\lambda, \varepsilon, \rho$*) as* $\Psi^{\varrho}_{\lambda,\varepsilon,\rho}$. *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*

$$R^{\varrho}_{\varepsilon}(\boldsymbol{\psi}) := \bigcup_{k=0}^{\lceil \lambda/((1-\rho)\varepsilon) \rceil} R^{\varrho}_{\varepsilon}(\psi_k) \tag{131}$$

and we denote its workspace projection as

$$\bar{R}^{\varrho}_{\varepsilon}(\boldsymbol{\psi}) := \bigcup_{k=0}^{\lceil \lambda/((1-\rho)\varepsilon) \rceil} \bar{R}^{\varrho}_{\varepsilon}(\psi_k). \tag{132}$$

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* $\pi \in \Pi^{\varrho}_{\lambda}$ *such that* $\pi(0) \in [\mathcal{X}_f]^{\mathcal{Q}}$ *is represented by some* $\boldsymbol{\psi} \in \Psi^{\varrho}_{\lambda,\varepsilon,\rho}$.

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 $Q^{\varrho}_{(1-\rho)\varepsilon}(\boldsymbol{q})$ such that $|Q^{\varrho}_{\varepsilon}(\boldsymbol{q})| \leqslant b$ and

$$R^{\varrho}_{(2-\rho)\varepsilon}(\boldsymbol{q}) = R^{\varrho}_{2(1-\rho/2)\varepsilon}(\boldsymbol{q}) \tag{133}$$

$$\subseteq \bigcup_{\boldsymbol{q}' \in Q^{\varrho}_{(1-\rho/2)\varepsilon}(\boldsymbol{q})} R^{\varrho}_{(1+\rho_2(\zeta))(1-\rho/2)\varepsilon}(\boldsymbol{q}') \subseteq \bigcup_{\boldsymbol{q}' \in Q^{\varrho}_{(1-\rho/2)\varepsilon}(\boldsymbol{q})} R^{\varrho}_{\varepsilon}(\boldsymbol{q}'). \tag{134}$$

We then construct $\boldsymbol{\psi} \in \Psi_{\lambda,\varepsilon,\rho}^{\wp}$ such that $\boldsymbol{\psi} \sqsupset \pi$ inductively.

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

Inductive step: For any $k > 0$, given ψ_{k-1} such that $\pi^{\wp}((k-1)(1-\rho)\varepsilon) \in R_{\varepsilon}^{\wp}(\psi_{k-1})$, we need to choose $\psi_k \in Q_{(1-\rho/2)\varepsilon}^{\wp}(\psi_{k-1})$ such that $\pi^{\wp}(k(1-\rho)\varepsilon) \in R_{\varepsilon}^{\wp}(\psi_k)$. We do this by noting that by definition $\pi^{\wp}(k(1-\rho)\varepsilon) \in R_{\varepsilon}^{\wp}(\pi^{\wp}((k-1)(1-\rho)\varepsilon))$ (since they are only cost $(1-\rho)\varepsilon$ apart on valid trajectory π). Thus, $\pi^{\wp}(k(1-\rho)\varepsilon) \in R_{(2-\rho)\varepsilon}^{\wp}(\psi_{k-1})$ (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 (133) we know that $R_{(2-\rho)\varepsilon}^{\wp}(\psi_{k-1})$ is covered by the ε-radius cost reachable sets from points in $Q_{(1-\rho/2)\varepsilon}^{\wp}(\psi_{k-1})$, and hence we can choose the one containing $\pi^{\wp}(k(1-\rho)\varepsilon) \in R_{(2-\rho)\varepsilon}^{\wp}(\psi_{k-1})$ to be ψ_k, thus maintaining $\psi_k \in Q_{(1-\rho/2)\varepsilon}^{\wp}(\psi_{k-1})$ and $\pi^{\wp}(k(1-\rho)\varepsilon) \in R_{\varepsilon}^{\wp}(\psi_k)$.

Thus, we can build $\psi_0, \psi_1, \ldots$ like this until we have covered all of π. Since $\pi^{\wp}$ only takes in inputs from 0 to λ, we only need to do this up to $\pi^{\wp}(\lfloor \lambda/((1-\rho)\varepsilon) \rfloor \varepsilon)$ 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 $Q_{\varepsilon}^{\wp}(\psi_{\lfloor \lambda/((1-\rho)\varepsilon) \rfloor})$). $\qquad\square$

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

Proof. For any $\pi^{\wp}(t)$, 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 $\pi^{\wp}(t) \in R_{2\varepsilon}^{\wp}(\psi_k)$ since $\pi^{\wp}(k(1-\rho)\varepsilon)$ is within ε cost from ψ_k and $\pi^{\wp}(t)$ is within $\tau < (1-\rho)\varepsilon < \varepsilon$ cost from $\pi^{\wp}(k(1-\rho)\varepsilon)$.

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 $\pi^{\wp}(t)$ is at most $\tau \leqslant \varepsilon/2$ cost (moving forwards) from $\pi^{\wp}(k(1-\rho)\varepsilon)$, which is at most ε cost from ψ_k, hence

$$\pi^{\wp}(t) \in R_{(3/2)\varepsilon}^{\wp}(\psi_k) \subseteq R_{(3/2)\varepsilon}^{\wp}(\boldsymbol{\psi}). \tag{135}$$

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

from ψ_{k+1}, hence

$$\pi^{\mathfrak{e}}(t) \in R^{\mathfrak{e}}_{(3/2)\varepsilon}(\psi_{k+1}) \subseteq R^{\mathfrak{e}}_{(3/2)\varepsilon}(\boldsymbol{\psi}). \tag{136}$$

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

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

Lemma 13. *Let b be the branching factor and $P^{\mathfrak{e}}$ the polynomial from Lemma 10(ii),*

$$|\Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho}| \leqslant P^{\mathfrak{e}}(1/((1-\rho)\varepsilon))b^{\lceil\lambda/((1-\rho)\varepsilon)\rceil} \tag{137}$$

Additionally, for any fixed $\delta > 0$, there is some $\rho^{(\delta)} > 0$ and $\varepsilon^{(\delta)} > 0$ such that for all $0 < \varepsilon \leqslant \varepsilon^{(\delta)}$,

$$|\Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho^{(\delta)}}| \leqslant b^{(1+\delta)\lambda/\varepsilon}. \tag{138}$$

Proof. This follows from the fact that for any $\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_{\lceil\lambda/((1-\rho)\varepsilon)\rceil}) \in \Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho}$, we have $\psi_0 \in Q^{\mathfrak{e},0}_\varepsilon$ and $\psi_k \in Q^{\mathfrak{e}}_\varepsilon(\psi_{k-1})$ for all $k > 0$. Thus, we have $|Q^{\mathfrak{e},0}_{(1-\rho)\varepsilon}| \leqslant P^{\mathfrak{e}}(1/((1-\rho)\varepsilon))$ choices for ψ_0, and then for each $k > 0$ (iterating from 1 to $\lceil\lambda/((1-\rho)\varepsilon)\rceil$) we have $|Q^{\mathfrak{e}}_{(1-\rho/2)\varepsilon}(\psi_{k-1})| \leqslant b$ choices. Thus the total number of ways to construct $\boldsymbol{\psi} \in \Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho}$ is at most $P^{\mathfrak{e}}(1/((1-\rho)\varepsilon))b^{\lceil\lambda/((1-\rho)\varepsilon)\rceil}$.

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

$$b^{(1+\delta)\lambda/\varepsilon}/b^{\lceil\lambda/((1-\rho^{(\delta)})\varepsilon)\rceil} \geqslant b^{((1+\delta)-1/(1-\rho^{(\delta)}))\lambda/\varepsilon-1} \geqslant b^{(\delta/2)\lambda/\varepsilon-1} \tag{139}$$

grows faster than any polynomial in $1/\varepsilon$, and specifically it must grow faster than $P^{\mathfrak{e}}(1/((1-\rho^{(\delta)})\varepsilon))$. Thus, for sufficiently small ε,

$$|\Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho^{(\delta)}}| \leqslant P^{\mathfrak{e}}(1/((1-\rho)\varepsilon))b^{\lceil\lambda/((1-\rho^{(\delta)})\varepsilon)\rceil} \leqslant b^{(1+\delta)\lambda/\varepsilon} \tag{140}$$

and we are done. $\qquad\square$

Furthermore, given any constant $r \geqslant 0$, the probability mass of any $\bar{R}^{\mathfrak{e}}_{r\varepsilon}(\boldsymbol{\psi})$ (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 $\boldsymbol{\psi} \in \Psi^{\mathfrak{e}}_{\lambda,\varepsilon,\rho^{(\delta)}}$,*

$$\mathbb{P}_{X \sim f}[X \in \bar{R}^{\mathfrak{e}}_{r\varepsilon}(\boldsymbol{\psi})] \leqslant (1+\delta)r^\gamma\lambda\varepsilon^{\gamma-1}. \tag{141}$$

Proof. We know that for any $\psi_k \in \mathcal{Q}$,

$$\mathbb{P}_{X \sim f}[X \in \bar{R}_\varepsilon^{\mathcal{c}}(\psi_k)] \leqslant (1 + o(1))\varepsilon^\gamma \tag{142}$$

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

$$\lceil \lambda/((1 - \rho^{(\delta)})\varepsilon) \rceil + 1 \leqslant (1 + \delta/2)\lambda/\varepsilon + 2 \leqslant (1 + (3/4)\delta)\lambda/\varepsilon \tag{143}$$

possible values of k. We can then take ε sufficiently small so that $\mathbb{P}_{X \sim f}[X \in \bar{R}_{r\varepsilon}^{\mathcal{c}}(\psi_k)] \leqslant (1 + \delta^*)r^\gamma\varepsilon^\gamma$, 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

$$\mathbb{P}_{X \sim f}[X \in \bar{R}_{r\varepsilon}^{\mathcal{c}}(\boldsymbol{\psi})] \leqslant \bigcup_{k=0}^{\lceil \lambda/((1 - \rho^{(\delta)})\varepsilon) \rceil} \mathbb{P}_{X \sim f}[X \in \bar{R}_{r\varepsilon}^{\mathcal{c}}(\psi_k)] \tag{144}$$

$$\leqslant (\lceil \lambda/((1 - \rho^{(\delta)})\varepsilon) \rceil + 1)(1 + \delta^*)r^\gamma\varepsilon^\gamma \tag{145}$$

$$\leqslant (1 + \delta)r^\gamma\lambda\varepsilon^{\gamma-1} \tag{146}$$

as we wanted. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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 14. *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 $\mathcal{c}$). If Π is nonsymmetric, let $r = 2$; if Π is symmetric, let $r = 3/2$; then*

$$\mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) := \max_{\boldsymbol{\psi} \in \Psi_{\lambda, \varepsilon, \rho}^{\mathcal{c}}} |\{X_1, \ldots, X_n\} \cap \bar{R}_{r\varepsilon}^{\mathcal{c}}(\boldsymbol{\psi})| \tag{147}$$

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 $\Psi_{\lambda, \varepsilon, \rho}^{\mathcal{c}}$. 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 $\varepsilon > 0$,*

$$\mathrm{CBO}_\Pi(X_1, \ldots, X_n; \lambda) \leqslant \mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) \tag{148}$$

Proof. This follows from Lemmas 11 and 12: for any $\pi \in \Pi_\lambda^{\mathscr{C}}$, there exists $\psi \in \Psi_{\lambda,\varepsilon,\rho}^{\mathscr{C}}$ such that $\psi \sqsupset \pi$, and this implies $\bar{\pi} \subseteq \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)$. Thus,

$$\left|\{X_1, \ldots, X_n\} \cap \bar{\pi}\right| \leqslant \left|\{X_1, \ldots, X_n\} \cap \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)\right| \tag{149}$$

and so the maximum of the former over $\pi \in \Pi$ is at most the maximum of the latter over $\psi \in \Psi_{\lambda,\varepsilon,\rho}^{\mathscr{C}}$. $\qquad\square$

We now prove Proposition 6. 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 ψ (arbitrarily) before $X_1, \ldots, X_n \overset{iid}{\sim} f$ are chosen and then looking at $\left|\{X_1, \ldots, X_n\} \cap \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)\right|$ as a random variable (since it takes random inputs $X_1, \ldots, X_n$). We let $p^{(\psi)} := \mathbb{P}_{X \sim f}[X \in \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)]$. Plugging into Lemma 14 yields

$$p^{(\psi)} = \mathbb{P}_{X \sim f}[X \in \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)] \leqslant (1+\delta)\lambda r^\gamma n^{-(1-\frac{1}{\gamma})} \tag{150}$$

$$\implies \mathbb{E}_{X_i \overset{iid}{\sim} f}\left[\left|\{X_1, \ldots, X_n\} \cap \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)\right|\right] \leqslant (1+\delta)\lambda r^\gamma n^{\frac{1}{\gamma}} \tag{151}$$

Additionally, with $\psi \in \Psi_{\lambda,\varepsilon,\rho}^{\mathscr{C}}$ fixed, we can let $Z_i := Z_i^{(\psi)} := \mathbb{1}\{X_i \in \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)\}$ and

$$Z := Z^{(\psi)} := \sum_{i=1}^n Z_i^{(\psi)} = \left|\{X_1, \ldots, X_n\} \cap \bar{R}_{r\varepsilon}^{\mathscr{C}}(\psi)\right| \tag{152}$$

We define $p^* = (1+\delta)r^\gamma n^{-(1-\frac{1}{\gamma})}$. Then we know that $Z_i \overset{iid}{\sim} \mathrm{Bern}(p^{(\psi)})$, and that for sufficiently small ε (which translates to sufficiently large n since $\varepsilon \to 0$ as $n \to \infty$) we have $p^{(\psi)} \leqslant p^*$ for all ψ; thus, we can WLOG assume $p^{(\psi)} \leqslant p^*$ for all $\psi \in \Psi_{\lambda,\varepsilon,\rho}^{\mathscr{C}}$.

We then define $Z_1^*, \ldots, Z_i^* \overset{iid}{\sim} \mathrm{Bern}(p^*)$ and $Z^* = \sum_{i=1}^n Z_i^*$. Since $p^{(\psi)} \leqslant p^*$, for any $A > 0$,

$$\mathbb{P}[Z^{(\psi)} \geqslant A] \leqslant \mathbb{P}[Z^* \geqslant A] \tag{153}$$

Let $\mu^* := \mathbb{E}[Z^*] = (1+\delta)r^\gamma \lambda n^{\frac{1}{\gamma}}$.

Therefore we use Lemma 15, the union bound, equation (153), and Lemma 13 (in that order) to get that for all $A > 0$ (where the probabilities are over $X_i \overset{iid}{\sim} f$):

$$\mathbb{P}\big[\mathrm{CBO}_\Pi(X_1, \ldots, X_n; f, \lambda) \geqslant A\big] \leqslant \mathbb{P}\big[\mathrm{SCP}_\Pi(X_1, \ldots, X_n; \lambda, \varepsilon, \delta) \geqslant A\big] \tag{154}$$

$$= \mathbb{P}\big[\exists\, \boldsymbol{\psi} \in \Psi^e_{\lambda,\varepsilon,\rho} : Z^{(\boldsymbol{\psi})} \geqslant A\big] \tag{155}$$

$$\leqslant \sum_{\boldsymbol{\psi} \in \Psi^e_{\lambda,\varepsilon,\rho}} \mathbb{P}\big[Z^{(\boldsymbol{\psi})} \geqslant A\big] \tag{156}$$

$$\leqslant |\Psi^e_{\lambda,\varepsilon,\rho}|\, \mathbb{P}\big[Z^* \geqslant A\big] \tag{157}$$

$$\leqslant b^{(1+\delta)\lambda/\varepsilon}\, \mathbb{P}\big[Z^* \geqslant A\big] \tag{158}$$

$$= b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}}\, \mathbb{P}\big[Z^* \geqslant A\big] \tag{159}$$

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

$$\mathbb{P}[Z^* \geqslant A] = \mathbb{P}[Z^* \geqslant (1 + \xi)\mu^*] \tag{160}$$

$$\leqslant e^{-\frac{\xi^2}{2+\xi}\mu^*} \tag{161}$$

$$= e^{-\frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda n^{\frac{1}{\gamma}}} \tag{162}$$

We then note that equation (159), which is an upper bound on the probability that the CBO problem yields a result $\geqslant A$, is

$$b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}}\, \mathbb{P}[Z^* \geqslant A] \leqslant e^{\log(b)\,(1+\delta)\,\lambda n^{\frac{1}{\gamma}} - \frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda n^{\frac{1}{\gamma}}} \tag{163}$$

$$= e^{\left(\log(b) - \frac{\xi^2}{2+\xi}r^\gamma\right)\lambda(1+\delta)n^{\frac{1}{\gamma}}} \tag{164}$$

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 (15): 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

$$\log(b) - \frac{\xi^2}{2+\xi}r^\gamma = \log(b) - \frac{9(\log(b)/r^\gamma)^2}{2 + 3(\log(b)/r^\gamma)}r^\gamma \tag{165}$$

$$\leqslant \log(b) - \frac{9(\log(b)/r^\gamma)^2}{5(\log(b)/r^\gamma)}r^\gamma \tag{166}$$

$$= -\frac{4}{5}\log(b) \tag{167}$$

Similarly, if $\log(b) \leqslant r^\gamma$ we get

$$\log(b) - \frac{\xi^2}{2 + \xi} r^\gamma = \log(b) - \frac{9(\log(b)/r^\gamma)}{2 + 3\sqrt{\log(b)/r^\gamma}} r^\gamma \tag{168}$$

$$\leqslant \log(b) - \frac{9(\log(b)/r^\gamma)}{5} r^\gamma \tag{169}$$

$$= -\frac{4}{5} \log(b) \tag{170}$$

and hence we have in either case the bound

$$\log(b) - \frac{\xi^2}{2 + \xi} r^\gamma \leqslant -\frac{4}{5} \log(b) \tag{171}$$

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

$$\mathbb{P}[\mathrm{CBO}_\Pi(\{X_i\}; f, \lambda) \geqslant (1 + \delta)\beta\lambda n^{\frac{1}{\gamma}}] \leqslant \mathbb{P}[\mathrm{SCP}_\Pi(\{X_i\}; \lambda, \varepsilon, \rho) \geqslant (1 + \delta)\beta\lambda n^{\frac{1}{\gamma}}] \tag{172}$$

$$\leqslant |\Psi^c_{\lambda, \varepsilon, \rho}| \; \mathbb{P}[Z^* \geqslant (1 + \xi)\mu^*] \tag{173}$$

$$\leqslant b^{(1+\delta)\lambda n^{\frac{1}{\gamma}}} e^{\left(-\frac{\xi^2}{2+\xi}(1+\delta)r^\gamma \lambda \varepsilon^{\gamma-1} n\right)} \tag{174}$$

$$\leqslant e^{-\frac{4}{5}(1+\delta)\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{175}$$

thus proving Propositions 5 and 6.

C. 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

$$\Pi^{\{X_i\}} := \{\pi \in \Pi : X_i \in \bar{\pi} \text{ for all } i\} \tag{176}$$

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

$$\sum_{i=1}^{n} \ell_{j(i)} \approx \sum_{i=1}^{n} \lambda / \mathfrak{c}_\zeta(X_i)^{-1} \approx n\lambda \mathbb{E}_{X\sim f}[\mathfrak{c}_\zeta(X)^{-1}] \approx (1-\delta)n\lambda \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx. \qquad (177)$$

However, this may well overcount the total length since each segment is counted once per X_i that falls on it; but by Proposition 6 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,

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \geqslant (1-\delta)\beta^{-1} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \qquad (178)$$

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

$$\mathrm{CBO}^{(n)} := \mathrm{CBO}_\Pi(X_1, X_2, \ldots, X_n; \lambda, f) \qquad (179)$$

Let β and ξ be as defined in Definition 5. 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 6 holds for all $\delta > 0$ for all sufficiently large n (where 'sufficiently large' may depend on δ),

$$\mathbb{P}[\mathrm{CBO}^{(n)} \leqslant (1 - \delta/2)\beta\lambda \, n^{\frac{1}{\gamma}}] \geqslant 1 - e^{-\frac{4}{5} \frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1 - \delta/2}} \geqslant 1 - e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} \qquad (180)$$

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:

$$k := \lceil \ell^{\mathfrak{e}}(\pi)/\lambda \rceil \quad \text{and} \quad \lambda' := \ell^{\mathfrak{e}}(\pi)/k \qquad (181)$$

Thus, $\lambda - 1/k \leqslant \lambda' \leqslant \lambda$. Furthermore, since the bound on $\text{CBO}^{(n)}$ 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

$$k \geqslant \frac{n}{(1 - \delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}}} = (1 - \delta/2)\beta^{-1}\lambda^{-1}n^{1 - \frac{1}{\gamma}} \tag{182}$$

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

$$t_j' := \min\left(t : \int_0^t \mathcal{c}_\zeta(\pi(\tau))\, d\tau = j\lambda'\right) \tag{183}$$

i.e. the trajectory reaches cost $j\lambda'$ at time t_j'. This also means that

$$\ell^{\mathcal{c}}(\pi_{[t_{j-1}', t_j']}) = \lambda' \tag{184}$$

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 $\text{CBO}^{(n)}$ is bounded and by (184), we know that for sufficiently large n, all $j \in [k]$ satisfy

$$|A_j| \leqslant (1 - \delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}}. \tag{185}$$

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 $\ell^{\mathcal{c}}(\pi)$ as described above). Note that the collection $\{A_j\}_{j\in[k]}$ partitions $[n]$.

Since $\mathcal{c}_\zeta(x) \geqslant \alpha_1$ everywhere and $\mathcal{c}_\zeta$ 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 $\ell^{\mathcal{c}}(\pi') = \lambda'$ and $X \in \bar{\pi}'$ then $\ell(\pi') \geqslant (1 - \eta)\lambda'/\mathcal{c}_\zeta(X)$. We then derive the following:

$$\ell(\pi) = \sum_{j=1}^k \ell_j \tag{186}$$

$$= \frac{1 - \delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{j=1}^k (1 - \delta/2)^{-1}\beta\lambda n^{\frac{1}{\gamma}} \ell_j \tag{187}$$

$$\geqslant \frac{1 - \delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{j=1}^k |A_j| \ell_j \tag{188}$$

$$= \frac{1 - \delta/2}{\beta\lambda n^{\frac{1}{\gamma}}} \sum_{i=1}^n \ell_{j(i)} \tag{189}$$

This then leads to the intuitive steps (note that we use $\lambda \approx \lambda'$ and remove the δ terms):

$$\ell(\pi) \gtrsim \frac{1}{\beta \lambda n^{\frac{1}{\gamma}}} \sum_{i=1}^{n} \lambda / \mathfrak{c}_\zeta(X_i) \text{ with very high probability} \tag{190}$$

$$\gtrsim \frac{1}{\beta} n^{1-\frac{1}{\gamma}} \mathbb{E}_{X \sim f}\left[\mathfrak{c}_\zeta(X)^{-1} \right] \text{ with very high probability} \tag{191}$$

$$\approx \frac{1}{\beta} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{192}$$

The two '$\gtrsim$' steps (190) and (191) need to be precisely stated and justified: (190) 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 (191) is shown with a concentration bound (and because when n is large $\lambda' \approx \lambda$). The approximation in (192) 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 (190) and (191), we discuss our main tool for showing them, which is the *one-sided Bernstein condition* and *Bernstein's inequality* (see [20]). 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 7. *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 8. *Let $Y_1, \ldots Y_n$ 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 $\delta > 0$,*

$$\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^{n} Y_i - \mathbb{E}[Y] \geqslant \delta \right] \leqslant e^{-\frac{n\delta^2/2}{\mathrm{Var}[Y] + a\delta}} \tag{193}$$

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 9. *Let $Y_1, \ldots, Y_n$ be iid nonnegative random variables (each with expectation $\mathbb{E}[Y]$ and variance $\mathrm{Var}[Y]$). Then for any $\delta > 0$,*

$$\mathbb{P}\Big[\sum_{i=1}^{n} Y_i \leqslant (1-\delta)n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n\mathbb{E}[Y]^2\delta^2/2}{\mathrm{Var}[Y]+\mathbb{E}[Y]^2\delta/2}} \tag{194}$$

Proof. By Proposition 7, if Y is nonnegative then $-Y$ satisfies the Bernstein condition with parameter $a = \mathbb{E}[Y]/3$. Then we can re-write

$$\sum_{i=1}^{n} Y_i \leqslant (1-\delta)n\mathbb{E}[Y] \iff \sum_{i=1}^{n}(-Y_i) \geqslant (1-\delta)n\mathbb{E}[-Y] \tag{195}$$

$$\iff \frac{1}{n}\sum_{i=1}^{n}(-Y_i) - \mathbb{E}[-Y] \geqslant \delta\,\mathbb{E}[Y] \tag{196}$$

Plugging this into Proposition 8 then yields the result. $\qquad\square$

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*

$$(1-\delta)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \leqslant \mathbb{E}[Y] \leqslant \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \tag{197}$$

and $\quad \mathrm{Var}[Y] \leqslant \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}}g(x)^{-\frac{2}{\gamma}}\,dx - (1-\delta)^2\left(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right)^2 \tag{198}$

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:

$$(1-\delta)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \leqslant \mathbb{E}[Y] \leqslant \alpha_1^{-1} \tag{199}$$

and $\quad \mathrm{Var}[Y] \leqslant \alpha_1^{-2}/4\,. \tag{200}$

Proof. We recall that the lucrativity function (of which the cost function $\mathcal{c}_\zeta$ is an approximation) is $\mathcal{c}^*(x) = (f(x)g(x))^{\frac{1}{\gamma}}$ and that $\gamma \geqslant 2$. Let $Y^* = \mathcal{c}^*(X)$ where $X \sim f$. Then we have

$$\mathbb{E}[Y^*] = \int (\mathcal{c}^*)^{-1} \, df \tag{201}$$

$$= \int_{\mathcal{X}_f} f(x)(f(x)g(x))^{-\frac{1}{\gamma}} \, dx \tag{202}$$

$$= \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{203}$$

$$\leqslant \int_{\mathcal{X}_f} (1 + f(x)) g_{\min}^{-\frac{1}{\gamma}} \, dx \tag{204}$$

$$= (\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f) + 1) g_{\min}^{-\frac{1}{\gamma}} \tag{205}$$

$$< \infty \tag{206}$$

where (204) follows because $g(x) \geqslant g_{\min}$ for all x and

$$f(x)^{1-\frac{1}{\gamma}} \leqslant \max(1, f(x)) \leqslant 1 + f(x) \tag{207}$$

(and $\int_{\mathcal{X}_f} f(x) \, dx = 1$ by definition). Similarly,

$$\mathbb{E}[(Y^*)^2] = \int_{\mathcal{X}_f} f(x)^{1-\frac{2}{\gamma}} g(x)^{-\frac{2}{\gamma}} \, dx \tag{208}$$

$$\leqslant \int_{\mathcal{X}_f} (1 + f(x)) g_{\min}^{-\frac{2}{\gamma}} \, dx \tag{209}$$

$$= (\mathrm{Vol}_{\mathcal{X}}(\mathcal{X}_f) + 1) g_{\min}^{-\frac{2}{\gamma}} \tag{210}$$

$$< \infty \tag{211}$$

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 $\mathcal{c}_\zeta(x) \geqslant \mathcal{c}^*(x)$ and hence if $\gamma \geqslant 2$, we have $\mathcal{c}_\zeta(x)^{-1} \leqslant \mathcal{c}^*(x)^{-1}$ 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 $\mathcal{c}_\zeta(x) \geqslant \alpha_1$, and hence $\mathbb{E}[\mathcal{c}_\zeta(X)^{-1}] \leqslant \alpha_1^{-1}$ and $\mathrm{Var}[\mathcal{c}_\zeta(X)^{-1}] \leqslant \alpha_1^{-2}/4$ (since $\mathcal{c}_\zeta(X)^{-1} \in (0, \alpha_1^{-1}]$ its variance is at most $\alpha_1^{-2}/4$). Thus, we are done. $\square$

Returning to the main proof, we first address step (190):

Lemma 17. *For any $\delta_1 > 0$, there is a sufficiently small $\lambda > 0$ such that (when $X_i \overset{iid}{\sim} f$),*

$$\mathbb{P}\left[\exists\, \pi \in \Pi^{\{X_i\}} \text{ s.t. } \sum_{i=1}^{n} \ell_{j(i)} < (1 - \delta_1)\lambda n\mathbb{E}[Y]\right] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y] + (1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{212}$$

for all sufficiently large n.

Proof. For any $x \in \mathcal{X}_f$ let

$$\ell^*(x; \lambda) = \min_{\pi' \in \Pi}(\ell(\pi') : \ell^{\mathfrak{e}}(\pi') = \lambda, x \in \bar{\pi}') \tag{213}$$

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 $0 < \lambda \leqslant \lambda_\eta^*$,

$$\mathbb{P}_{X\sim f}[\exists\, \pi' \in \Pi : X \in \bar{\pi}',\ \ell^{\mathfrak{e}}(\pi') = \lambda \text{ and } \ell(\pi') < (1 - \eta)\lambda/\mathfrak{c}_\zeta(X)] \leqslant \eta \tag{214}$$

We define $\mathcal{X}(\eta) \subseteq \mathcal{X}_f$ to be the region in which

$$x \in \bar{\pi}',\ \ell^{\mathfrak{e}}(\pi') = \lambda \implies \ell(\pi') \geqslant (1 - \eta)\lambda/\mathfrak{c}_\zeta(x) \tag{215}$$

for $\lambda \leqslant \lambda_\eta^*$ breaks down, i.e.

$$\mathcal{X}(\eta) = \{x \in \mathcal{X}_f : \exists\, \pi' \in \Pi : x \in \bar{\pi}',\ \ell^{\mathfrak{e}}(\pi') \leqslant \lambda_\eta^* \text{ and } \ell(\pi') < (1 - \eta)\ell^{\mathfrak{e}}(\pi')/\mathfrak{c}_\zeta(x)\} \tag{216}$$

Then, if we have some $x \in \mathcal{X}_f$ and a trajectory π' s.t. $x \in \bar{\pi}'$ and $\ell^{\mathfrak{e}}(\pi') = \lambda < \lambda_\eta^*$,

$$\ell(\pi') \geqslant \begin{cases} (1 - \eta)\lambda'/\mathfrak{c}_\zeta(x) & \text{if } x \notin \mathcal{X}(\eta) \\ 0 & \text{if } x \in \mathcal{X}(\eta) \end{cases} \tag{217}$$

Therefore, we define the random variables

$$Z_i := Z_i^{(\eta)} = \begin{cases} (1 - 2\eta)\lambda/\mathfrak{c}_\zeta(X_i) & \text{if } X_i \notin \mathcal{X}(\eta) \\ 0 & \text{if } X_i \in \mathcal{X}(\eta) \end{cases} \tag{218}$$

Since these are iid (depending only on iid X_i) let Z be the generic version. We know from (185) that

$$\lambda' \geqslant \lambda - 1/k \geqslant (1 - \beta n^{-(1-\frac{1}{\gamma})})\lambda \tag{219}$$

$$\implies (1 - \eta)\lambda' \geqslant (1 - 2\eta)\lambda \quad \text{for} \quad n \geqslant (\beta/\eta)^{1 + \frac{1}{\gamma - 1}} \tag{220}$$

Since $\ell_{j(i)}$ is a cost-λ' trajectory passing through X_i, we know (fixing η) that for any sufficiently large n

$$\ell_{j(i)} \geqslant Z_i \text{ for all } i \tag{221}$$

Additionally, since $\mathbb{P}_{X \sim f}[X \in \mathcal{X}(\eta)] \leqslant \eta$ and because $\mathbb{E}[Y] = \mathbb{E}_{X \sim f}[\mathcal{c}_\zeta(X)^{-1}]$ and $\mathbb{E}[Y^2] = \mathbb{E}_{X \sim f}[\mathcal{c}_\zeta(X)^{-2}]$ are finite, we have

$$\lim_{\eta \to 0} \mathbb{E}[Z^{(\eta)}] = \lambda \mathbb{E}[Y] \quad \text{and} \quad \lim_{\eta \to 0} \mathrm{Var}[Z^{(\eta)}] = \lambda \mathrm{Var}[Y] \tag{222}$$

Therefore, for any $\delta_1 > 0$ we can select some $\eta > 0$ such that

$$\mathbb{E}[Z^{(\eta)}] \geqslant (1 - \delta_1/2)\lambda \mathbb{E}[Y] \text{ and } \mathbb{E}[Z^{(\eta)}] \leqslant (1 + \delta_1)\lambda \mathrm{Var}[Y] \tag{223}$$

Fixing such η, we apply Proposition 9 (noting that Z_i's are nonnegative iid) to get

$$\mathbb{P}\Big[\sum_{i=1}^{n} Z_i \leqslant (1 - \delta_1/2)n\mathbb{E}[Z]\Big] \leqslant e^{-\frac{n\mathbb{E}[Z]^2(\delta_1/2)^2/2}{\mathrm{Var}[Z]+\mathbb{E}[Z]^2(\delta_1/2)/3}} \tag{224}$$

$$\implies \mathbb{P}\Big[\sum_{i=1}^{n} Z_i \leqslant (1 - \delta_1)\lambda n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y]+(1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{225}$$

$$\implies \mathbb{P}\Big[\sum_{i=1}^{n} \ell_{j(i)} \leqslant (1 - \delta_1)\lambda n\mathbb{E}[Y]\Big] \leqslant e^{-\frac{n(1-\delta_1)\mathbb{E}[Y]^2\delta_1^2/8}{(1+\delta_1)\mathrm{Var}[Y]+(1-\delta_1)\mathbb{E}[Y]^2\delta_1/6}} \tag{226}$$

for all $\lambda < \lambda_\eta^*$, which holds since we can apply

$$\mathbb{E}[Z]^2 \geqslant (1 - \delta_1/2)^2\lambda \mathbb{E}[Y] \geqslant (1 - \delta_1)\lambda \mathbb{E}[Y] \quad \text{and} \quad \mathrm{Var}[Z] \leqslant (1 + \delta_1)\mathrm{Var}[Y] \tag{227}$$

(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). $\qquad \square$

But this means that with very high probability, given the CBO bound and applying (189) (and using $1 + \delta_1 < (1 + \delta_1/2)^2$)

$$\sum_{i=1}^{n} \ell_{j(i)} > (1 - \delta_1/2)\lambda n\mathbb{E}[Y] \tag{228}$$

$$\implies \ell(\pi) \geqslant \frac{1 - \delta_1/2}{\beta \lambda n^{\frac{1}{\gamma}}} \sum_{i=1}^{n} \ell_{j(i)} > (1 - \delta_1)\beta^{-1}n^{1-\frac{1}{\gamma}}\mathbb{E}[Y] \tag{229}$$

with very high probability. Specifically (taking into account the probability that the CBO bound holds) we have

$$\mathbb{P}\big[\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1 - \delta_1)\beta^{-1}n^{1-\frac{1}{\gamma}}\mathbb{E}[Y]\big] \tag{230}$$

$$\geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta_1/2}} - e^{-\frac{n(1-\delta_1/2)\mathbb{E}[Y]^2\delta_1^2/32}{(1+\delta_1/2)\mathrm{Var}[Y]+(1-\delta_1/2)\mathbb{E}[Y]^2\delta_1/12}} \tag{231}$$

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):

$$\frac{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}{(1+\frac{\delta_1}{2})\mathrm{Var}[Y]+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}} = \frac{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}{(1+\frac{\delta_1}{2})(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2)+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}} \tag{232}$$

$$= \left(\frac{(1+\frac{\delta_1}{2})(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2)+(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1}{12}}{(1-\frac{\delta_1}{2})\mathbb{E}[Y]^2\frac{\delta_1^2}{32}}\right)^{-1} \tag{233}$$

$$= \left(\frac{1+\frac{\delta_1}{2}}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} + \frac{(1-\frac{\delta_1}{2})\frac{\delta_1}{12}-(1+\frac{\delta_1}{2})}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\right)^{-1} \tag{234}$$

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

$$\left(\frac{1+\frac{\delta_1}{2}}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} + \frac{(1-\frac{\delta_1}{2})\frac{\delta_1}{12}-(1+\frac{\delta_1}{2})}{(1-\frac{\delta_1}{2})\frac{\delta_1^2}{32}}\right)^{-1} = \frac{\delta_1^2}{32(\mathbb{E}[Y^2]/\mathbb{E}[Y]^2-1)} + o(\delta_1^3) \tag{235}$$

$$= \frac{\delta_1^2}{32(\mathrm{Var}[Y]/\mathbb{E}[Y]^2)} + o(\delta_1^3) \tag{236}$$

Then, given a desired approximation factor $1-\delta$, we choose $\delta = \delta_1/2$ and a regularization term $\zeta > 0$ producing

$$\mathbb{E}[Y] \geqslant (1-\delta_2)\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \tag{237}$$

where $\delta_2 = 1-(1-\delta)/(1-\delta/2)$, i.e. such that

$$1-\delta = (1-\delta/2)(1-\delta_2) = (1-\delta_1)(1-\delta_2) \tag{238}$$

Thus, our bound becomes

$$(1-\delta_1)\beta^{-1}n^{1-\frac{1}{\gamma}}\mathbb{E}[Y] = (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx \tag{239}$$

and substituting $\delta_1 = \delta/2$ into the very-high-probability bound yields

$$\mathbb{P}\left[\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{240}$$

$$\geqslant 1-e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/4}} - e^{-\frac{n(1-\delta/4)\mathbb{E}[Y]^2\delta^2/128}{(1+\delta/4)\mathrm{Var}[Y]+(1-\delta/4)\mathbb{E}[Y]^2\delta/24}} \tag{241}$$

$$\geqslant 1-e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} - e^{-\left(\frac{\delta^2}{128(\mathrm{Var}[Y]/\mathbb{E}[Y]^2)}+o(\delta^3)\right)n} \tag{242}$$

when n is sufficiently large. When $\gamma \geqslant 2$, this can be made to be arbitrarily close to

$$\mathbb{P}\left[\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \tag{243}$$

$$\geqslant 1 - e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}} - e^{-\frac{n(1-\delta/2)\mathbb{E}[Y^*]^2\delta^2/32}{(1+\delta/2)\mathrm{Var}[Y^*]+(1-\delta/2)\mathbb{E}[Y^*]^2\delta/12}} \tag{244}$$

$$\geqslant 1 - e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} - e^{-\left(\frac{\delta^2}{32(\mathrm{Var}[Y^*]/\mathbb{E}[Y^*]^2)}+o(\delta^3)\right)n} \tag{245}$$

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 $n \geqslant n^{(\delta)}$,

$$\mathbb{P}\left[\mathrm{TSP}_\Pi(\{X_i\}) \geqslant (1-\delta)\beta^{-1}n^{1-\frac{1}{\gamma}}\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right] \geqslant 1 - (1+\delta)e^{-\frac{4}{5}\frac{\log(b)\lambda n^{\frac{1}{\gamma}}}{1-\delta/2}} \tag{246}$$

$$\geqslant 1 - (1+\delta)e^{-\frac{4}{5}\log(b)\lambda n^{\frac{1}{\gamma}}} \tag{247}$$

with the caveat that as $\delta \to 0$, $n^{(\delta)}$ potentially goes to ∞.

VI. DSTSP UPPER BOUND FOR SYMMETRIC SYSTEMS

As in [18], we show an upper bound via an algorithm which, with very high probability, creates a trajectory $\pi \in \Pi$ which successfully visits the targets $X_1, \ldots, X_n$ with a trajectory of length at most $\Theta(n^{1-\frac{1}{\gamma}})$. However, while the $\Theta(n^{1-\frac{1}{\gamma}})$ growth rate holds in general, the more precise growth rate of

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \approx cn^{1-\frac{1}{\gamma}}\left(\int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}}g(x)^{-\frac{1}{\gamma}}\,dx\right) \qquad wvhp \tag{248}$$

does not hold in general for non-symmetric systems. Hence, in this section we deal exclusively with the case where Π is symmetric.

A. *The Hierarchical Collection Problem*

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 system 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

$$C(v_{(\boldsymbol{q},\varepsilon)}) = \{v_{(\boldsymbol{q}^{(1)},\varepsilon/2)}, \ldots, v_{(\boldsymbol{q}^{(b)},\varepsilon/2)}\} \quad \text{satisfying} \quad R_\varepsilon(\boldsymbol{q}) \subseteq \bigcup_{j=1}^b R_{\varepsilon/2}(\boldsymbol{q}^{(j)}) \tag{249}$$

as given in Assumption 8. We also assume WLOG that

$$R_\varepsilon(\boldsymbol{q}) \cap R_{\varepsilon/2}(\boldsymbol{q}^{(j)}) \neq \varnothing \quad \text{for all} \ \ j \tag{250}$$

(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

$$T_k := \{v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0, \varepsilon_0) : \varepsilon = \varepsilon_0/2^k\}. \tag{251}$$

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 18. *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),*

$$d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant 3\varepsilon/2. \tag{252}$$

iii. For any $v_{(\boldsymbol{q},\varepsilon)} \in T(\boldsymbol{q}_0,\varepsilon_0)$,

$$d_\Pi(\boldsymbol{q}_0,\boldsymbol{q}) \leqslant 3\varepsilon_0 \tag{253}$$

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 (252)), 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

$$d_\Pi(\boldsymbol{q},\boldsymbol{q}') \leqslant d_\Pi(\boldsymbol{q},\boldsymbol{q}'') + d_\Pi(\boldsymbol{q}',\boldsymbol{q}'') \leqslant 3\varepsilon/2 \tag{254}$$

(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

$$d_\Pi(\boldsymbol{q}_0,\boldsymbol{q}_k) \leqslant \sum_{j=1}^{k} d_\Pi(\boldsymbol{q}_{j-1},\boldsymbol{q}_j) \leqslant \sum_{j=1}^{k} 3\varepsilon_0 2^{-j} \leqslant 3\varepsilon_0 \tag{255}$$

Hence we are done. $\qquad\square$

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

$$n(v) := |\{i : v_k(X_i) = v\}| \tag{256}$$

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 system 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 15. *The* hierarchical collection problem (HCP) *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 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*

$$w(e) = s^{-\min(k(v), k(v'))} \tag{257}$$

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

$$\mathrm{HCP}^*(n; \bar{b}, s) := \max_{\tau_1, \ldots, \tau_n} \mathrm{HCP}(\tau_1, \ldots, \tau_n; \bar{b}, s). \tag{258}$$

The targets $\tau_1, \ldots, \tau_n$ in the HCP 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 10. *The HCP 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 10. *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 7) 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 7).

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:

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant \varepsilon_0 \mathrm{HCP}(\tau_1, \ldots, \tau_n; b, s) \leqslant \varepsilon_0 \mathrm{HCP}^*(n; b, s). \tag{259}$$

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 HCP 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 HCP, we are at $\boldsymbol{q}(v) \in \mathcal{Q}$ in the TSP. If $v' \in C(v)$, then

$$d_\pi(\boldsymbol{q}(v), \boldsymbol{q}(v')) \leqslant \varepsilon(v) = \varepsilon_0 s^{-k(v)} \tag{260}$$

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,

$$v \in \tau_i \implies X_i \in S(\boldsymbol{q}(v), \varepsilon(v)) \implies d_\pi(\boldsymbol{q}(v), X_i) \leqslant \varepsilon_0 s^{-k(v)} \tag{261}$$

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 HCP. Thus every solution to the HCP 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 $\mathrm{HCP}^*(n; \bar{b}, s)$. $\qquad\square$

We now define a few notions which will help us discuss the HCP.

Definition 16. *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 HCP 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. *$c_{\mathrm{cost}}(\theta') = c_{\mathrm{cost}}(\theta)$ 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 HCP in general but is made for convenience, particularly for Lemma 20. We now consider two questions: (i) what is the optimal algorithm for the HCP?; (ii) what is the upper bound $\mathrm{HCP}^*(n; b, s)$?

Lemma 19. *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. $\qquad\square$

Lemma 20. *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

$$2s^{-(k+k')}n_v + 2\sum_{j=0}^{k'-1} s^{-(k+j)} \tag{262}$$

  (collection costs, minimized if they are maximally deep, and minimum possible movement costs to reach level $k + k'$ and return).

But then

$$n_v < \frac{s}{s-1} \implies n_v < 1 + \frac{n_v}{s} \tag{263}$$

and we can then collapse (262) because:

$$2s^{-(k+k')}n_v + 2\sum_{j=0}^{k'-1} s^{-(k+j)} = 2s^{-k}\left(s^{-k'}n_v + \sum_{j=0}^{k'-1} s^{-j}\right) \tag{264}$$

$$> 2s^{-k}\left(s^{-(k'-1)}n_v + \sum_{j=0}^{k'-2} s^{-j}\right) \tag{265}$$

$$> \ldots > 2s^{-k}n_v \tag{266}$$

where each step takes the last entry in the sum and the n_v term and applies (263). 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

$$2s^{-k(v')} = 2s^{-(k-k')} \tag{267}$$

Alternatively, we can consider plan θ' which instead of collecting at v' go to v and collect there. This incurs cost

$$2s^{-k(v)}n_v + 2\sum_{j=0}^{k'-1} s^{-(k(v')+j)} = 2s^{-k}n_v + 2\sum_{j=0}^{k'-1} s^{-(k-k'+j)} \tag{268}$$

$$= 2s^{-k}\left(n_v + \sum_{j=1}^{k'} s^{-j}\right) \tag{269}$$

$$\leqslant 2s^{-k}s^{k'} = 2s^{-(k-k')} \tag{270}$$

where the sum collapses because at each step we apply

$$n_v \geqslant \frac{s}{s-1} \implies s\,n_v \geqslant s + n_v \tag{271}$$

Hence, as before, the modified plan θ' costs at most as much as θ (and visits more vertices) so θ cannot be optimal. $\qquad\square$

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

$$V_{\geqslant \frac{s}{s-1}} := \left\{v \in V : n_v \geqslant \frac{s}{s-1}\right\} \tag{272}$$

which must include the root and must be connected since

$$n_v \geqslant \frac{s}{s-1} \implies n_{P(v)} \geqslant n_v \geqslant \frac{s}{s-1}. \tag{273}$$

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 11. *For any $n, \bar{b}, s$ such that $s \geqslant 2$ and $\gamma = \log_s(\bar{b}) \geqslant 2$,*

$$\mathrm{HCP}^*(n; \bar{b}, s) \leqslant 6sn^{1-\frac{1}{\gamma}} \tag{274}$$

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

$$2 \sum_{k=0}^{k^*-1} \bar{b}(\bar{b}/s)^k = 2\bar{b}\frac{(\bar{b}/s)^{k^*} - 1}{(\bar{b}/s) - 1} = 2\frac{(\bar{b}/s)^{k^*} - 1}{(1/s) - (1/\bar{b})} \tag{275}$$

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

$$(\bar{b}/s)^{k^*} \leqslant (\bar{b}/s)^{\log_{\bar{b}} n} = ns^{-\log_{\bar{b}} n} = n^{1-\frac{1}{\gamma}} \tag{276}$$

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

$$2\frac{(\bar{b}/s)^{k^*} - 1}{(1/s) - (1/\bar{b})} \leqslant 4sn^{1-\frac{1}{\gamma}}. \tag{277}$$

Our collection cost is additionally

$$2ns^{-k^*} \leqslant 2sns^{-\log_{\bar{b}} n} = 2sn^{1-\frac{1}{\gamma}}. \tag{278}$$

Adding these together gives a total cost of $\leqslant 6sn^{1-\frac{1}{\gamma}}$. $\qquad \square$

Remark 11. *While the abstraction provided by the HCP is useful for providing an upper bound and an algorithm for the Symmetric DTSP, 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 10, 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 HCP 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.*

B. Hierarchical Collection Problems and the TSP

We now consider the following related 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

$$\nu := 1 - \frac{1}{\gamma} \in (0, 1) \quad \text{and} \quad Y_j := n_j^\nu \text{ and } Y = \sum_{j=1}^m Y_j \tag{279}$$

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 11, the time it takes to visit all the n_j targets in S'_j is at most

$$6s\varepsilon_0 n^{1-\frac{1}{\gamma}} = 6s\varepsilon_0 n_j^\nu = 6s\varepsilon_0 Y_j \tag{280}$$

To visit all n targets thus takes at most

$$C + \sum_{j=1}^{m} 6s\varepsilon_0 Y_j = C + 6s\varepsilon_0 Y \tag{281}$$

where C is a constant (in n) which depends on the choices of the S_j, representing the time it takes for the system 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 21. $\mathbb{E}[Y_j] \leqslant (p_j n)^\nu$ 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,

$$\mathbb{E}[Y_j] = \mathbb{E}[n_j^\nu] \leqslant (\mathbb{E}[n_j])^\nu = (p_j n)^\nu \tag{282}$$

and the result from $\mathbb{E}[Y]$ follows by summing the $\mathbb{E}[Y_j]$. $\qquad\square$

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:

$$Y_j(i) = \mathbb{E}[Y_j \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y_j] \quad \text{and} \quad Y(i) = \sum_{j=1}^{m} Y_j(i) = \mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y] \tag{283}$$

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.

$$\mathbb{E}[Y(i) \mid Y(i-1)] = Y(i-1) \tag{284}$$

(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.

$$n_j(i) = \sum_{i'=1}^{i} \mathbf{1}\{Z_{i'} = j\}. \tag{285}$$

We then show a few results concerning $Y(i)$:

Lemma 22. $|Y(i) - Y(i-1)| \leqslant 1$.

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 $Z_i' = j'$.

$$|Y(Z^{(n)}) - Y(Z_{-i}^{(n)}(Z_i'))| = |n_j^\nu + n_{j'}^\nu - (n_j - 1)^\nu - (n_{j'} + 1)^\nu| \tag{286}$$

$$= |(n_j^\nu - (n_j - 1)^\nu) - ((n_{j'} + 1)^\nu - n_{j'}^\nu)| \leqslant 1 \tag{287}$$

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',

$$|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']| = \mathbb{E}_{Z_{i+1}, \ldots, Z_n}[|Y(Z^{(n)}) - Y(Z_{-i}^{(n)}(Z_i'))|] \tag{288}$$

$$\leqslant 1 \tag{289}$$

But then we get that

$$|Y(i) - Y(i-1)| = |\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}]| \tag{290}$$

$$= |\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}_{Z_i'}[\mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']]| \tag{291}$$

$$\leqslant \mathbb{E}_{Z_i'}[|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z_i']|] \tag{292}$$

$$\leqslant 1 \tag{293}$$

$\square$

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 23. *Given the definitions above,*

$$\mathbb{E}\left[Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{2n}} = e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^m p_j^\nu)^2} \tag{294}$$

Furthermore, $\sum_{j=1}^m p_j^\nu \geqslant 1$, so we can even find a bound independent of $\boldsymbol{p} = (p_1, \ldots, p_m)$:

$$\mathbb{E}\left[Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}} \tag{295}$$

Proof. This is a direct application of the Azuma-Hoeffding Inequality with the bounded differences as shown in Lemma 22.

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 $\boldsymbol{1}_1, \ldots, \boldsymbol{1}_m$. Then we note that the function $\theta(\boldsymbol{p}) = \sum_{j=1}^m p_j^\nu$ is concave and $\theta(\boldsymbol{1}_j) = 1$ for all j, and hence by Jensen's Inequality

$$\theta(\boldsymbol{p}) \geqslant \sum_{j=1}^m p_j \theta(\boldsymbol{1}_j) = 1 \tag{296}$$

and thus we have the result we wanted. $\qquad\square$

Proposition 12. *Given $Y = \sum_{j=1}^m n_j^\nu$, we have*

$$\mathbb{P}\left[Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^m p_j^\nu)^2} \tag{297}$$

which implies the following bound independent of $\boldsymbol{p}$:

$$\mathbb{P}\left[Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}} \tag{298}$$

Proof. Using the definition of the centered martingale $Y(i)$, we have $Y(n) = Y - \mathbb{E}[Y]$, thus giving us:

$$Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu \implies Y - n^\nu \sum_{j=1}^m p_j^\nu \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{299}$$

$$\implies Y - \mathbb{E}[Y] \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{300}$$

$$\implies Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu \tag{301}$$

which in turn means by Lemma 23

$$\mathbb{P}\left[Y \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant \mathbb{P}\left[Y(n) \geqslant n^\nu \sum_{j=1}^m p_j^\nu\right] \leqslant e^{-\frac{1}{2}n^{2\nu-1}(\sum_{j=1}^m p_j^\nu)^2} \tag{302}$$

The second bound independent of p follows as in the proof of Lemma 23. $\qquad\square$

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 22: 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 24. *Let* $W \sim \mathrm{Bin}(p, n)$ *for some* $p \in (0, 1)$ *and* n, *and* $\nu \in (0, 1)$; *then*

$$\mathbb{E}[(W + 1)^\nu] - \mathbb{E}[W^\nu] \leqslant e^{-\frac{3}{28}np} + 2\nu(np)^{\nu-1} \tag{303}$$

Proof. Since $\nu \in (0, 1)$, $\cdot^\nu$ is concave, and so the first-order approximation is an upper bound, yielding

$$(W + 1)^\nu - W^\nu \leqslant W^\nu + \nu W^{\nu-1} - W^\nu = \nu W^{\nu-1} \tag{304}$$

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

$$(W + 1)^\nu - W^\nu \leqslant \min(\nu W^{\nu-1}, 1) \tag{305}$$

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

$$\mathbb{E}[(W + 1)^\nu] - \mathbb{E}[W^\nu] = \mathbb{E}[(W + 1)^\nu - W^\nu] \tag{306}$$

$$\leqslant \mathbb{P}[W \leqslant np/2] + \mathbb{P}[W > np/2](\nu(np/2)^{\nu-1}) \tag{307}$$

$$\leqslant \mathbb{P}[W \leqslant np/2] + 2\nu(np)^{\nu-1} \tag{308}$$

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

$$W_i' = p - W_i \quad \text{(which also means } W_i = p - W_i'\text{)} \tag{309}$$

which are zero-mean and $|W_i'| \leqslant 1$ always. Thus,

$$W = \sum_{i=1}^{n} W_i = np - \sum_{i=1}^{n} W_i' \tag{310}$$

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

$$\mathbb{P}\left[\sum_{i=1}^{n} W_i' \geqslant \frac{np}{2}\right] \leqslant e^{-\frac{(np)^2/8}{np(1-p)+np/6}} = e^{-\frac{np/8}{(1-p)+1/6}} \leqslant e^{-\frac{3}{28}np} \tag{311}$$

and hence we can put it all together and get

$$\mathbb{E}[(W+1)^\nu] - \mathbb{E}[W^\nu] \leqslant e^{-\frac{3}{28}np} + 2\nu(np)^{\nu-1} \tag{312}$$

$\square$

Lemma 25. *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*

$$|Y(i) - Y(i-1)| \leqslant \min(e^{-\frac{3}{28}(n-i)p_1} + 2\nu((n-i)p_1)^{\nu-1}, 1) \tag{313}$$

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

$$\left|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}]\right| \tag{314}$$

$$= \left|\mathbb{E}[Y \mid Z_1, \ldots, Z_i] - \mathbb{E}_{Z \sim \boldsymbol{p}}[\mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z]]\right| \tag{315}$$

$$\leqslant \max_{Z \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z' \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z'] \tag{316}$$

(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

$$\max_{Z_1, \ldots, Z_{i-1}} (\max_{Z \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z' \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z']) \tag{317}$$

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 $W_j \sim \text{Bin}(p_j, n - i)$, we have :

$$\max_{Z_1, \ldots, Z_{i-1}} \left(\max_{Z \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z] - \min_{Z' \in [m]} \mathbb{E}[Y \mid Z_1, \ldots, Z_{i-1}, Z'] \right) \tag{318}$$

$$= \max_{n_j(i), n_{j'}(i)} \left(\mathbb{E}[(W_j + n_j(i) + 1)^\nu] - \mathbb{E}[(W_j + n_j(n))^\nu] \right) \tag{319}$$

$$- \left(\mathbb{E}[(W_{j'} + n_{j'}(i) + 1)^\nu] - \mathbb{E}[(W_{j'} + n_{j'}(n))^\nu] \right) \tag{320}$$

$$\leqslant \max_{n_j(i)} \mathbb{E}[(W_j + n_j(i) + 1)^\nu] - \mathbb{E}[(W_j + n_j(n))^\nu] \tag{321}$$

$$= \max_j \mathbb{E}[(W_j + 1)^\nu] - \mathbb{E}[W_j^\nu] \tag{322}$$

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 24 (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. $\qquad\square$

We now adjust the Azuma-Hoeffding argument for the new difference bound:

Proposition 13. *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*

$$n \geqslant \frac{280}{3} \log(1/\nu) p_1^{-1} \tag{323}$$

we have the following concentration bounds:

$$\mathbb{E}\left[Y(n) \geqslant 2n^\nu \sum_{j=1}^m p_j^\nu \right] \leqslant \begin{cases} e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{13 + \frac{8}{2\nu-1} n^{2\nu-1}}} & \text{for } \nu \in (1/2, 2/3) \\[2em] e^{-\frac{(2/9)p_1 n(\sum_{j=1}^m p_j^{1/2})^2}{127 - \log(1/p_1) + \log(n)}} & \text{for } \nu = 1/2 \\[2em] e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^m p_j^\nu)^2}{\frac{560}{3}\log(1/\nu) + 4 + 18\nu^2 \frac{1}{1-2\nu}(\frac{280}{3}\log(1/\nu))^{2\nu-1}}} & \text{for } \nu \in (0, 1/2) \end{cases} \tag{324}$$

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

$$\mathbb{E}\left[Y(n) \geqslant n^\nu \sum_{j=1}^{m} p_j^\nu\right] \leqslant \begin{cases} e^{-\Theta(p_1 n(2\nu-1)(\sum_{j=1}^{m} p_j^\nu)^2)} & \textit{for } \nu \in (1/2, 2/3) \\ e^{-\Theta(p_1 (n/\log(n))(\sum_{j=1}^{m} p_j^\nu)^2)} & \textit{for } \nu = 1/2 \\ e^{-\Theta(p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2(\log(1/\nu)+(1-2\nu)^{-1})^{-1})} & \textit{for } \nu \in (0, 1/2) \end{cases} \tag{325}$$

Proof. Let

$$c_i = \min(e^{-\frac{3}{28}(n-i)p_1} + 2\nu((n-i)p_1)^{\nu-1}, 1) \tag{326}$$

We now want to compute $\sum_{i=1}^{n} c_i^2$. We first define

$$a_i = c_{n-i} = \min(e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1}, 1) \tag{327}$$

where i now goes from 0 to $n-1$. Then we know that

$$\sum_{i=0}^{n-1} a_i^2 = \sum_{i=1}^{n} c_i^2 \tag{328}$$

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

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \text{ for all } z \geqslant z^* \tag{329}$$

We compute:

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \tag{330}$$

$$\iff -\frac{3}{28}z \leqslant (\nu-1)\log(z) + \log(2\nu) \tag{331}$$

$$\iff \frac{3}{28}z \geqslant (1-\nu)\log(z) - \log(2\nu) \tag{332}$$

Then note that if $z \geqslant \frac{28}{3}\frac{1}{1-\nu}$,

$$\frac{d}{dz}\frac{3}{28}z = \frac{3}{28} \geqslant (1-\nu)\frac{1}{z} = \frac{d}{dz}(1-\nu)\log(z) - \log(2\nu) \tag{333}$$

Thus, if $z^* \geqslant \frac{28}{3}\frac{1}{1-\nu}$ and

$$\frac{3}{28}z^* \geqslant (1-\nu)\log(z^*) - \log(2\nu) \tag{334}$$

then we can conclude that for all $z \geqslant z^*$,

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \,. \tag{335}$$

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

$$\frac{3}{28}z^* = 10\log(1/\nu) \geqslant (1-\nu)\log\left(\frac{280}{3}\log(1/\nu)\right) - \log(2\nu) \tag{336}$$

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

$$e^{-\frac{3}{28}z} \leqslant 2\nu z^{\nu-1} \,. \tag{337}$$

Then, recalling that $z = ip_1$, we have for all $i \geqslant z^*/p_1 = \frac{280}{3}\log(1/\nu)p_1^{-1}$,

$$e^{-\frac{3}{28}ip_1} \leqslant 2\nu(i\,p_1)^{\nu-1} \implies e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1} \leqslant 3\nu(i\,p_1)^{\nu-1} \tag{338}$$

Thus, we have the following:

$$a_i = \min(e^{-\frac{3}{28}ip_1} + 2\nu(i\,p_1)^{\nu-1}, 1) \leqslant \begin{cases} 1 & \text{if } i \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} \\ 3\nu(i\,p_1)^{\nu-1} & \text{if } i \geqslant \frac{280}{3}\log(1/\nu)p_1^{-1} \end{cases} \tag{339}$$

We now assume that $n \geqslant \frac{280}{3}\log(1/\nu)p_1^{-1}$. In that case,

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + \sum_{i=\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} 9\nu^2(i\,p_1)^{2\nu-2} \tag{340}$$

$$\leqslant \frac{280}{3}\log(1/\nu)p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2} \int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\, dt \tag{341}$$

We now have to split this according to $\nu > 1/2$, $\nu = 1/2$, and $\nu < 1/2$. This yields the following: for $\nu \in (1/2, 2/3)$,

$$9\nu^2 p_1^{2\nu-2} \int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2}\, dt \tag{342}$$

$$= 9\nu^2 p_1^{2\nu-2}\frac{1}{2\nu-1}\left((n-1)^{2\nu-1} - \left(\frac{280}{3}\log(1/\nu)p_1^{-1}\right)^{2\nu-1}\right) \tag{343}$$

$$\leqslant 4p_1^{-1}\frac{1}{2\nu-1}\left(n^{2\nu-1} - \left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1}\right) \tag{344}$$

and hence we can conclude that (again, for $\nu \in (1/2, 2/3)$)

$$\sum_{i=0}^{n-1} a_i^2 \leqslant p_1^{-1}\left(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}\left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1}\right) + 2 + \frac{4}{2\nu-1}p_1^{-1}n^{2\nu-1} \tag{345}$$

$$= O(p_1^{-1}\log(1/\nu) + 1) + O(p_1^{-1}(2\nu-1)^{-1})n^{2\nu-1} \tag{346}$$

Then, for $\nu = 1/2$, we have

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3} \log(1/\nu) p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2} \int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2} \, dt \tag{347}$$

$$= \frac{280}{3} \log(2) p_1^{-1} + 2 + (9/4) p_1^{-1} \left(\log(n-1) - \log\left(\frac{280}{3}\log(2)\right) + \log(1/p_1) \right) \tag{348}$$

$$\leqslant \left(\frac{280}{3} \log(2) - (9/4) \log\left(\frac{280}{3}\log(2)\right) \right) p_1^{-1} + 2 + (9/4) p_1^{-1} \left(\log(n) - \log(1/p_1) \right) \tag{349}$$

$$< (56 - (9/4)\log(1/p_1)) p_1^{-1} + 2 + (9/4) p_1^{-1} \log(n) \tag{350}$$

And finally, for $\nu \in (0, 1/2)$, we have

$$\sum_{i=0}^{n-1} a_i^2 \leqslant \frac{280}{3} \log(1/\nu) p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2} \int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{n-1} t^{2\nu-2} \, dt \tag{351}$$

$$\leqslant \frac{280}{3} \log(1/\nu) p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2} \int_{\frac{280}{3}\log(1/\nu)p_1^{-1}}^{\infty} t^{2\nu-2} \, dt \tag{352}$$

$$= \frac{280}{3} \log(1/\nu) p_1^{-1} + 2 + 9\nu^2 p_1^{2\nu-2} \frac{1}{1-2\nu} \left(\frac{280}{3}\log(1/\nu)p_1^{-1} \right)^{2\nu-1} \tag{353}$$

$$= \left(\frac{280}{3} \log(1/\nu) + 2 + 9\nu^2 \frac{1}{1-2\nu} \left(\frac{280}{3}\log(1/\nu) \right)^{2\nu-1} \right) p_1^{-1} \tag{354}$$

(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 $\nu \in (1/2, 2/3)$,

$$\mathbb{E}\left[Y(n) \geqslant 2n^\nu \sum_{j=1}^{m} p_j^\nu \right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2}{2\sum_{i=1}^{n} a_i^2}} \tag{355}$$

$$\leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2}{2p_1^{-1}\left(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}\left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1} + 4 + \frac{8}{2\nu-1}p_1^{-1}n^{2\nu-1}\right)}} \tag{356}$$

$$= e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2}{2\left(\frac{280}{3}\log(1/\nu) - \frac{4}{2\nu-1}\left(\frac{280}{3}\log(1/\nu)\right)^{2\nu-1} + 4p_1 + \frac{8}{2\nu-1}n^{2\nu-1}\right)}} \tag{357}$$

$$\leqslant e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2}{9 + 4p_1 + \frac{8}{2\nu-1}n^{2\nu-1}}} \tag{358}$$

$$\leqslant e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^\nu)^2}{13 + \frac{8}{2\nu-1}n^{2\nu-1}}} \tag{359}$$

$$= e^{-\Theta(p_1 n(2\nu-1)(\sum_{j=1}^{m} p_j^\nu)^2)} \tag{360}$$

where the second-to-last step is simply from numerically taking the maximum of

$$\frac{280}{3} \log(1/\nu) - \frac{4}{2\nu-1} \left(\frac{280}{3}\log(1/\nu) \right)^{2\nu-1} \tag{361}$$

over the domain $\nu \in (1/2, 2/3)$.

For $\nu = 1/2$, we have

$$\mathbb{E}\left[Y(n) \geqslant 2n^{1/2} \sum_{j=1}^{m} p_j^{1/2}\right] \leqslant e^{-\frac{n(\sum_{j=1}^{m} p_j^{1/2})^2}{2 \sum_{i=1}^{n} a_i^2}} \tag{362}$$

$$\leqslant e^{-\frac{n(\sum_{j=1}^{m} p_j^{1/2})^2}{2((56-(9/4)\log(1/p_1))p_1^{-1}+2+(9/4)p_1^{-1}\log(n))}} \tag{363}$$

$$= e^{-\frac{p_1 n(\sum_{j=1}^{m} p_j^{1/2})^2}{(28-(9/2)\log(1/p_1))+4p_1+(9/2)\log(n)}} \tag{364}$$

$$\leqslant e^{-\frac{(2/9)p_1 n(\sum_{j=1}^{m} p_j^{1/2})^2}{(126-\log(1/p_1))+p_1+\log(n)}} \tag{365}$$

$$\leqslant e^{-\frac{(2/9)p_1 n(\sum_{j=1}^{m} p_j^{1/2})^2}{127-\log(1/p_1)+\log(n)}} \tag{366}$$

$$= e^{-\Theta(p_1(n/\log(n))(\sum_{j=1}^{m} p_j^{\nu})^2)} \tag{367}$$

Finally, for $\nu \in (0, 1/2)$ we have

$$\mathbb{E}\left[Y(n) \geqslant 2n^{\nu} \sum_{j=1}^{m} p_j^{\nu}\right] \leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^{m} p_j^{\nu})^2}{2 \sum_{i=1}^{n} a_i^2}} \tag{368}$$

$$\leqslant e^{-\frac{n^{2\nu}(\sum_{j=1}^{m} p_j^{\nu})^2}{(\frac{560}{3}\log(1/\nu)+4+18\nu^2 \frac{1}{1-2\nu}(\frac{280}{3}\log(1/\nu))^{2\nu-1})p_1^{-1}}} \tag{369}$$

$$= e^{-\frac{p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^{\nu})^2}{\frac{560}{3}\log(1/\nu)+4+18\nu^2 \frac{1}{1-2\nu}(\frac{280}{3}\log(1/\nu))^{2\nu-1}}} \tag{370}$$

$$= e^{-\Theta(p_1 n^{2\nu}(\sum_{j=1}^{m} p_j^{\nu})^2(\log(1/\nu)+(1-2\nu)^{-1})^{-1})} \tag{371}$$

$\square$

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 12 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.

C. 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 HCS's that cover $\mathcal{X}_f$ (at scale ε_0), and let $x_j = [\boldsymbol{q}_j]_{\mathcal{X}}$ for all j; we define

$$p_j := \mathbb{P}_{X \sim f}[X \in S(\boldsymbol{q}_j, \varepsilon_0)] \tag{372}$$

Note that p_j depend on which HCS'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 E). 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 HCS covers a certain amount of probability mass, except a negligible fraction:

Lemma 26. *For any $\rho_1 > 0$, let a ρ_1-good cell S_j be defined as one such that*

$$p_j \geqslant (1 - \rho_1)\alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma. \tag{373}$$

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

$$\mathbb{P}_{X \sim f}[S_{j(X)} \text{ is } \rho_1\text{-good}] \geqslant 1 - \rho_2. \tag{374}$$

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

$$\lim_{\zeta \to 0} \mathbb{P}_{X \sim f}[\check{f}^{(\zeta)}(X) \leqslant (3/2)\rho^*] = \mathbb{P}_{X \sim f}[f(X) \leqslant (3/2)\rho^*] \leqslant p_2/2 \tag{375}$$

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

$$X \in S_{j(X)} \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_j) \subseteq \mathcal{B}_{c_\Pi \varepsilon_0}(x_j) \tag{376}$$

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

$$|\check{f}^{(\zeta)}(X) - \check{f}^{(\zeta)}(x_j)| \leqslant \rho^*/2 \tag{377}$$

$$\implies \check{f}^{(\zeta)}(x_j) \geqslant \rho^* \tag{378}$$

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,

$$\mathbb{P}_{f \sim X}[X \in S_j] = \int_{S_j} f(x)\, dx \tag{379}$$

$$\geqslant \int_{S_j} \check{f}^{(\zeta)}(x)\, dx \tag{380}$$

$$\geqslant \int_{S_j} (1 - \rho_1/2)\check{f}^{(\zeta)}(x_j)\, dx \tag{381}$$

$$= (1 - \rho_1/2)\check{f}^{(\zeta)}(x_j)\mathrm{Vol}_{\mathcal{X}}(S_j) \tag{382}$$

$$\geqslant (1 - \rho_1/2)^2 \alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma \tag{383}$$

$$\geqslant (1 - \rho_1)\alpha \check{f}^{(\zeta)}(x_j)\check{g}^{(\zeta)}(x_j)\varepsilon_0^\gamma. \tag{384}$$

Finally, claims (i)-(iii) show our result because

$$\mathbb{P}_{X \sim f}[S_{j(X)} \text{ is } \rho_1\text{-good}] \geqslant \mathbb{P}_{X \sim f}[S_{j(X)} \text{ is } \rho^*\text{-sufficient}] \tag{385}$$

$$\geqslant \mathbb{P}_{X \sim f}[\check{f}^{(\zeta)}(X) > (3/2)\rho^*] \tag{386}$$

$$\geqslant \rho_2 \tag{387}$$

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 E that this assumption can be made without loss of generality, i.e. that the same very high probability bound can be achieved.

Given covering HCS's with roots $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$, we define

$$C_{\varepsilon_0} = \mathrm{TSP}_\Pi(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m) \tag{388}$$

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 system 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 system, 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 12 or Proposition 13 (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

$$\sum_{j=1}^m p_j^{1-\frac{1}{\gamma}} = \mathbb{E}_{Z \sim \boldsymbol{p}}\left[p_Z^{-\frac{1}{\gamma}} \right] \tag{389}$$

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 26 applies and

$$\int_{\mathcal{X}_f} f(x) \check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}} dx \approx \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \tag{390}$$

(this approximation can be made arbitrarily close by setting ζ small) to get

$$\sum_{j=1}^m p_j^{1-\frac{1}{\gamma}} = \mathbb{E}_{Z \sim \boldsymbol{p}}\left[p_Z^{-\frac{1}{\gamma}} \right] \tag{391}$$

$$= \mathbb{E}_{X \sim f}\left[p_{j(X)}^{-\frac{1}{\gamma}} \right] \tag{392}$$

$$\leqslant \mathbb{E}_{X \sim f}\left[(1-\rho_1)^{-\frac{1}{\gamma}} \alpha^{-\frac{1}{\gamma}} \varepsilon_0^{-1} \check{f}^{(\zeta)}(X)^{-\frac{1}{\gamma}} \check{g}^{(\zeta)}(X)^{-\frac{1}{\gamma}} \right] \tag{393}$$

$$= (1-\rho_1)^{-\frac{1}{\gamma}} \alpha^{-\frac{1}{\gamma}} \varepsilon_0^{-1} \int_{\mathcal{X}_f} f(x) \check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}} dx \tag{394}$$

$$\leqslant (1+\delta) \alpha^{-\frac{1}{\gamma}} \varepsilon_0^{-1} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \tag{395}$$

By Proposition 11, 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,

$$\text{TSP}_\Pi(X_1, \ldots, X_n) \leqslant C_{\varepsilon_0} + 6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} \tag{396}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + 12s\varepsilon_0 n^{1-\frac{1}{\gamma}} \sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} \tag{397}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + 12s\varepsilon_0 n^{1-\frac{1}{\gamma}} (1-\rho_1)^{-\frac{1}{\gamma}} \alpha^{-\frac{1}{\gamma}} \varepsilon_0^{-1} \int_{\mathcal{X}_f} f(x) \check{f}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \check{g}^{(\zeta)}(x)^{-\frac{1}{\gamma}} \, dx \tag{398}$$

$$wvhp \quad \leqslant C_{\varepsilon_0} + (1+\delta) 12s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{399}$$

Exactly what 'wvhp' means here is dependent on what regime from Propositions 12 and 13 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 12. *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 $n \to \infty$.

VII. 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 tour length. We refer to this as *adversarial* target placement. While our adversarial target lower bound will hold for non-symmetric dynamics, as in Section VI our focus here is on symmetric Π.

Remark 13. *When* Π *is nonsymmetric, there are counterexamples in which adversarial target placement results in a tour length* $\Theta(n)$; *for instance, if we have a Dubins car with turning radius* 1, *we can distribute our targets evenly over the boundary of a circle with radius* $1/100$. *Then at each 'pass' through the circle, the system can visit at most* 2 *target points, and it takes* $\approx 2\pi$ *time to return, thus yielding a tour of length* $\approx \pi n$.

As discussed in Section III, 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 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 $\mathcal{X}^* \subseteq \mathcal{X}_f$ 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

$$cn^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{400}$$

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

$$\int_{\mathcal{X}*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{401}$$

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.

A. Adversarial randomness

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:

$$\text{maximize} \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \quad \text{(where } f : \mathcal{X}^* \to \mathbb{R} \text{ is integrable)} \tag{402}$$

$$\text{subject to} \int_{\mathcal{X}^*} f(x)dx = 1, \text{ and } f(x) \geqslant 0 \text{ for all } x \in \mathcal{X}^* \tag{403}$$

(i.e. the constraint is that f is a probability density function over $\mathcal{X}^*$).

Lemma 27. *Objective* (402) *under constraints* (403) *is maximized by* $f^*(x) \propto g(x)^{-1}$, *i.e.*

$$f^*(x) := \frac{g(x)^{-1}}{\int_{\mathcal{X}^*} g(y)^{-1} dy}. \tag{404}$$

Proof. First, we note that f^* trivially satisfies the constraints in (403): 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 $\int_{\mathcal{X}^*} g(y)^{-1} dy \leqslant g_{\min}^{-1} \text{Vol}_{\mathcal{X}}(\mathcal{X}^*) < \infty$.

We then note that scaling $g(x)$ by a constant c yields

$$\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} (c\, g(x))^{-\frac{1}{\gamma}} dx = c^{-\frac{1}{\gamma}} \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \tag{405}$$

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 27 using Hölder's Inequality. We first define:

$$\tilde{f}(x) := f(x)^{1-\frac{1}{\gamma}} \text{ and } \tilde{g}(x) = g(x)^{-\frac{1}{\gamma}} \tag{406}$$

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:

$$\int_{\mathcal{X}^*} \tilde{f}(x)\,\tilde{g}(x)dx \leqslant \left(\int_{\mathcal{X}^*} \tilde{f}(x)^{\eta} dx \right)^{\frac{1}{\eta}} \left(\int_{\mathcal{X}^*} \tilde{g}(x)^{\gamma} dx \right)^{\frac{1}{\gamma}}. \tag{407}$$

But, using the definitions from above,

$$\tilde{f}(x)^\eta = \left(f(x)^{1-\frac{1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} = f(x) \text{ and } \tilde{g}(x)^\gamma = \left(g(x)^{-\frac{1}{\gamma}}\right)^\gamma = g(x)^{-1} \tag{408}$$

so we can rewrite the inequality as

$$\int_{\mathcal{X}^*} \tilde{f}(x)\, \tilde{g}(x) dx \leqslant \left(\int_{\mathcal{X}^*} f(x) dx\right)^{\frac{1}{\eta}} \left(\int_{\mathcal{X}^*} g(x)^{-1} dx\right)^{\frac{1}{\gamma}}; \tag{409}$$

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

$$\int_{\mathcal{X}^*} f(x) dx = \int_{\mathcal{X}^*} g(x)^{-1} dx = 1 \tag{410}$$

implying that the right hand side of expression (409) is just 1. Thus,

$$\int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}}\, g(x)^{-\frac{1}{\gamma}} dx = \int_{\mathcal{X}^*} \tilde{f}(x)\, \tilde{g}(x) dx \leqslant 1 \tag{411}$$

for any probability density function f. But, using f^* as defined in Lemma 27, it is trivial to see that

$$\int_{\mathcal{X}^*} f^*(x)^{1-\frac{1}{\gamma}}\, g(x)^{-\frac{1}{\gamma}} dx = 1 \tag{412}$$

thus showing that f^* is the maximizing density function. $\square$

Remark 14. *Density f^* makes the lucrativity function constant over $\mathcal{X}^*$, as for all $x \in \mathcal{X}^*$,*

$$\varrho^*(x) = (f^*(x)g(x))^{\frac{1}{\gamma}} = \left(\int_{\mathcal{X}^*} g(y)^{-1}\, dy\right)^{-\frac{1}{\gamma}}. \tag{413}$$

Lemma 28. *The optimal value of the problem given in (402) and (403) is*

$$\max_f \int_{\mathcal{X}^*} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}}\, dx = \left(\int_{\mathcal{X}^*} g(x)^{-1}\, dx\right)^{-\frac{1}{\gamma}} \tag{414}$$

Proof. Plugging f^* into expression (402) gives

$$\int_{\mathcal{X}^*} f^*(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx = \int_{\mathcal{X}^*} \left(\frac{g(x)^{-1}}{\int_{\mathcal{X}^*} g(y)^{-1} dy}\right)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} dx \tag{415}$$

$$= \int_{\mathcal{X}^*} \frac{g(x)^{-1}}{\left(\int_{\mathcal{X}^*} g(y)^{-1} dy\right)^{1-\frac{1}{\gamma}}} dx \tag{416}$$

$$= \frac{\int_{\mathcal{X}^*} g(x)^{-1} dx}{\left(\int_{\mathcal{X}^*} g(y)^{-1} dy\right)^{1-\frac{1}{\gamma}}} \tag{417}$$

$$= \left(\int_{\mathcal{X}^*} g(x)^{-1} dx\right)^{\frac{1}{\gamma}}. \tag{418}$$

□

Thus, if we distribute the points $X_1, X_2, \ldots, X_n$ according to f^* (iid), by Theorem 1, for any $\delta > 0$

$$\mathrm{TSP}_\Pi(X_1, \ldots, X_n) \leqslant (1 + \delta)\beta^{-1} n^{1 - \frac{1}{\gamma}} \left(\int_{\mathcal{X}^*} g(x)^{-1} dx \right)^{\frac{1}{\gamma}} \quad wvhp \tag{419}$$

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.

B. Adversarial targets upper bound

We now prove the upper bound in Theorem 2. Consider a Hierarchical Cell Structure over $\mathcal{X}^*$ with scaling parameter s and overlap parameter α at scale ε_0, which we know exists. The HCS induces a specific algorithm, which by Proposition 11 generates a tour of length bounded above by

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1 - \frac{1}{\gamma}} + C_{\varepsilon_0} \tag{420}$$

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

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1 - \frac{1}{\gamma}} = 6s\varepsilon_0 n^{1 - \frac{1}{\gamma}} \sum_{j=1}^{m} p_j^{1 - \frac{1}{\gamma}} \tag{421}$$

Note that in contrast with Section VI, 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

$$\sum_{j=1}^{m} p_j^{1 - \frac{1}{\gamma}} = \mathbb{E}_{Z \sim \boldsymbol{p}}[p_Z^{-\frac{1}{\gamma}}] \tag{422}$$

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 is

$$p_j^* = \mathbb{P}_{X \sim f^*}[X \in S_j] = \int_{S_j} f^*(x)\, dx = \frac{\int_{S_j} g(x)^{-1}\, dx}{\int_{\mathcal{X}^*} g(y)^{-1}\, dy} \tag{423}$$

Therefore, we can re-write

$$\mathbb{E}_{Z \sim \boldsymbol{p}^*}[(p_Z^*)^{-\frac{1}{\gamma}}] = \mathbb{E}_{Z \sim \boldsymbol{p}^*}\left[\left(\frac{\int_{S_Z} g(x)^{-1}\,dx}{\int_{\mathcal{X}^*} g(y)^{-1}\,dy}\right)^{-\frac{1}{\gamma}}\right] \tag{424}$$

$$= \left(\int_{\mathcal{X}^*} g(y)^{-1}\,dy\right)^{\frac{1}{\gamma}} \mathbb{E}_{Z \sim \boldsymbol{p}^*}\left[\left(\int_{S_Z} g(x)^{-1}\,dx\right)^{-\frac{1}{\gamma}}\right] \tag{425}$$

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 $\breve{g}^{(\eta)}$ 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 $d_\Pi(\boldsymbol{q}, \boldsymbol{q}') \leqslant \varepsilon_0$,

$$\mathrm{Vol}_{\mathcal{X}}(S_j) \geqslant (1 - \delta_1)\alpha \mathrm{Vol}_{\mathcal{X}}(\bar{R}_{\varepsilon_0}(\boldsymbol{q}_j)) \geqslant (1 - \delta_2)(1 - \delta_1)\alpha \breve{g}^{(\eta)}([\boldsymbol{q}]_{\mathcal{X}})\varepsilon_0^\gamma \tag{426}$$

$$\implies \breve{g}^{(\eta)}(x) \leqslant (1 - \delta_1)^{-1}(1 - \delta_2)^{-1}\alpha^{-1}\varepsilon_0^{-\gamma}\mathrm{Vol}_{\mathcal{X}}(S_j) \tag{427}$$

$$\implies \breve{g}^{(\eta)}(x)^{-1} \geqslant (1 - \delta_1)(1 - \delta_2)\alpha\varepsilon_0^\gamma\mathrm{Vol}_{\mathcal{X}}(S_j)^{-1} \tag{428}$$

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,

$$\int_{S_j} g(x)^{-1}\,dx \geqslant (1 - \delta_3)\int_{S_j} \breve{g}^{(\eta)}(x)^{-1}\,dx \tag{429}$$

But then we can conclude that for any j,

$$\int_{S_j} g(x)^{-1}\,dx \geqslant (1 - \delta_3)\int_{S_j} \breve{g}^{(\eta)}(x)^{-1}\,dx \tag{430}$$

$$\geqslant (1 - \delta_1)(1 - \delta_2)(1 - \delta_3)\alpha\varepsilon_0^\gamma\int_{S_j} \mathrm{Vol}_{\mathcal{X}}(S_j)^{-1}\,dx \tag{431}$$

$$= (1 - \delta_1)(1 - \delta_2)(1 - \delta_3)\alpha\varepsilon_0^\gamma \tag{432}$$

and hence we can flip it around and get

$$\left(\int_{S_j} g(x)^{-1}\,dx\right)^{-\frac{1}{\gamma}} \leqslant (1 - \delta_1)^{-\frac{1}{\gamma}}(1 - \delta_2)^{-\frac{1}{\gamma}}(1 - \delta_3)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}. \tag{433}$$

Since this holds for all j, we know that

$$\mathbb{E}_{Z \sim \boldsymbol{p}^*}\left[\left(\int_{S_Z} g(x)^{-1}\,dx\right)^{-\frac{1}{\gamma}}\right] \leqslant (1 - \delta_1)^{-\frac{1}{\gamma}}(1 - \delta_2)^{-\frac{1}{\gamma}}(1 - \delta_3)^{-\frac{1}{\gamma}}\alpha^{-\frac{1}{\gamma}}\varepsilon_0^{-1}. \tag{434}$$

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 (421) to get

$$6s\varepsilon_0 n^{1-\frac{1}{\gamma}} \sum_{j=1}^{m}(p_j^*)^{1-\frac{1}{\gamma}} \leqslant 6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} \left((1-\delta_1)(1-\delta_2)(1-\delta_3)\right)^{-\frac{1}{\gamma}} \quad (435)$$

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 $n \to \infty$, they $p_j = n_j/n$ 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

$$\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} \leqslant (1+\delta_4) \sum_{j=1}^{m}(p_j^*)^{1-\frac{1}{\gamma}} \quad (436)$$

Then, for any $\delta_5 > 0$, we can set $\delta_1, \delta_2, \delta_3, \delta_4 > 0$ such that

$$1 + \delta_5 = (1-\delta_1)^{-\frac{1}{\gamma}}(1-\delta_2)^{-\frac{1}{\gamma}}(1-\delta_3)^{-\frac{1}{\gamma}}(1+\delta_4) \quad (437)$$

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

$$6s\varepsilon_0 \sum_{j=1}^{m} n_j^{1-\frac{1}{\gamma}} \leqslant (1+\delta_5)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} \quad (438)$$

Thus, for sufficiently large n, the HCS induces an algorithm which collects all the target points in at most $(1+\delta_5)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} + C_{\varepsilon_0}$ length; and for any $\delta > 0$ we can choose $0 < \delta_5 < \delta$ and choose n sufficiently large that

$$(1+\delta_5)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} + C_{\varepsilon_0} \leqslant (1+\delta)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} \quad (439)$$

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

$$\sup_{X_1, \ldots, X_n \in \mathcal{X}*} \text{TSP}_\Pi(X_1, \ldots, X_n) \leqslant (1+\delta)6s\alpha^{-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}*} g(y)^{-1}\, dy \right)^{\frac{1}{\gamma}} \quad (440)$$

and we are done with the upper bound to Theorem 2.

Remark 15. *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*

$$\sum_{j=1}^{m} p_j^{1-\frac{1}{\gamma}} = m^{\frac{1}{\gamma}}. \quad (441)$$

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.

VIII. CONCLUSION

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. ([17], 2005) and Adler, Karaman ([18], 2016), as well as a novel Dynamic TSP algorithm for symmetric dynamically-constrained systems 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 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

$$n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{442}$$

where $\mathcal{X}_f$ is the support of f, and for *symmetric* dynamics a matching upper bound which scales according to (442). 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 systems, we show similar (but deterministic) matching lower and upper bounds which scale according to

$$n^{1-\frac{1}{\gamma}} \left(\int_{\mathcal{X}^*} g(x)^{-1} \, dx \right)^{\frac{1}{\gamma}} \tag{443}$$

with the lower bound also applying to non-symmetric systems.

A. *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, where s (typically $s = 2$ for control-affine systems) and α are parameters measuring the effectiveness of our cell-tiling algorithm, and β 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 system can perform the TSP, or whether it is a product of our specific Hierarchical Cell Structure algorithm.

Theorem 1 almost completely extends the result of Beardwood et al. ([6], 1959) on the Euclidean Stochastic TSP to general symmetric dynamic constraints. In particular, 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

$$\text{TSP}(X_1, \ldots, X_n) \to \beta_d n^{1-\frac{1}{d}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{d}} \, dx \tag{444}$$

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 $X_1, X_2, \ldots \overset{iid}{\sim} f$

$$\lim_{n\to\infty} \frac{\text{TSP}_\Pi(X_1, \ldots, X_n)}{n^{1-\frac{1}{\gamma}} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx} = \beta_\Pi \tag{445}$$

would complete the extension of the Beardwood-Halton-Hammersley Theorem to our setting, 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 F, 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 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 space 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 [17], 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 [16] 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 algorithm achieves a constant factor approximation of the shortest tour with very high probability; this in turn shows that it also achieve a constant factor approximation of the expected length of the shortest tour (see Corollary 1 and Appendix A). Finally, Theorem 2 shows that our 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 longer ones (though when the targets are random this becomes vanishingly unlikely). A guaranteed approximation algorithm (as opposed to an algorithm like the HCP algorithm which finds an approximate solution with very high probability or in expectation) would find a tour whose length is within a constant multiplicative factor of the shortest tour on *any* set of n target points.

References

[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] 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.

[7] 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.

[8] Dimitris Bertsimas, Philippe Chervi, and Michael Peterson, "Computational approaches to stochastic vehicle routing problems," *Transportation Science*, vol. 29, no. 4, pp. 342–352, 1995.

[9] Ludger Rüschendorf Birgit Anthes, "On the weighted euclidean matching problem in r^d," *Applicationes Mathematicae*, vol. 28, no. 2, pp. 181–190, 2001.

[10] 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.

[11] 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.

[12] 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.

[13] Ketan Savla, Emilio Frazzoli, and Francesco Bullo, "On the dubins traveling salesperson problems: Novel approximation algorithms," in *Robotics: Science and Systems*, 2006.

[14] 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.

[15] 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.

[16] Sleiman Itani, *Dynamic systems and subadditive functionals*, Ph.D. thesis, Massachusetts Institute of Technology, 2012.

[17] 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.

[18] 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.

[19] Frédéric Jean, *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning*, SpringerBriefs in Mathematics. Springer International Publishing, 2014.

[20] Patrick Rebeschini, "Algorithmic foundations of learning, lecture 7 notes," September 2020.

APPENDIX A

WITH VERY HIGH PROBABILITY

Recall that a sequence of events $\{A_n\}_{n \in \mathbb{Z}_{\geq 0}}$ happens *with very high probability* if there are $c_1, c_2, c_3 > 0$ such that for all sufficiently large n, $\mathbb{P}[A_n] \geq 1 - c_1 e^{-c_2 n^{c_3}}$. We will explore important implications for this, as well as ways in which it differs from the more common definition of 'with high probability' (in which $\{A_n\}_{n \in \mathbb{Z}_{\geq 0}}$ happens with high probability if $\lim_{n \to 0} \mathbb{P}[A_n] = 1$, without any additional condition on the speed of convergence).

Essentially, wvhp converges to 1 faster than any inverse polynomial, which yields additional information for which 'with high probability' is not sufficient:

Lemma 29. *If $\{A_n\}$ happens with very high probability, then*

$$\mathbb{P}[(\max n : \neg A_n) < \infty] = 1 \tag{446}$$

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

$$\mathbb{P}[(\max n : E_n) < \infty] = 1 \tag{447}$$

But letting $E_n = \neg A_n$, we have

$$\mathbb{P}[E_n] \leq c_1 e^{-c_2 n^{c_3}} \tag{448}$$

for sufficiently large n (say, $n > n'$). Of course, for $n \leq n'$, we have the trivial bound $\mathbb{P}[E_n] \leq 1$. Thus,

$$\sum_{n=1}^{\infty} \mathbb{P}[E_n] \leq n' + \sum_{n > n'} c_1 e^{-c_2 n^{c_3}} < \infty \tag{449}$$

since $c_1 e^{-c_2 n^{c_3}}$ (which decreases faster than any n^{-k}) is a convergent series.

Therefore, $\mathbb{P}[(\max n : \neg A_n) < \infty] = 1$, as we wanted. $\square$

We also define a related notion of a random variable growing "with very high probability":

Definition 17. *If $\{Z_n\}$ is a sequence of random variables and $\alpha(n) : \mathbb{Z}_{>0} \to \mathbb{R}_{\geq 0}$ 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 $A_n = \{Z_n \leq c\alpha(n)\}$ 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 nonnegative sequence δ_n satisfying $\lim_{n\to\infty} \delta_n = 0$ such that

$$B_n = \{1 - \delta_n \leqslant Z_n/\alpha(n) \leqslant 1 + \delta_n\} \tag{450}$$

happens with very high probability.

The following then hold:

Lemma 30. *If $Z_1, Z_2, \ldots$ and $Z_1', Z_2', \ldots$ are sequences of random variables parameterized by n such that $Z_n \geq Z_n'$ for all n, then*

$$Z_n \overset{wvhp}{=} O(\alpha(n)) \implies Z_n' \overset{wvhp}{=} O(\alpha(n)) \tag{451}$$

$$and \quad Z_n' \overset{wvhp}{=} \Omega(\alpha(n)) \implies Z_n \overset{wvhp}{=} \Omega(\alpha(n)) \tag{452}$$

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

$$\mathbb{P}[Z_n > c\alpha(n)] \leqslant c_1 e^{-c_2 n^{c_3}} \tag{453}$$

Since $Z_n \geq Z_n'$ implies $\mathbb{P}[Z_n' > c\alpha(n)] \leqslant \mathbb{P}[Z_n > c\alpha(n)]$, we are done with Eq. (451). The proof of Eq. (452) is analogous. $\qquad \square$

Lemma 31. *If $Z_n \overset{wvhp}{\to} \alpha(n)$, then $Z_n/\alpha(n) \to 1$ almost surely.*

Proof. By Lemma 29, with probability 1 there is some finite n^* such that B_n happens for all $n > n^*$. Thus, for $n > n^*$, we have

$$1 - \delta_n \leqslant Z_n/\alpha(n) \leqslant 1 + \delta_n \tag{454}$$

and since $\delta_n \to 0$ as $n \to \infty$, the result follows. $\qquad \square$

The next lemma (along with Theorem 1) shows that our main results on the very-high-probability behavior of the DSTSP translate directly to expected value.

Lemma 32. *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 $A_n^{(1)} = \{Z_n < c_1 n^k\}$; $A_n^{(2)} = \{c_1 n^k \leqslant Z_n \leqslant c_2 n^k\}$ (where $0 < c_1 < c_2$ as in Definition 17); and $A_n^{(3)} = \{c_2 n^k < Z_n\}$. Then:

$$\mathbb{E}[Z_n] = \sum_{i=1}^{3} \mathbb{P}[A_n^{(i)}] \, \mathbb{E}[Z_n \mid A_n^{(i)}] \tag{455}$$

We can then upper bound and lower bound these since: $\mathbb{P}[A_n^{(2)}] \geqslant 1 - b_1 e^{-b_2 n^{b_3}}$ (for $b_1, b_2, b_3 > 0$ as in Definition 4) we know that $\mathbb{P}[A_n^{(1)}], \mathbb{P}[A_n^{(3)}] \leqslant b_1 e^{-b_2 n^{b_3}}$; and $\mathbb{E}[Z_n \mid A_n^{(i)}] = O(n^\ell)$ (and are nonnegative), and $\mathbb{E}[Z_n \mid A_n^{(2)}] = \Theta(n^k)$ (by definition).

Then since $b_1 e^{-b_2 n^{b_3}} \to 0$ faster than $n^{-\ell}$, terms $i = 1$ and 3 vanish as $n \to \infty$ while term $i = 2$ goes to $\Theta(n^k)$ and we are done. $\qquad\square$

This shows that Corollary 1 follows directly from Theorem 1 and Proposition 1.

APPENDIX B
EQUIREGULAR CONTROL-AFFINE DYNAMICS

In this appendix we show how, if Π is symmetric and Meta-Assumption 1 holds (that Π is control-affine and equiregular and $[\mathcal{X}_f]^{\mathcal{Q}}$ is contained in a compact subset of $\mathcal{Q}$), many of the assumptions in Section IV follow, either directly or from some other simpler and easier-to-verify condition. We will rely heavily on the notation and terminology given in *Control of Nonholonomic Systems* by Frederic Jean [19]. We make use of the following notions:

- We have control vector fields $h_1^{\mathrm{cont}}, \ldots, h_m^{\mathrm{cont}}$ (see Section II-C) 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 $h_1^{\mathrm{cont}}, \ldots, h_m^{\mathrm{cont}}$, 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 $q \in \mathcal{Q}$ (and which span $T_q(\mathcal{Q})$), 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 $q \in \mathcal{Q}$ in which the Lie algebra behaves smoothly in a neighborhood around 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 $q \in \mathcal{Q}^*$ 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 (q, ε),*

$$\text{Box}_{\boldsymbol{H}}(q, c\varepsilon) \subset R_\varepsilon(q) \subset \text{Box}_{\boldsymbol{H}}(q, C\varepsilon) \tag{456}$$

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

This states that at small scales ε the configuration ε-reachable set $R_\varepsilon(q)$ 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 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 $\text{Box}_{\boldsymbol{H}}(q, \varepsilon)$. This then yields, intuitively, a way to cover $R_\varepsilon(q)$ with a number of configuration $(\varepsilon/2)$-reachable sets, by taking the circumscribed box of $R_\varepsilon(q)$ and tiling it with the inscribed boxes of $R_{\varepsilon/2}(q')$ for a set of well-chosen q'.

We begin with Lemma 4, which asserts that if Meta-Assumption 1 holds for symmetric Π, 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 $q \in \mathcal{Q}$ and

$0 < \varepsilon \leqslant \varepsilon^*$ there is a set $Q_\varepsilon(\boldsymbol{q})$ of configurations such that

$$R_{2\varepsilon}(\boldsymbol{q}) \subseteq \bigcup_{\boldsymbol{q}' \in Q_\varepsilon(\boldsymbol{q})} R_\varepsilon(\boldsymbol{q}') \text{ and } |Q_\varepsilon(\boldsymbol{q})| \leqslant b. \tag{457}$$

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,

$$R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon). \tag{458}$$

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

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon/2) \subset R_{\varepsilon/2}(\boldsymbol{q}'). \tag{459}$$

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

$$b = \prod_{i=1}^{d} \lceil (2C/c)^{|I_i|} \rceil \tag{460}$$

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

$$R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \subseteq \bigcup_{j=1}^{b} \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}_j, c\varepsilon/2) \subseteq \bigcup_{j=1}^{b} R_{\varepsilon/2}(\boldsymbol{q}_j) \tag{461}$$

and we are done. $\qquad\square$

We then consider Lemma 5, which asserts that if Meta-Assumption 1 holds for symmetric Π, 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 $\boldsymbol{q} \in [\mathcal{X}_f]^{\mathcal{Q}}$,

$$\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) . \tag{462}$$

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)

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{\boldsymbol{q} \in [\mathcal{X}_f]^{\mathcal{Q}}} R_{\varepsilon*(\boldsymbol{q})}(\boldsymbol{q}) \tag{463}$$

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

$$[\mathcal{X}_f]^{\mathcal{Q}} \subseteq \bigcup_{i=1}^{S} R_{\varepsilon*(\boldsymbol{q}_i)}(\boldsymbol{q}_i) \tag{464}$$

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

$$k_\varepsilon := \min(k : \varepsilon \leqslant 2^{-k}\varepsilon^*_{\max}) = \lceil \log_2(\varepsilon^*_{\max}/\varepsilon) \rceil \leqslant \log_2(\varepsilon^*_{\max}/\varepsilon) + 1 \tag{465}$$

This also implies that

$$\varepsilon \geqslant 2^{-k_\varepsilon}\varepsilon^*_{\max} \tag{466}$$

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} configuration ε-reachable sets. But we can bound

$$b^{k_\varepsilon} \leqslant b^{\log_2(\varepsilon^*_{\max}/\varepsilon)+1} = b(\varepsilon^*_{\max}/\varepsilon)^{\log_2 b} = c^* \varepsilon^{-\log_2 b} \tag{467}$$

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. $\qquad\square$

A. Hierarchical Cell Structures under equiregularity

In this appendix we 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 Hierarchical Cell Structure (HCS) about any given $x \in \mathcal{X}_f$ with $s = 2$, i.e. we prove Lemma 6. We use the notation from the first part of this appendix, including letting $d = \dim(\mathcal{Q})$; and let $d' = \dim(\mathcal{X}) \leqslant d$.

We want to define equivalent notions to the Lie algebra but for the projection of the dynamics onto $\mathcal{X}$; we say that $\boldsymbol{q} \in \mathcal{Q}$ is *projection-regular* if the projections of the Lie bracket vector fields to $\mathcal{X}$ at $\boldsymbol{q}$ (or rather the projections $[h_I(\boldsymbol{q})]_{\mathcal{X}}$ of $h_I(\boldsymbol{q}) \in T_{\boldsymbol{q}}\mathcal{Q}$ to $T_x\mathcal{X}$ where $x = [\boldsymbol{q}]_{\mathcal{X}}$) behave smoothly. In this case, γ is consistent within a neighborhood around $\boldsymbol{q}$. We assume:

Assumption 11. *For any $x \in \mathcal{X}_f$ and any $\delta > 0$, there is a projection-regular $\boldsymbol{q}$ such that $[\boldsymbol{q}]_{\mathcal{X}} = x$ and $g(\boldsymbol{q}) \geqslant (1 - \delta)g(x)$.*

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

Definition 18. *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 $\alpha \leqslant 1$, 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*

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ and } \mathrm{Vol}_{\mathcal{X}}(S(\boldsymbol{q}_0, \varepsilon_0)) \geqslant \alpha\check{g}^{(\zeta)}(\boldsymbol{q}_0)\varepsilon_0^{\gamma} \qquad (468)$$

  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*

$$S(\boldsymbol{q}_0, \varepsilon_0) \subseteq \bigcup_{j=1}^{s^{\gamma}} S(\boldsymbol{q}_j, \varepsilon_0/s) \quad \text{and} \quad \boldsymbol{q}_j \in \bar{R}_{\varepsilon_0}(\boldsymbol{q}_0) \text{ for all } j \qquad (469)$$

The difference between Definition 18 and Definition 7 is that the volume of the cell is only expected to be a certain size relative to $\breve{g}^{(\zeta)}(\boldsymbol{q}_0)\varepsilon_0^\gamma$ (which bounds the volume of the workspace reachable set at $\boldsymbol{q}_0$ for small ε_0) as opposed to $\breve{g}^{(\zeta)}([\boldsymbol{q}_0]_\mathcal{X})\varepsilon_0^\gamma$; 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 HCS 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 $\breve{\cdot}^{(\zeta)}$ doesn't affect this as $\zeta \to 0$ except at a closed set of measure 0, see Section D).

We want to use Meta-Assumption 1 with the Uniform Ball-Box Theorem as in Section B, in order to create a coordinate map in a neighborhood around $\boldsymbol{q}_0$ in which reachable sets can be inscribed with a rectangle of given dimensions (and circumscribed by a rectangle whose dimensions are fixed multiples of the inscribed rectangle). While in Lemma 4 the focus was on covering the *circumscribed* rectangle of a configuration ε-reachable set $R_\varepsilon(\boldsymbol{q})$ with a set number of *inscribed* rectangles of configuration $(\varepsilon/2)$-reachable sets $R_{\varepsilon/2}(\boldsymbol{q}_1), \ldots, R_{\varepsilon/2}(\boldsymbol{q}_b)$ (thus proving that $R_\varepsilon(\boldsymbol{q}) \subseteq \bigcup_{j=1}^b R_{\varepsilon/2}(\boldsymbol{q}_j)$), in our case we will use inscribed rectangles as our cells; thus, we will want to cover inscribed rectangles of reachable sets (cells) with the inscribed rectangles of smaller reachable sets (sub-cells).

One additional wrinkle is that in Lemma 4 we needed to perform this covering for configuration-reachable sets; however, HCS's and LHCS's involve cells in the workspace. Hence, we will need to carefully consider the projection function $[\cdot]_\mathcal{X}$. Additionally, it was sufficient in Lemma 4 to achieve the covering with any fixed number b of smaller rectangles, for HCS's and LHCS's it is required to achieve the covering with exactly 2^γ smaller cells (since $s = 2$ for what follows).

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 HCS and LHCS definitions) and constants $0 < c < C$ such that for any $\boldsymbol{q}$ in a neighborhood of $\boldsymbol{q}_0$ and any $\varepsilon < \varepsilon_0^*$

$$\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon) \subset R_\varepsilon(\boldsymbol{q}) \subset \text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon) \tag{470}$$

(see the Uniform Ball-Box Theorem).

Given $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon)$, we note that we can cover it with 2^r instances of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c(\varepsilon/2))$, 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 HCS 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 16. *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 $\boldsymbol{q}$*

$$\frac{\text{Vol}_{\mathcal{Q}}(\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, c\varepsilon))}{\text{Vol}_{\mathcal{Q}}(\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}, C\varepsilon))} \geqslant \frac{1}{2}(c/C)^r \tag{471}$$

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 II, 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 $\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', c\varepsilon)$ 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 $\boldsymbol{H}(\boldsymbol{q}) = \{h_{I_1}(\boldsymbol{q}), \dots, h_{I_d}(\boldsymbol{q})\}$ is a basis of $T_{\boldsymbol{q}}\mathcal{Q}$; therefore

$$[\boldsymbol{H}(\boldsymbol{q})]_\mathcal{X} = \{[h_{I_1}(\boldsymbol{q})]_\mathcal{X}, \dots, [h_{I_d}(\boldsymbol{q})]_\mathcal{X}\} \tag{472}$$

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

$$\bar{\boldsymbol{H}}(\boldsymbol{q}) := [\boldsymbol{H}(\boldsymbol{q})]_\mathcal{X} = \{\bar{h}_1(\boldsymbol{q}), \dots, \bar{h}_d(\boldsymbol{q})\}. \tag{473}$$

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 17. *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 $x' = [\boldsymbol{q}']_\mathcal{X}$:

$$\mathrm{Box}_{\boldsymbol{H}_A}(\boldsymbol{q}', \varepsilon) = \{\exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}') : |z_i| \leqslant \varepsilon^{|I_i|} \ \forall i \in A \text{ and } z_i = 0 \ \forall i \notin A\} \tag{474}$$

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,

$$\mathrm{Vol}_\mathcal{X}([\mathrm{Box}_{\boldsymbol{H}_A}(\boldsymbol{q}', \varepsilon)]_\mathcal{X}) = \Theta(\varepsilon^{r_A}). \tag{475}$$

We first consider a 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 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 18. *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:

$$\text{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon) = \tag{476}$$

$$\{\exp(z_1 h_{I_1}) \circ \cdots \circ \exp(z_d h_{I_d})(\boldsymbol{q}') : |z_i| \leqslant \varepsilon^{|I_i|} \ \forall i \in A, z_i = a_i \varepsilon^{|I_i|} \ \forall i \notin A\} \tag{477}$$

This denotes a d'-dimensional facet of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ 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

$$[\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}} = \bigcup_{A \subseteq [d]:|A|=d'} \bigcup_{\boldsymbol{a} \in \{-1,1\}^{[d]\backslash A}} [\text{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}} \tag{478}$$

which is just saying that the projection of $\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$ onto $\mathcal{X}$ is the same as the union of the projections of all its dimension-d' facets onto $\mathcal{X}$. This means

$$\text{Vol}_{\mathcal{X}}([\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \leqslant \sum_{A \subseteq [d]:|A|=d'} \sum_{\boldsymbol{a} \in \{-1,1\}^{[d]\backslash A}} \text{Vol}_{\mathcal{X}}([\text{Box}_{\boldsymbol{H}_A^{(\boldsymbol{a})}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \tag{479}$$

$$= \Theta(\varepsilon^{r_{A*}}) \tag{480}$$

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 $\text{Box}_{\boldsymbol{H}_{A*}}(\boldsymbol{q}', \varepsilon) \subseteq \text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)$, we have

$$\Theta(\varepsilon^{r_{A*}}) = \text{Vol}_{\mathcal{X}}([\text{Box}_{\boldsymbol{H}_{A*}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \leqslant \text{Vol}_{\mathcal{X}}([\text{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) \tag{481}$$

and hence we have

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) = \Theta(\varepsilon^{r_{A*}}) \tag{482}$$

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

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) = \Theta(\varepsilon^{\gamma}) \tag{483}$$

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

$$\mathrm{Vol}_{\mathcal{X}}([\mathrm{Box}_{\boldsymbol{H}_{A'}}(\boldsymbol{q}', \varepsilon)]_{\mathcal{X}}) > 0 \tag{484}$$

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

$$\min(r_A : A \subseteq [d] \text{ s.t. } |A| = d' \text{ and } \bar{\boldsymbol{H}}_A(\boldsymbol{q}') \text{ is linearly independent}) \leqslant r'_{A'}. \tag{485}$$

Therefore, we may choose

$$A^* := \arg\min(r_A : A \subseteq [d] \text{ s.t. } |A| = d' \text{ and } \bar{\boldsymbol{H}}_A(\boldsymbol{q}') \text{ is linearly independent}) \tag{486}$$

Then the arguments from above (particularly equations (478) through (483)) 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 HCS around x. There exists a efficiency parameter $\alpha > 0$ which, for sufficiently small ε, holds for all the sub-HCS'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.

APPENDIX C

EXISTENCE OF ACCURATE HCS COVERS

In this appendix we show formally the existence of ρ-accurate HCS covers, i.e. we prove Lemma 7 and Lemma 8 from Assumption 10. Recall that Assumption 10 states that every point in $\mathcal{X}_f$ is contained inside some HCS; Lemma 7 states that we can therefore cover $\mathcal{X}_f$ with a finite number of HCS's of some scale ε_0; and Lemma 8 states that we can do so at arbitrarily small scales ε_0 with arbitrarily small accuracy parameter $\rho > 0$.

Proof of Lemma 7. By Assumption 10 each $x \in \mathcal{X}_f$ is contained in the interior of some HCS 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 2, $\mathcal{X}_f$ is compact, and hence there is a finite subset of these S_x's whose interiors cover $\mathcal{X}_f$. $\qquad\square$

Proof of Lemma 8. 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 HCS, the sub-HCS's are all disjoint, and that any scale-ε HCS is contained in some ball in $\mathcal{X}$ with radius $c_\Pi \varepsilon$.

By Lemma 7 we have a HCS 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

$$\text{Int}_\eta(A) := \{x \in A : d_{\mathcal{X}}(x, x') \geq \eta \text{ for all } x' \notin A\} \tag{487}$$

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 $\text{Int}_\eta(A) = \bigcap_{j=1}^k \text{Int}_\eta(A_j)$ and $\text{Int}(A) = \bigcap_{j=1}^k \text{Int}(A_j)$. Since intersections of finitely many Jordan-measurable sets are Jordan-measurable, we also have $\text{Vol}_{\mathcal{X}}(A) = \text{Vol}_{\mathcal{X}}(\text{Int}(A))$.

We now subdivide each $S_j^{(\text{start})}$ them k times, for any integer k to get a cover of $m := m^{\text{start}} s^{\gamma k}$ HCS's at scale $\varepsilon_0 := \varepsilon_0^{(\text{start})} s^{-k}$. Let these be denoted $S_1, \ldots, S_m$; let $S_1^{(j)}, \ldots, S_{s^{\gamma k}}^{(j)}$ denote the HCS's which were created by subdividing $S_j^{(\text{start})}$. For each $S_i^{(j)}$, we then remove it if there is some $j' < j$ such that $S_i^{(j)} \subseteq S_{j'}^{(\text{start})}$, i.e. if it is completely contained within a different HCS

from our starting cover with a lower index. This is a simple way to remove redundant HCS's, and preserves the covering property; to see this, we can perform this action starting with the descendants of $S_{m^{(\text{start})}}^{(\text{start})}$, then the descendants of $S_{m^{(\text{start})}-1}^{(\text{start})}$ 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 HCS's) to an arbitrarily small set. Let

$$S_k(j, j') := \{x : x \in S_i^{(j)} \cap S_{i'}^{(j')} \text{ for some } i, i'\} \tag{488}$$

i.e. the region in which descendants of $S_j^{(\text{start})}$ and $S_{j'}^{(\text{start})}$ 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 $S_0(j, j') = S_j^{(\text{start})} \cap S_{j'}^{(\text{start})}$, and that

$$S_k(j, j') \subseteq S_{k'}(j, j') \text{ for all } k > k'. \tag{489}$$

We now let

$$\eta := 2c_\Pi \varepsilon_0 = 2c_\Pi \varepsilon_0^{(\text{start})} s^{-k} \tag{490}$$

and claim that

$$S_k(j, j') \cap \text{Int}_\eta(S_0(j, j')) = \varnothing \tag{491}$$

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. $x = [\boldsymbol{q}]_{\mathcal{X}}$)

$$S_i^{(j)} \subseteq \bar{R}_{\varepsilon_0}(\boldsymbol{q}) \subseteq \mathcal{B}_{c_\Pi \varepsilon_0}(x) \tag{492}$$

This means that any cell $S_i^{(j)}$ rooted at x is necessarily deleted if $x \in \text{Int}_{c_\Pi \varepsilon_0}(S_0(j, j'))$; and this additionally means that if $S_i^{(j)}$ *contains* x where $x \in \text{Int}_{2c_\Pi \varepsilon_0}(S_0(j, j')) = \text{Int}_\eta(S_0(j, j'))$, then $S_i^{(j)}$ is removed as well. Thus we get (491), which further implies

$$S_k(j, j') \subseteq S_0(j, j') \backslash \text{Int}_\eta(S_0(j, j')). \tag{493}$$

But since $S_0(j, j') = S_j^{(\text{start})} \cap S_{j'}^{(\text{start})}$, 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

$$S_k' := \{x \in \mathcal{X}_f : x \in S_i^{(j)} \cap S_{i'}^{(j')} \text{ for some } i, j, i', j' \text{ where } j \neq j'\} \tag{494}$$

we get that

$$S'_k = \bigcup_{j \neq j'} S_k(j, j') \tag{495}$$

$$\implies \mathrm{Vol}_{\mathcal{X}}(S'_k) \leqslant \sum_{j \neq j'} \mathrm{Vol}_{\mathcal{X}}(S_k(j, j')) \tag{496}$$

$$\implies \lim_{k \to \infty} \mathrm{Vol}_{\mathcal{X}}(S'_k) \leqslant \lim_{k \to \infty} \sum_{j \neq j'} \mathrm{Vol}_{\mathcal{X}}(S_k(j, j')) = 0 \tag{497}$$

Thus for any $\rho > 0$, we can set k sufficiently large so that

$$\mathbb{P}_{X \sim f}[X \in S'_k] \leqslant \rho \tag{498}$$

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^*. $\qquad\square$

APPENDIX D
APPROXIMATELY EVERYWHERE

We now prove Proposition 4, which states that our main results (Proposition 2 and Proposition 3) hold even when our assumptions are true only approximately everywhere (Definition 9) rather than everywhere.

Proof of Proposition 4. 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 8, 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

$$x \in \mathcal{X}'_\eta \iff \inf_{x' \notin \mathcal{X}_\eta} d_{\mathcal{X}}(x, x') \geqslant \eta \tag{499}$$

$$\iff \inf_{i \in [k]} \inf_{x' \notin \mathcal{X}_\eta^{(i)}} d_{\mathcal{X}}(x, x') \geqslant \eta \tag{500}$$

$$\iff x \in \mathcal{X}_\eta^{(i)} \text{ for all } i \in [k] \tag{501}$$

$$\iff x \in \bigcap_{i=1}^k \mathcal{X}_\eta^{(i)'} \tag{502}$$

Then we can compute:

$$\text{Vol}_{\mathcal{X}}(\mathcal{X}_f \backslash \mathcal{X}'_\eta) = \text{Vol}_{\mathcal{X}}\left(\mathcal{X}_f \backslash \bigcap_{i=1}^k \mathcal{X}_\eta^{(i)'}\right) \tag{503}$$

$$= \text{Vol}_{\mathcal{X}}\left(\bigcup_{i=1}^k (\mathcal{X}_f \backslash \mathcal{X}_\eta^{(i)'})\right) \tag{504}$$

$$\leqslant \sum_{i=1}^k \text{Vol}_{\mathcal{X}}(\mathcal{X}_f \backslash \mathcal{X}_\eta^{(i)'}) \tag{505}$$

$$\leqslant k\eta \tag{506}$$

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

$$\lim_{\eta \to 0} \mathbb{P}_{X \sim f}[X \in \mathcal{X}'_\eta] = 1 \text{ and } \lim_{\eta \to 0} \int_{\mathcal{X}'_\eta} f(x)^{1 - \frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx = \int_{\mathcal{X}_f} f(x)^{1 - \frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{507}$$

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 $X_1, \ldots, X_n$,

$$\text{TSP}_\Pi(\{X_1, \ldots, X_n\} \cap \mathcal{X}'_\eta) \leqslant \text{TSP}_\Pi(X_1, \ldots, X_n) \tag{508}$$

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

$$\mathbb{P}_{X \sim f}[X \notin \mathcal{X}'_\eta] \leqslant \delta_1/2 \tag{509}$$

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

$$f|_{\mathcal{X}'_\eta}(x) = \begin{cases} f(x)/\mathbb{P}_{X \sim f}[X \in \mathcal{X}'_\eta] & \text{if } x \in \mathcal{X}'_\eta \\ 0 & \text{if } x \notin \mathcal{X}'_\eta \end{cases} \tag{510}$$

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

$$f(x) \leqslant f|_{\mathcal{X}'_\eta}(x) \leqslant (1 - \delta_1/2)^{-1} f(x) \tag{511}$$

which in turn means that

$$\int_{\mathcal{X}'_\eta} f|_{\mathcal{X}'_\eta}(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \leqslant (1 - \delta_1/2)^{-(1-\frac{1}{\gamma})} \int_{\mathcal{X}'_\eta} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{512}$$

$$\leqslant (1 - \delta_1/2)^{-(1-\frac{1}{\gamma})} \int_{\mathcal{X}_f} f(x)^{1-\frac{1}{\gamma}} g(x)^{-\frac{1}{\gamma}} \, dx \tag{513}$$

We can then turn to the Orienteering problem, in particular the Cost-Balanced Orienteering problem Definition 12 and the very high probability upper bound Proposition 6 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 6 and set

$$\lambda < \frac{\eta \zeta}{c_\Pi} \tag{514}$$

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 6 holds and we obtain the bound

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

$$\tag{515}$$

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. □

APPENDIX E

GOOD TARGET BAD TARGET

We now claim that even though our HCS cover might not be entirely overlap-free, because it can be arbitrarily overlap-free (see Lemma 8) we can derive the main results in the same way as in Section VI. 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 26 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 $\varepsilon^{(\mathrm{bad})}$ with $m^{(\mathrm{bad})}$ different HCS's, as in Lemma 7; this is allowed to overlap to any amount, but note that it is fixed independent of ρ^*.

Let $n^{(\mathrm{bad})}$ be the number of bad target points, and $n_j^{(\mathrm{bad})}$ be the number of bad target points in the jth cell of the fixed HCS cover with $m^{(\mathrm{bad})}$ cells at scale $\varepsilon^{(\mathrm{bad})}$. Note by Proposition 11 and the concavity of $x^{1-\frac{1}{\gamma}}$ that this means that

$$\sum_{j=1}^{m^{(\mathrm{bad})}} \mathrm{HCP}^*(n^{(\mathrm{bad})}; \hat{b}, s) \leqslant \sum_{j=1}^{m^{(\mathrm{bad})}} 6s(n^{(\mathrm{bad})}/m^{(\mathrm{bad})})^{1-\frac{1}{\gamma}} = 6s(m^{(\mathrm{bad})})^{\frac{1}{\gamma}}(n^{(\mathrm{bad})})^{1-\frac{1}{\gamma}} \tag{516}$$

Thus, letting $\boldsymbol{q}_1^{(\mathrm{bad})}, \ldots, \boldsymbol{q}_{m^{(\mathrm{bad})}}^{(\mathrm{bad})}$ be the roots of this HCS cover and letting

$$C^{(\mathrm{bad})} := \mathrm{TSP}_\Pi(\boldsymbol{q}_1^{(\mathrm{bad})}, \ldots, \boldsymbol{q}_{m^{(\mathrm{bad})}}^{(\mathrm{bad})}). \tag{517}$$

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

$$\text{TSP}_\Pi(\{X_i : X_i \text{ is bad}\}) \leqslant C^{(\text{bad})} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}} \varepsilon^{(\text{bad})} (n^{(\text{bad})})^{1-\frac{1}{\gamma}} . \tag{518}$$

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

$$\text{TSP}_\Pi(\{X_i : X_i \text{ is bad}\}) \leqslant C^{(\text{bad})} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}} \varepsilon^{(\text{bad})} 2^{1-\frac{1}{\gamma}} (\rho^*)^{1-\frac{1}{\gamma}} n^{1-\frac{1}{\gamma}} \tag{519}$$

By Lemma 26, for any $\delta^* > 0$, we can then fix ρ^* sufficiently small so that for sufficiently large n,

$$C^{(\text{bad})}/n^{1-\frac{1}{\gamma}} + 6s(m^{(\text{bad})})^{\frac{1}{\gamma}} \varepsilon^{(\text{bad})} 2^{1-\frac{1}{\gamma}} (\rho^*)^{1-\frac{1}{\gamma}} \leqslant \delta^* \tag{520}$$

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.

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).

APPENDIX F

EUCLIDEAN TSP FOR FRACTALLY-DISTRIBUTED 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 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 $x \in \mathcal{X}$.[2] 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.

---

[2]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 $n \to \infty$.

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 collection 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

$$\dot{x} = h(x)\boldsymbol{u} \tag{521}$$

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

$$\phi(x) = \lim_{\varepsilon \to 0} \frac{\mathbb{P}_{X \sim f}[X \in \bar{R}_\varepsilon(x)]}{\varepsilon^\gamma} \tag{522}$$

(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 14. *For noninteger γ, we get the following bounds: if $\gamma \geqslant 3$ is noninteger, then bounds (29), (30), and (32) from Proposition 3 still hold. If $\gamma \in (2,3)$, then for all sufficiently large n,*

$$\mathbb{P}\big[\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\mathbb{E}_{X \sim f}[\phi(x)^{-\frac{1}{\gamma}}]\big] \tag{523}$$

$$\geqslant 1 - e^{-\frac{(\gamma-2)p_1 n (\sum_{j=1}^m p_j^{1-1/\gamma})^2}{13\gamma}} \tag{524}$$

And, finally, if $\gamma \in (1,2)$, then for all sufficiently large n,

$$\mathbb{P}\big[\mathrm{TSP}_\Pi(\{X_i\}) \leqslant (1+\delta)\big(12s\alpha^{-\frac{1}{\gamma}}\big)n^{1-\frac{1}{\gamma}}\mathbb{E}_{X \sim f}[\phi(x)^{-\frac{1}{\gamma}}]\big] \tag{525}$$

$$\geqslant 1 - e^{-\frac{p_1 n^{2(1-1/\gamma)} (\sum_{j=1}^m p_j^{1-1/\gamma})^2}{\frac{560}{3}\log(\gamma/(\gamma-1))+4+18(1-1/\gamma)^2 \frac{\gamma}{\gamma-2}(\frac{280}{3}\log(\gamma/(\gamma-1)))^{1-2/\gamma}}} \tag{526}$$

Proof. This is proved in the same manner as the upper bounds in Section VI using the HCS cover and the collection problem. □