Conditional Density Growth (CDG) model: a simplified model of RRT coverage for kinematic systems

Joel M. Esposito*

Department of Systems Engineering, United States Naval Academy, Annapolis, MD 21402, USA

SUMMARY
It has been shown before that the Rapidly Exploring Random Tree (RRT) algorithm is probabilistically and resolution complete; and that the probability of finding a particular solution path can be related to the number of nodes. However, little analysis exists on the rate at which the tree covers the configuration space. In this paper, we present a stochastic difference equation which models how the tree covers the configuration space as a function of the number of nodes in the tree. Using two simplifying assumptions, appropriate for holonomic, kinematic systems in expansive configuration spaces, we derive closed-form solutions for the expected value and variance of configuration space coverage, which only depend on two easily computable parameters. Using a grid-based coverage measurement, we present experimental evidence supporting this model across a range of dimensions, obstacle densities, and parameter choices. Collecting data from 1000 RRTs, we provide evidence that configuration space coverage concentrates tightly around the expected coverage predicted by the model; and the results of the Chi-squared test suggest that the distribution of coverage across these runs is highly Gaussian. Together these results enable one to predict the expected coverage, along with a confidence interval, after a certain number of nodes have been added to the tree. We also applied the model to an example with extremely narrow passages and to a system with non-holonomic kinematics. The expected value prediction is still qualitatively accurate; but the rate constant is reduced and the variance is higher. Overall, in addition to its theoretical value, the model may find future application as an online measure of search-progress and problem difficulty, useful for adaptive variants of the basic RRT algorithm.

KEYWORDS: Path planning for manipulators; Motion control; Motion planning; Randomized methods.

1. Introduction
Randomized algorithms, such as Probabilistic Road Maps (PRM),[2,17] Rapidly Exploring Random Trees (RRTs)[24,25] and their variants,[3,4,21] are a powerful and practically important class of motion planning methods that have been used to address a variety of real-world problems successfully. Their appeal lies in their ability to address large

* Corresponding author. E-mail: esposito@usna.edu

or complex problems in an incremental fashion.[22] In contrast, deterministic, computational geometry-type approaches to motion planning (for example ref. [6]) suffer from well-known complexity limitations because they attempt to explicitly construct and manipulate the free configuration space. However, the price of this incremental approach is a reduction in completeness. Most computational geometry-type algorithms are algorithmically complete – meaning that they are guaranteed to find a solution to a problem if one exists; or report that none exists upon termination. They are also guaranteed to terminate in finite time. Randomized methods sacrifice algorithmic completeness for weaker *Probabilistic Completeness*[23] – meaning that if a solution exists, the probability that it is found approaches one as the number of iterations approaches infinity. While establishing an algorithm's probabilistic completeness is of great theoretical importance, it tells us nothing about the rate at which the configuration space is covered. Little work has been done to enhance the theoretical foundation of RRTs in this regard.

From a practical point of view, understanding the expected rate of coverage would be useful for planning problems in which it is not known *a priori* if a solution exists. Because, as long as no solution is found, the algorithm has no standard termination criteria. Therefore, the roboticist must use experience and intuition to decide when to stop the search. It would be nice to be able to say, for example, that after 100, 000 nodes have been added, we are 95% confident that 99.99% of the space has been explored.

This paper presents an idealized model of RRT growth, called the Conditional Density Growth (CDG) model, primarily suited to holonomic systems operating in expansive configuration spaces. From this model we are able to derive various statistical properties of the RRT's configuration space coverage such as the expected value, variance, and distribution properties.[1] Section 2 provides a problem statement, reviews the RRT algorithm, and surveys related work. Section 3 presents the main analysis. First, coverage rate (also known as growth) is defined here as the incremental change in the volume fraction of the configuration space that

---

[1] An earlier version of this work was presented at the 2012 IEEE/RSJ International Conference on Intelligent Robot and Systems, during the invited session *50 Years of Robot Motion Planning: Achievements and Emerging Approaches*. The current version has been significantly expanded to include analysis of coverage variance and concentration, as well as more examples.

could be connected to the tree in the next iteration. Based on two simplifying assumptions, we develop a stochastic difference equation which models the coverage of holonomic systems operating in expansive spaces. Our main result is that the expected value of coverage is exponential in the number of nodes; and that the variance is a cubic function of coverage, peaking when two-thirds of the space has been explored. The model and assumptions are then validated against experimental computational data in Section 4. A grid-based method is used to approximately measure the configuration space coverage; and the model is fit with good agreement across a variety of dimensions and parameter values. Evidence suggests that, over many trials, configuration space coverage concentrates tightly around the expected value predicted by the model in a Gaussian-like fashion. This is significant because it suggests that the expected value and the variance give sufficient information to place a confidence interval on the expected coverage after a given number of nodes have been added to the tree. Section 5 examines the application of the model to non-holonomic systems and poorly expansive workspaces. Section 6 discusses the significance of the model, as well as ways in which it could be extended. While this class of systems is admittedly limited, the CDG model introduced in this paper enhances our theoretical understanding of the algorithm and can be used as a base line for comparing algorithmic enhancements.

2. Background

Problem 1. Given: the free configuration space $\mathcal{X}$; a compact time interval, $T = [t^{\text{init}}, t^{\text{final}}]$; a compact set of possible initial states $\mathcal{X}^{\text{init}} \subset \mathcal{X}$; a compact set of input values U; dynamics $\dot{x} = f(x, u)$, Lipschitz in $x \in \mathcal{X}$ and $u \in U$; and a compact goal set $\mathcal{S} \subset \mathcal{X}$. Find a function $u : T \to U$ (if one exists) such that, when the state evolves from $x \in \mathcal{X}^{\text{init}}$, and is restricted to lie in $\mathcal{X}$, it produces a trajectory such that $\exists t^* \in T \, | \, x(t^*) \in \mathcal{S}$.

Note that in the first part of this paper we strictly focus on versions of this problem in which the dynamics $f(x, u)$ are holonomic; and $\mathcal{X}$, while not necessarily simply connected, is expansive as defined in ref. [12].

A RRT is an incremental randomized search algorithm that is designed for a broad class of motion planning problems.[24] It utilizes a tree-like data structure, $\mathcal{T}$, whose vertices (also known as nodes) are points in the free configuration space $x \in \mathcal{X}$, and edges are constant inputs $u \in U$ connecting two vertices in time step Δt. Many variants of the algorithm have been introduced. While we consider kinematic systems in this paper, our ultimate goal is to extend the analysis to a richer class of problems. Among other things this often precludes bi-directional searches and requires the use of a local planner, numerical integration method, and limited time step size. To that end, we have chosen the forward search version of the RRT method presented in ref. [23, Ch. 14, p. 834] to implement and analyze. This extension of the original RRT is capable of solving problems with differential constraints.

Consider the outlines provided in Algorithms 1 and 2. Algorithm 1 simply calls Algorithm 2 Extend, repeatedly

adding vertices and edges to the tree until a solution is found. As long as no solution has been found there is no accepted termination criteria. Algorithm 2 is the heart of the RRT method. It generates a sample, x^{rand}, from a random variable X^{rand}, with sample space $\mathcal{X}$ and density function $p(\cdot)$. We assume the uniform distribution is used, since it is by far the most common choice in the literature. It then uses $\rho(\cdot, \cdot)$, a distance function defined according to some suitable (pseudo) metric, to select x^{near}, the closest existing vertex to the sample x^{rand}. A local motion planner is then used to select a new input u^{new}, which can steer the system as close as possible to x^{rand} in time Δt beginning from x^{near}. A local planner is a quick, but possibly incomplete, algorithm that attempts to connect two configurations in close proximity to each other (small Δt). In the case of holonomic, kinematic systems, the local planner may simply use line segments to connect x^{rand} and x^{near}. For systems with more complex dynamics, possible local planning techniques include steering controllers which exploit the structure of the dynamics, heuristic approaches such as potential fields, or discretization techniques which approximate the set of admissible input values U with a countable set of inputs $\bar{U}$. Once an input is selected, any numerical integration technique that can simulate the system's response to piecewise constant inputs can then be used, with u^{new} as an input and x^{near} as an initial condition to generate x^{new}. If the new state is not already a vertex in the tree, it will be added along with u^{new} as an edge. It is possible for the local planner to fail to generate a *new* tree node, for example if all possible values in $\bar{U}$ have been explored from that particular x^{near}, but we do not consider such resolution issues in this paper.

Algorithm 1 Generate RRT: $\mathcal{T}$ (Based on ref. [23, Ch. 14])

Initialize RRT: $\mathcal{T}$.addVertex($x^1 \in \mathcal{X}^{\text{init}}$)
while $\nexists x \in \mathcal{T} \bigcap \mathcal{S}$ **do**
    Extend($\mathcal{T}$)
end while

Algorithm 2 Extend($\mathcal{T}$) (Based on ref. [23, Ch. 14])

$x^{\text{rand}} \leftarrow X^{\text{rand}}$
$x^{\text{near}} \leftarrow \arg\min_{x \in \mathcal{T}} \rho(x, x^{\text{rand}})$
$u^{\text{new}} = \arg\min_{u \in U} \{\rho(x^{\text{near}} + \int^{\Delta t} f(x, u) dt, x^{\text{rand}})\}$
$x^{\text{new}} = x^{\text{near}} + \int^{\Delta t} f(x, u^{\text{new}}(t)) dt$
if $x^{\text{new}} \notin \mathcal{T}$ **then**
    $\mathcal{T}$.addVertex(x^{new})
    $\mathcal{T}$.addEdge(u^{new}, $x^{\text{near}} \to x^{\text{new}}$)
end if

The advantage of the RRT algorithm is that it works directly with the set of admissible inputs and is therefore applicable to systems with complex dynamics. It is well suited to the problem of quickly searching high-dimensional spaces that have both algebraic and differential constraints. A second advantage is that it biases the exploration toward unexplored portions of the space by randomly sampling points in the state space and incrementally pulling the search tree toward them, reducing the size of the largest Voronoi

regions as the tree grows.[25,26] The RRT algorithm was shown to be probabilistically complete in ref. [24], with some discussion of the odds of discovering particular solution paths in ref. [25]. In ref. [7] the RRT algorithm, with some technical modifications, was shown to be resolution complete – meaning that accounting for discretization effects in time, input, and the state space, the algorithm is guaranteed to find goal sets of a certain minimum size. Recently the quality of the solutions emerging from the RRT have been shown to be asymptotically sub-optimal; and a new algorithm, RRT*, has been introduced, which is asymptotically optimal.[15]

From a theoretical point of view, the solution time (vice coverage) has been as analyzed before – both for the RRT[25] and other randomized planning algorithms.[16] Typically in such "path finding" approaches, a particular path of interest is identified (ex. a path connecting a specific initial state to a specific goal state). Given some constants capturing the properties of this path, such as length and obstacle clearance within the configuration space, one determines the odds of discovering a homotopically equivalent path as a function of the number of sample nodes. A similar result bounding the probability of failure was reported in ref. [8] in terms of the number of milestones and lookout sets. The approach was generalized in ref. [20] to include more abstract metric spaces. The approach is of great theoretical importance since it explains known challenges in finding paths through narrow passageways in the free space. Several interesting observations regarding the variance in coverage rate were reported in ref. [28]. However, surprisingly little attention has been directed at developing predictive models for the *rate* at which the free space is covered by the tree. Of course, this is partially because, in practice, the goal of such methods is not necessarily to explore the entire state space but rather to find a sequence of inputs from x^1 to S as quickly as possible. Indeed many adaptive methods and enhancements to the RRT (such as refs. [1, 5, 19]) improve performance without necessarily improving the coverage of the state space.

Despite its natural connection to pragmatic goal of finding a solution quickly, "path finding" analysis has several limitations. First, in realistic applications, one can rarely identify such paths of interest in advance. Second, assuming a path is identified, many of the constants used in establishing convergence bounds may be impractical to compute. Finally, the approach is extremely sensitive to the problem definition. In particular, a small perturbation in the location of the initial state or goal state can dramatically alter which areas of the state space are considered "interesting."

We believe that a space-covering approach to analysis overcomes some of these challenges. Space-covering approaches attempt to determine the probability of finding *all possible* paths.[12,13] Therefore, this type of analysis finds application in the so-called verification or testing problems, such as those considered in refs. [18, 19]. Since a "path of interest" does not need to be identified, limited *a priori* knowledge is required. More importantly, the approach is insensitive to perturbations in the initial and goal configurations. Analyzing coverage is also akin to considering worst case solution times.
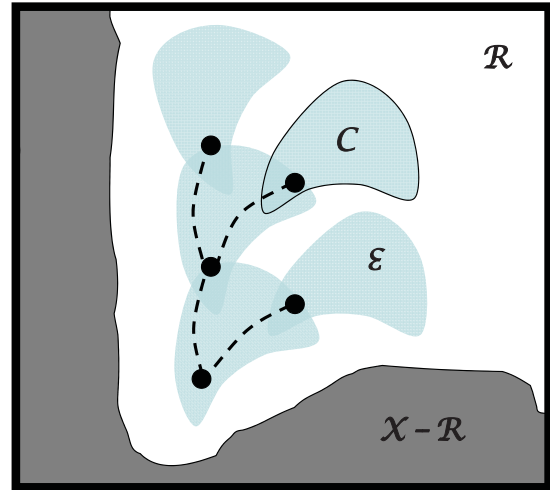

Fig. 1. (Colour online) Subsets of the configuration space: the reachable set, $\mathcal{R}$ (white); the explored set, $\mathcal{E}$ (blue); and the connection set, $\mathcal{C}$, from x^{near} (blue outlined in black).

3. Analysis
In this section, we define what we mean by "exploring" the state space, and present a stochastic model of this process.

3.1. Coverage definition
We would like to determine what fraction of the reachable set a tree, $\mathcal{T}$, has explored. However, because a tree node x^k is a point and an edge is a line, the tree itself has measure zero and, therefore, can never fill the state space. Extending some concepts and notation from refs. [19, 23], define the set of states R that could be connected to the tree from a node x^k in a single iteration, using time step Δt, as

$$R(x^k, \Delta t) = \{x \in X | \exists u \in U, x = x^k + \int^{\Delta t} f(x, u)dt\}.$$

Note that this set has a non-zero measure, allowing us to discuss its volume, which can be related to the fraction of the space that can be explored from that node. This notation has the following three principle uses, as illustrated in Fig. 1:

1. The *Connection Set* $\mathcal{C} = R(x^{\text{near}}, \Delta t)$ is a set of states that can be connected to x^{near} in a single iteration.
2. The *Explored Set* is defined by $\mathcal{E}_K = R(\mathcal{T}_K, \Delta t) = \bigcup_{k=1}^{K} R(x^k, \Delta t)$. It is the set of states that can be connected to, from any existing node in the tree, in a single iteration. It is used to define coverage. Note that if the intersection of the explored set and the goal set $\mathcal{S}$ are non-empty, the algorithm can terminate.
3. The *Reachable Set* is defined as $\mathcal{R} = R(x^1, T)$. It is the set of all states that can possibly be explored from initial condition x^1 within the given time interval $T = [t^{\text{init}}, t^{\text{final}}]$. It is a super set of the Explored Set.

Using this notation we define the *Covered Volume Fraction*, or *coverage* for short, of a tree with K nodes as

$$C_K = \frac{\mu(\mathcal{E}_K)}{\mu(\mathcal{R})}, \tag{3.1}$$

the ratio of the volume of the explored set to that of the reachable set, where $\mu(\cdot)$ is used to denote the volume (measure) of a set.

When a new node, x^K is added to the tree, the change in the covered volume fraction is the *growth*

$$G_K = \frac{\mu\left(\mathcal{E}_K - \mathcal{E}_{K-1}\right)}{\mu(\mathcal{R})} = C_K - C_{K-1}. \qquad (3.2)$$

Remark 1. Note that obstacles in the configuration space are explicitly accounted for in the coverage measures in Eqs. (3.1) and (3.2) because they reduce the volume of $\mathcal{R}$ in the denominator.

3.2. Coverage model: holonomic systems in expansive spaces

In the remainder of this paper we use the following notation.

If X is a random variable with probability density function $p(x)$, then $Pr[|x| < c]$ is the probability the absolute value of x is less than some positive number c. $E(X)$ and $V(X)$ denote the expected value and variance of X. The notation $|$ is used to denote probabilities, expected value, or variances conditioned on some other event.

Since Algorithm 2 is driven by a random variable X^{rand}, the coverage of a tree with K nodes, $C_K \in [0, 1]$, is the state of a random, discrete time process,

$$C_K = C_{K-1} + G_K, \qquad (3.3)$$

with the exception of C_1, which is determined solely by the initial condition. We choose to model the growth as a function of two random variables and the parameter Δt, $G_K(X^{\mathrm{rand}}, C_{K-1}; \Delta t)$.

Assumption 1. Resolution factors, such as the choice of integration method, do not significantly influence the expected value of G_K. For a more complete discussion of resolution, the reader is referred to ref. [7]

The assumptions and model that follow are motivated by some observations of users of the RRT (see refs. [4,19, 22, 25]):

1. When random states are generated in large Voronoi regions (typically far away from the existing nodes), the tree *usually* experiences significant expansion.
2. When random states are generated in small Voronoi regions, the tree *usually does not* expand far from the existing nodes.
3. The size of the largest Voronoi regions decreases very quickly at first, then more slowly as more nodes are added.

In light of these observations, we propose the CDG model:

$$p(G_K|x^{\mathrm{rand}}) \approx \begin{bmatrix} p^{\mathrm{int}}(G_K), & \text{if } x^{\mathrm{rand}} \in \mathcal{E} \\ p^{\mathrm{ext}}(G_K), & \text{if } x^{\mathrm{rand}} \in \mathcal{X} - \mathcal{E} \end{bmatrix}, \qquad (3.4)$$

which assumes G_K is drawn from one of the two underlying probability distributions. Next we formally introduce two simplifying assumptions based on Observations 1 and 2, which enable us to develop a model of coverage, with
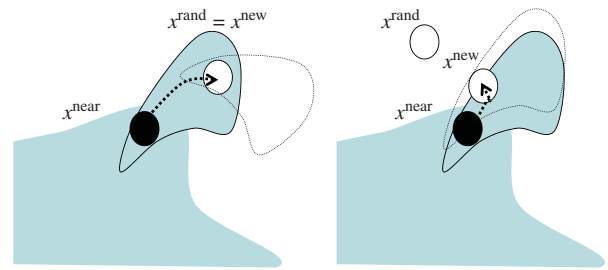

Fig. 2. (Colour online) Counterexamples. Assumption 2 (left), the tree can grow even if $x^{\mathrm{rand}} \in \mathcal{E}$, especially if x^{near} is on the boundary of the explored region. Regarding Assumption 3 (right), it is possible to have very small growth if $x^{\mathrm{rand}} \in \mathcal{X} - E$, especially when the metric poorly reflects the "cost to go."
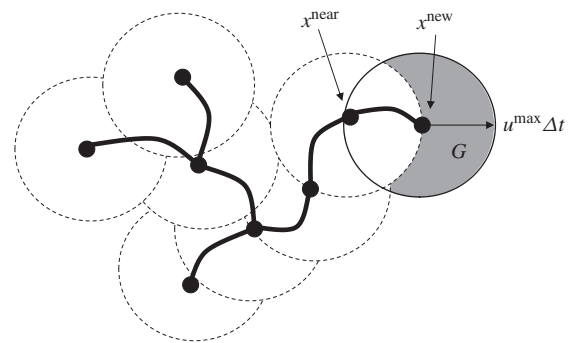

Fig. 3. For holonomic systems with no obstacles, the connection sets are balls. The maximum growth G^{max} is the volume of the gray area formed by $R(x^{\mathrm{new}}, \Delta t) - R(x^{\mathrm{near}}, \Delta t)$, which is the hypersphere minus the *vesica piscis*.

a closed form solution that reflects Observation 3. We explicitly acknowledge that our assumptions are, similar to all assumptions, not uniformly true. We claim that for the class of systems considered here they are accurate enough to produce a closed form coverage model with high fidelity. Later, in Section 4 we examine their validity in practice.

Assumption 2. If x^{rand} is generated *inside* the explored set, $\mathcal{E}_K$, the *conditional expected value* of the growth is very small, $E(G_K|x^{\mathrm{rand}} \in \mathcal{E}) = g(\Delta t) \approx 0$.

Remark 2. Note that the assumption does not state that the growth is zero *every* time $x^{\mathrm{rand}} \in \mathcal{E}$. In particular we expect it to be non-zero when x^{near} supports the boundary of $\mathcal{E}$ (see Fig. 2 for a counterexample). Also, it implicitly accounts for the role of the current tree configuration and density function because it is conditional on x^{rand} and $\mathcal{E}$. It simplifies the analysis by making the *conditional expected value* $E(G_K|x^{\mathrm{rand}} \in \mathcal{E}_K)$ independent of C_{K-1}, x^{rand}, or Δt (if one assumes $g = 0$). The experimental results tend to validate the accuracy of this assumption (see Section 4).

Assumption 3. If x^{rand} is generated *outside* the explored set, $\mathcal{X} - \mathcal{E}$, we assume the conditional expected value of growth is larger and solely a function of the step size, $E(G_K|x^{\mathrm{rand}} \in \mathcal{X} - \mathcal{E}_K) = \hat{G}(\Delta t) \gg g(\Delta t)$. See Fig. 3 for an example.

Remark 3. While the growth at a given iteration is determined by a complex relationship between the location

of x^{rand}, x^{near}, the configuration of the tree, the Lipschitz constant, and the time step Δt, the assumption allows one, on average, to replace them with a simplified function $\hat{G}$ when $x^{\text{rand}} \in \mathcal{X} - \mathcal{E}$. Obviously, the assumption is not uniformly true, especially for systems that are not small-time locally controllable (see Fig. 2 for counterexamples). In general, $\hat{G}$ is directly proportional to Δt and the local Lipschitz constant of the dynamics $f(x, u)$ under the set of admissible inputs U. Increasing either has the effect of accelerating coverage.

It follows directly from Eq. (3.4) that

$$E(G_K | x^{\text{rand}}) \approx \begin{bmatrix} g(\Delta t), & \text{if } x^{\text{rand}} \in \mathcal{E} \\ \hat{G}(\Delta t), & \text{if } x^{\text{rand}} \in \mathcal{X} - \mathcal{E} \end{bmatrix}. \quad (3.5)$$

Therefore, our first main result is that the expected value of the growth at each iteration can be modeled as follows:

$$E(G_K | \mathcal{E}) \approx g \cdot Pr[x^{\text{rand}} \in \mathcal{E}] + \hat{G} \cdot (1 - Pr[x^{\text{rand}} \in \mathcal{E}]).$$

Assuming x^{rand} is generated using a uniform distribution (the most common choice in the literature by far), $Pr[x^{\text{rand}} \in \mathcal{E}] = C_{K-1}$ and

$$E(G_K | C_{K-1}) \approx g \cdot C_{K-1} + (1 - C_{K-1}) \cdot \hat{G}. \quad (3.6)$$

Note that this formula predicts that the tree grows quickly at first (small C_{K-1}) and more slowly as C_{K-1} approaches one, which supports Observation 3.

Our second main result concerns the variance of the growth. Using the fact that the total variance of two sub-populations can be computed as the variance of the sub-population means, plus the mean of the sub-population variances,

$$V(G_K | C_{K-1}) \approx (\hat{G} - g)^2 C_{K-1}(1 - C_{K-1})$$
$$+ C_{K-1} V(G | x^{\text{rand}} \in \mathcal{E})$$
$$+ (1 - C_{K-1}) V(G | x^{\text{rand}} \in \mathcal{X} - \mathcal{E}). \quad (3.7)$$

The first term is the variance of the conditional means. Our experiments suggest that this term is dominant. The second and third terms are mean of the conditional variances. We neglect their effects in the sequel. Under this simplification, the variance vanishes when $C_{K-1} = 0$ or $C_{K-1} = 1$, and attains its maximum value of $(\hat{G} - g)^2/4$ at $C_{K-1} = 0.5$. It predicts the variance to be higher for systems with large values of $\hat{G}$, either due to a large time step or the Lipschitz constant. Another observation is that the ratio $V(G_K | C_{K-1})/C_{K-1}$ monotonically decreases as new nodes are added to the tree.

Our third result provides an estimate of the expected value of coverage. Assume $g = 0$. Substituting Eq. (3.6) into Eq. (3.3), the expected value of C_K is

$$E(C_K) \approx E(C_{K-1}) + (1 - E(C_{K-1})) \cdot \hat{G}, \quad (3.8)$$

whose solution is

$$E(C_K) \approx C_1 (1 - \hat{G})^K + \hat{G} \sum_{k=0}^{K-1} (1 - \hat{G})^k. \quad (3.9)$$

Note that C_1 is completely determined by the initial condition and is not a random variable; however, further simplification is possible by assuming $C_1 \approx \hat{G}$, allowing the two terms to be condensed into a single geometric series. Simplifying yields the third result

$$E(C_K) \approx 1 - (1 - \hat{G})^K. \quad (3.10)$$

Observe that in accordance with previous results on probabilistic completeness,[25] $E(C_K) \to 1$ as $K \to \infty$. Obviously, increasing $\hat{G}$ has the effect of hastening coverage. We remark that a similar analysis, relaxing the assumption that $g = 0$, yields a solution

$$E(C_K) = \frac{\hat{G}}{\hat{G} - g} [1 - (1 - \hat{G} - g)^K], \quad (3.11)$$

which only converges to 1 if $g \to 0$ as $C \to 1$. While this is true in practice, we prefer the simplicity of Eq. (3.10).

Our fourth result predicts the expected value of the number of nodes, K, required to achieve a particular coverage level, C,

$$E(K) \approx \frac{\ln\left(\frac{\hat{G}}{\hat{G}-g} - C\right)}{\ln\left(\frac{\hat{G}}{\hat{G}-g}\left(1 - \hat{G} - g\right)\right)}. \quad (3.12)$$

Example: to achieve 95% coverage ($C = 0.95$), with $\hat{G} = 0.01$ and $g = 0$, the expected number of nodes is $K = 298$.

Our last result concerns the variance of coverage. Using the definition of the covariance one can show the relationship between the growth during iterations i and j to be $Cov(G_i, G_j) = -(\hat{G} - g)^3 C_{\min(i, j)}(1 - C_{\min(i, j)})$. Then using the fact that $C_K = \sum_{i=1}^{K} G_i$ and the variance of a sum is equal to the sum of entries in the covariance matrix,

$$V(C_K) \approx \frac{(\hat{G} - g)}{2} C_K^2 \cdot (1 - C_K), \quad (3.13)$$

suggesting the maximum variance is $2/27 \cdot (\hat{G} - g)$, which occurs at $C_K = 2/3$. These results are summarized in Table IV.

4. Computational Experiments

4.1. Holonomic systems in expansive spaces

The results in this section were obtained for an N-dimensional holonomic system with no obstacles. $\mathcal{X} = [-0.5, 0.5]^N \subset \mathbb{R}^N$ is the simply connected unit hypercube, the initial condition x^1 is the zero vector, $U = \{u \in \mathbb{R}^N | \|u\|_2 \leq u^{\text{max}}\}$, and $\dot{x} = u$.

4.2. Parameter computation

For holonomic, kinematic systems, $R(x^k, \Delta t)$ is simply a ball centered at x^k of radius $u^{\text{max}} \cdot \Delta t$. Regarding Eq. (3.6), in order to generate an easily computable approximation, we begin by assuming $g = 0$ and $\hat{G}$ is equal to the upper bound of G_K, which is the fractional volume of the shaded piece of the ball shown in Fig. 3. In arbitrary dimension N

Table I. Results of non-dimensionalized numerical integration used with Eq. (4.2) to compute $G^{\max}$ in dimensions 1 through 30.

N	1	2	3	4	5	6
α_N	0.5	0.609	0.688	0.747	0.793	0.830
N	7	8	9	10	20	30
α_N	0.859	0.883	0.902	0.918	0.985	0.997

the fractional volume of the shaded set can be computed via numerical integration using

$$G^{\max} = \frac{1}{\mu(\mathcal{R})}\left\{ \frac{\pi^{N/2}}{\Gamma(N/2+1)}(u^{\max} \cdot \Delta t)^N \right. \qquad (4.1)$$

$$\left. - 2\int_{u^{\max}\cdot\Delta t/2}^{u^{\max}\cdot\Delta t} \sin(\cos^{-1}(r))^{N-1}dr\right\},$$

where Γ is the generalization of the factorial function. The first term is $\mu(\mathbb{B}^N(u^{\max} \cdot \Delta t))$ – the volume of a hypersphere with radius $u^{\max} \cdot \Delta t$. The second term is twice the volume of the hypersphere cap formed by an intersecting hyperplane whose distance from the sphere's center is half the radius – sometimes referred to as the *vesica piscis*. We compute this quantity by numerical integration. Readers interested in reproducing these results can quickly compute $G^{\max}$ using

$$\mu(\mathbb{B}^N(u^{\max} \cdot \Delta t)) = \frac{\pi^{N/2}}{\Gamma(N/2+1)}(u^{\max} \cdot \Delta t)^N, \quad (4.2)$$

$$G^{\max} = \alpha_N \cdot \frac{\mu(\mathbb{B}^N(u^{\max} \cdot \Delta t))}{\mu(\mathcal{R})},$$

where α_N can be found in Table I, which summarizes the results of our unitless numerical integration of Eq. (4.1) in dimensions $N = 1, \ldots, 30$. Note, as $N \to \infty$, $\alpha_N \to 1$.

4.3. RRT implementation
A uniform distribution was used to select x^{rand}. In all cases the tree was allowed to grow until $C_K = 0.995$. A 4th-order Runge–Kutta integration method was used with variable step size and a tolerance of 10^{-8}. The Euclidean metric was used, and a closed-form optimal planner generated u^{new}. There are no subsequent references to run-times or memory usage statistics, making the results machine-independent. All subsequent reported results are averaged over 1000 trials unless otherwise stated.

4.4. Discretized coverage measure
While the notions of growth and coverage in Section 3.1 are useful for analysis, computing them exactly for a given tree is intractable, especially in the case of high-dimensional systems with complex dynamics. Instead an approximation is used (Fig. 4, top). Given a free state space, define a uniform, axis-aligned grid with spacing δx. Each grid cell is a closed set $X_{i,j,\ldots}$ indexed by an integer in each dimension. As each new node is added to the tree, a cell $X_{i,j,\ldots}$ is considered occupied if $\exists x \in R(\mathcal{T}_K, \Delta t) \bigcap X_{i,j,\ldots}$. The coverage C is approximated as the fraction of grid
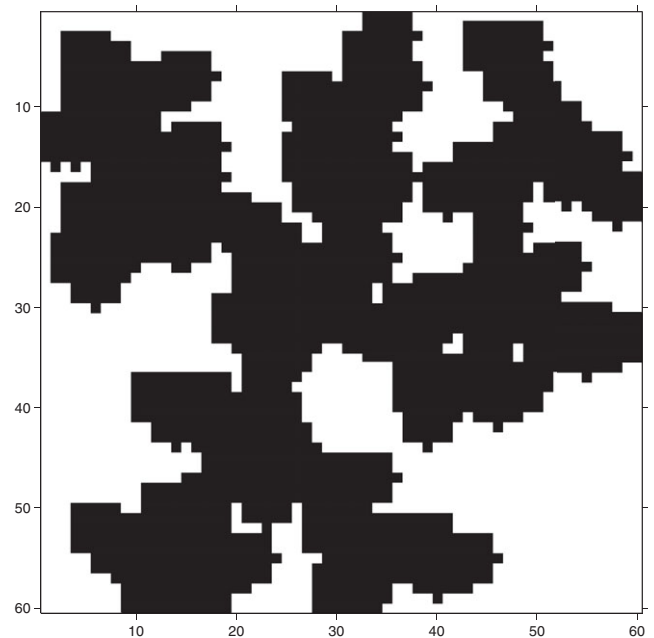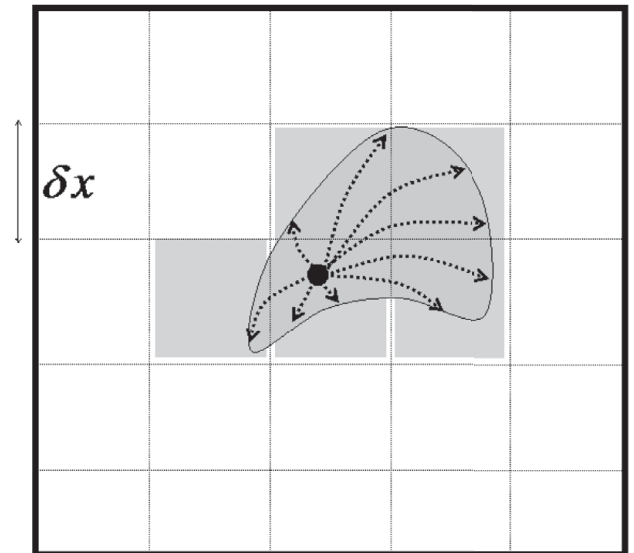

Fig. 4 (Top) An illustration of the grid-based approximate coverage measure. (Bottom) Visualization of the approximate explored set of a 2-dimensional tree (black). After 220 nodes are added, about 68 % of the space has been covered. A coarse discritization was used to illustrate the "pixelated" nature of the approximation.

cells which are occupied. For example, Fig. 4 (bottom) is a snapshot of the approximated $\mathcal{E}$ when $C_{220} = 0.68$. This is much less conservative, and easier to compute, than dispersion-based measures.[10,11]

4.5. Experimental results and model validity
Our initial experiments used $\Delta t \cdot u^{\max} = 0.05$, $N = 3$, and a grid size of $\delta x = 0.005$. Figures 5 and 6 compare the experimental mean and variance of the tree growth at each iteration with the values predicted by Eqs. (3.6) and (3.7). The value of $\hat{G}$ was assumed to be $G^{\max}$, and computed via Eq. (4.2) and Fig. 1, while g was assumed to be zero.
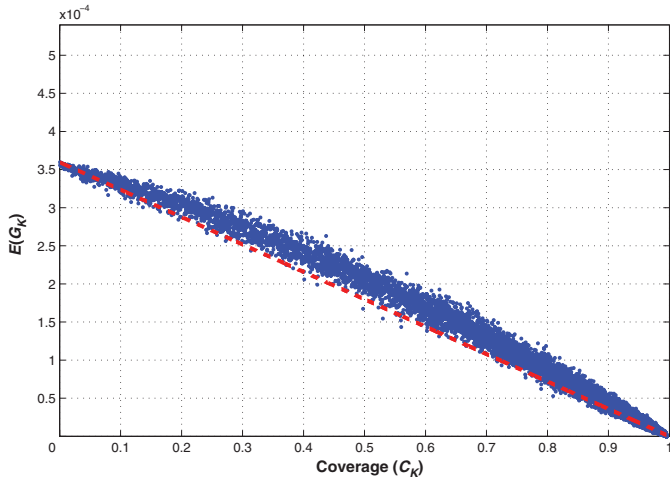
Fig. 5. (Colour online) The predicted $E(G_K)$ as a function of coverage (red dashed line). The experimental means of G_K are shown as blue dots.
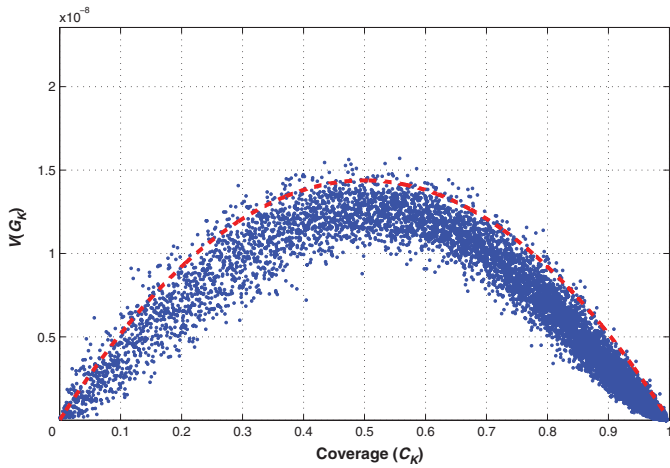

Fig. 6. (Colour online) The predicted $V(G_K)$ as a function of coverage (red dashed line). The experimental variances are shown as blue dots.

Fig. 7. (Colour online) (Top) Checking Assumption 2: $E(G_K|x^{\mathrm{rand}} \in \mathcal{E}) \approx 0$? The histogram of G_K when x^{rand} was generated *inside* the explored set indicates that the growth is indeed very small. (Bottom) Checking Assumption 3: $E(G_K|x^{\mathrm{rand}} \in \mathcal{X} - E) \approx \hat{G}$? The histogram of G_K when x^{rand} was generated *outside* the explored set indicates that the growth is significantly larger.

Regarding Assumption 2, $E(G_K|x^{\mathrm{rand}} \in \mathcal{E}) \approx 0$, consider Fig. 7 (top), which shows a histogram of experimental G_Ks when x^{rand} was generated *inside* the explored set. Each bin in the histogram includes a range of volumes equal to 10 grid cells. More than 1/3 of the time $G_K = 0$, making it the mode by a large margin; and the mean is 0.52×10^{-5} (variance 3.21×10^{-9}). Since the growth can never be negative, it is unlikely that the expected value would be zero in practice; however, it is indeed very small considering that G^{max} is 3.56×10^{-4}.

Regarding Assumption 3, $E(G_K|x^{\mathrm{rand}} \in \mathcal{X} - E) \approx \hat{G}$, consider Fig. 7 (bottom), which shows a histogram of actual G_Ks when x^{rand} was generated *outside* the explored set. The experimental mode is 3.60×10^{-4} and the mean is 2.78×10^{-4} (variance 8.93×10^{-9}). Recall G^{max} is 3.56×10^{-4}. More important is the relative scale of the growth when x^{rand} is inside versus outside the explored region. The fact that $E(G_K|x^{\mathrm{rand}} \in \mathcal{E}) \ll E(G_K|x^{\mathrm{rand}} \in \mathcal{X} - E)$ suggests that the assumptions in Section 3 are reasonable.

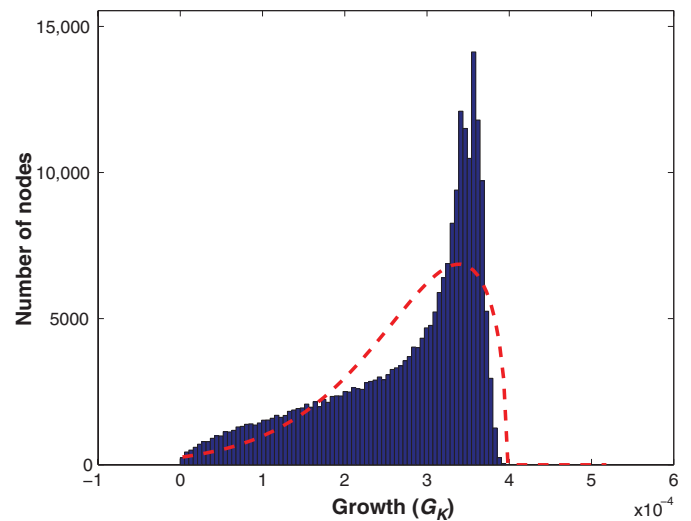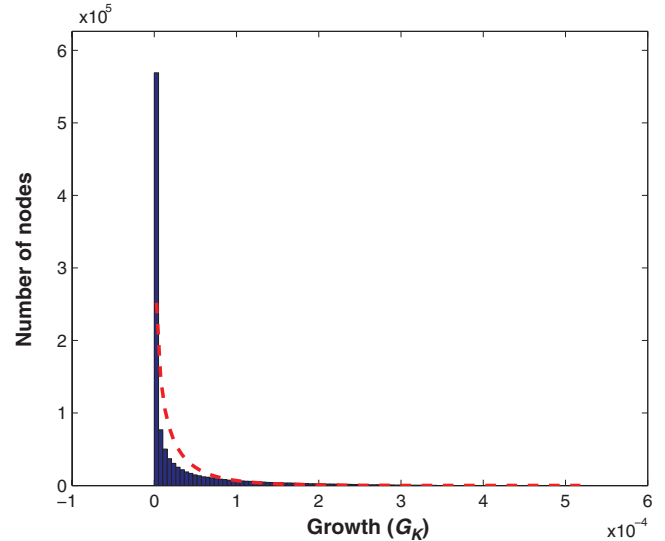One of our original goals was to identify the underlying distributions seen in Fig. 7 for the purpose of refining the model. Motivated by the fact that the selection of both x^{near} and u^{new} are minimizing in some sense (consider Algorithm 2), we focused on the family of extreme value distributions, including Beta, Weilbull, Gamma, Gumbel, Frechet, and Pareto. The red dashed lines in Fig. 7 depict scaled generalized extreme value (GEV) density functions plotted using the maximum likelihood estimates of distribution parameters. For each of these candidates, we applied the Chi-squared test to determine the likelihood that the data in Fig. 7 could have come from one of the distributions. To date none of the results have been statistically significant at the 0.05 level.

One observation worth mentioning is that, while the overall shape of histograms in Fig. 7 is fairly typical, the best-fit distribution parameters vary significantly as coverage evolves. For example, the ratio of the number of nodes in the "peak" of the histogram of $G_K|x^{\mathrm{rand}} \in \mathcal{E}$ (Fig. 7, top)
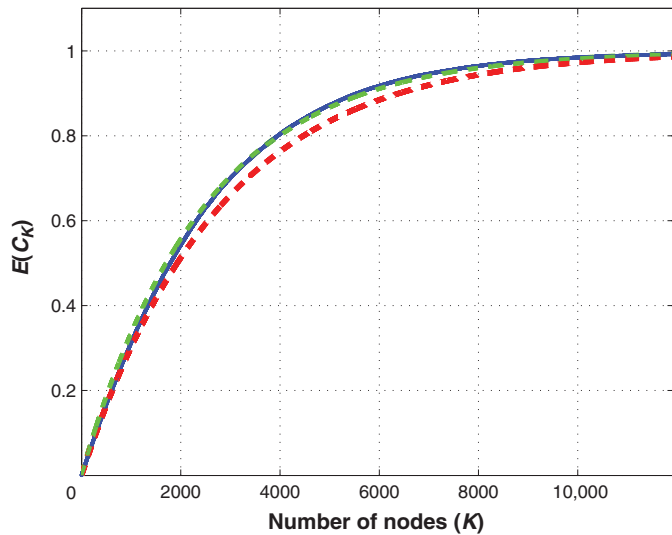
Fig. 8. (Colour online) Expected value of coverage. The blue dots show mean experimental coverage. Dashed red line shows expected value of coverage as predicted by Eq. (3.10) using $\hat{G} = G^{\text{max}}$. The dashed green line shows expected value of coverage as predicted by Eq. (3.10) using the best fit numerically computed value of $\hat{G}$.
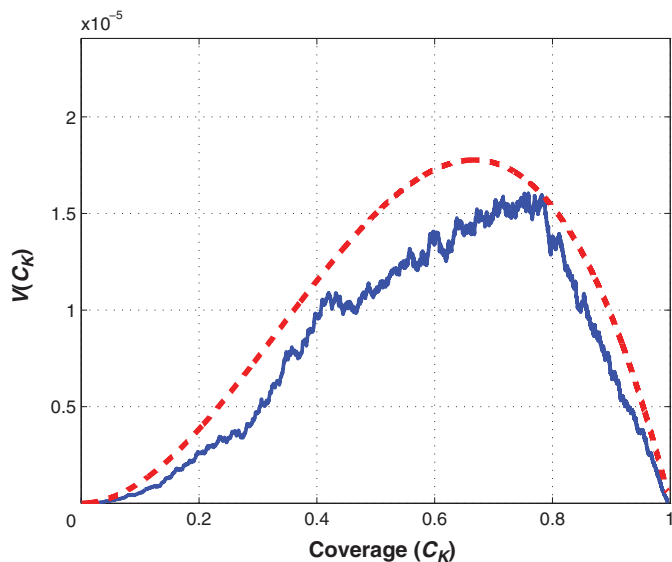

Fig. 9. (Colour online) Blue dots show experimentally determined variance of coverage $V(C_K)$ as a function of coverage. Red dashed line is the variance predicted by Eq. (3.13).

Fig. 10. (Colour online) Logarithmic curve fit. $-\ln(1 - C_K)$ should be linear in K. Note, the log is ill-conditioned as $C_K \to 1$.

0.989, indicating an excellent agreement. While intuitive to visualize this way, we caution that $-\ln(1 - C_K)$ is ill-conditioned as $C_K \to 1$ since $\ln 0$ is undefined. Instead, we use the fit of the raw data to Eq. (3.10) as discussed in the next paragraph as our primary accuracy measure.

To quantitatively test the accuracy of the coverage model, we used a trust-region-based nonlinear least-squares algorithm to numerically curve fit Eq. (3.10) to the experimental data with a relative tolerance of 10^{-10}, leaving $\hat{G}$ as a free parameter. The green dashed line in Fig. 8 was created with a computed fit-coefficient of $\bar{G} = (4.02 \pm 0.002) \times 10^{-4}$ (95% confidence). The "goodness of fit" was $r^2 = 0.9991$, indicating that the form of Eq. (3.10) described the shape of the coverage curve extremely well. For comparison, according to Fig. 3, $G^{\text{max}} = 3.56 \times 10^{-4}$.

Table II shows the result of similar analysis across examples with dimensions 2–8, holding $\Delta t \cdot u_{\text{max}} = 0.2$ constant. Note that $N = 7, 8$ use a coarser grid size due to memory limitations. First note that $r^2 \approx 0.999$ for all examples, indicating that the form of Eq. (3.10) holds well. Columns 4–6 show some dimensionless quantities that also speak to model fidelity. For a perfect model, $G^{\text{ext}}/\bar{G}$ would be 1 and $G^{\text{int}}/\bar{G}$ would be ≈ 0. An interesting trend is that $\bar{G} / G^{\text{max}}$ (column 4) gets smaller as N increases ($N = 7, 8$ are exceptions due to coarse grid size). One reason for this is that when x^{rand} is generated near the boundary of $\mathcal{E}$, the growth can be smaller than expected (Fig. 2, right). Note that in high-dimensional state spaces the volume of the boundary layer (thickness $\Delta t \cdot u^{\text{max}}$) becomes larger relative to the volume of the interior of $\mathcal{E}$.

Another issue that we investigated was the validity of Eq. (3.10) as $\Delta t \cdot u^{\text{max}}$ varies, which directly alters $\hat{G}$ and influences the rate at which the tree grows. Table III summarizes the results. First, note that for all trials $r^2 > 0.99$ indicating good agreement with the model for a wide range of step sizes. Second, as this parameter decreases, $G^{\text{ext}} = E(G_K | x^{\text{rand}} \notin \mathcal{E})$ increases relative to $\bar{G}$, indicating Assumption 3 becomes more accurate. However, the ratio in the final column increases slightly, implying that Assumption

as compared to the "tail" is much lower for the first few thousand nodes (small C_K), while the peak is significantly higher for the final few thousand nodes (C_K close to 1). This suggests that $g \to 0$ as $C_K \to 1$ as required by Eq. (3.11). Therefore, we do not expect that the distribution parameters are static, but rather a function of coverage.

Figures 8 and 9 show the experimental mean and variance of coverage as blue dots, while the red dashed lines show the values predicted by Eqs. (3.10) and (3.12) using G^{max}.

An interesting way to visualize the agreement of the model with the experimental data is to note that according to Eq. (3.10), the function $-\ln(1 - E(C_K))$ should vary linearly with K (solid line in Fig. 10). The dotted red line shows a linear fit to the log-transformed data, with $r^2 =$
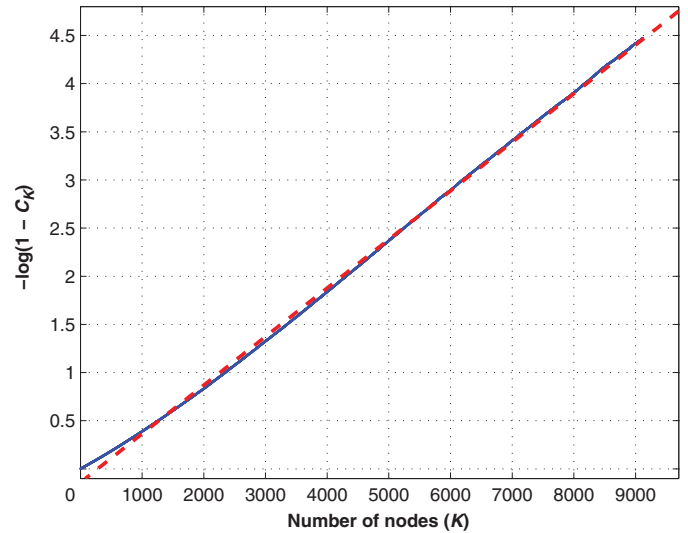
Table II. Experimental results for holonomic systems in expansive spaces. $\Delta t \cdot u_{\max} = 0.2$ and $\delta x = 0.066$ across $N = 2, \ldots, 6$. $N = 7, 8$ use a larger grid size (0.2) due to memory limitations. $G^{\text{ext}} \equiv E(G_K | x \notin \mathcal{E})$ and $G^{\text{int}} \equiv E(G_K | x \in \mathcal{E})$.

N	Nodes	r^2	$\frac{\bar{G}}{G^{\max}}$	$\frac{G^{\text{ext}}}{\bar{G}}$	$\frac{G^{\text{int}}}{\bar{G}}$
2	3.10×10^1	0.9991	1.2744	0.5859	0.1098
3	1.38×10^2	0.9995	1.0743	0.6035	0.0908
4	6.67×10^2	0.9997	0.8903	0.60732	0.1202
5	3.37×10^3	0.9993	0.8932	0.5748	0.1257
6	1.81×10^4	0.9983	0.7609	0.5622	0.1218
7*	2.76×10^5	0.9999	0.9831	0.9139	0.1232
8*	1.28×10^6	0.9995	0.8823	0.9948	0.0727

Table III. Experimental results for holonomic systems in expansive spaces. Here $N = 3$ as $\Delta t \cdot u^{\max}$ changes. Note $G^{\text{ext}} \equiv E(G_K | x \notin \mathcal{E})$ and $G^{\text{int}} \equiv E(G_K | x \in \mathcal{E})$.

$\Delta t u_{\max}$	Nodes	r^2	$\frac{\bar{G}}{\bar{G}}$	$\frac{G^{\text{ext}}}{\bar{G}}$	$\frac{G^{\text{int}}}{\bar{G}}$
0.200	151	0.999	1.097	0.6067	0.0884
0.100	1258	0.998	1.166	0.6380	0.1062
0.050	9127	0.996	1.236	0.6343	0.1209
0.025	69318	0.994	1.262	0.6335	0.1283

2 (no growth when $x^{\text{rand}} \in \mathcal{E}$) becomes weaker. Again, we feel this trend is related to the ratio of the boundary layer to the interior volume, which increases as $\Delta t \cdot u^{\max}$ becomes smaller. The surface area in Fig. 4 is much larger when this parameter is decreased. For large values the individual branches begin to merge and the surface area is smaller.

4.6. Concentration properties

While the mean and variance of coverage are important parameters, they are not sufficient to determine the shape of its probability distribution. One could use Chebyshev's inequality to produce a loose bound on the deviation from the expected value after a certain number of nodes have been added,

$$Pr\big[|C_K - E(C_K)| \leq m \cdot \sqrt{V(C_K)}\big] \geq 1 - 1/m^2, \quad (4.3)$$

where $m > 0$. However, considering that the coverage is a sum of many random variables, one would expect the distribution of coverage to be more tightly concentrated around its expected value; yet, most concentration inequalities are not applicable. For example, Hoeffding's inequality only applies to sums of independent random variables, while Azuma's only applies to martingales.

We grew 1000 RRTs using $\Delta t \cdot u^{\max} = 0.05$ in $\mathcal{X} \subset \mathbb{R}^3$, and a grid size of $\delta x = 0.005$. We observed that once $C_K > 0.03$, the distribution of coverage looked quite Gaussian (Fig. 11), and that approximately 68, 95, and 99% of the trees were within 1, 2, and 3 standard deviations, respectively, from the mean (Fig. 12) – suggesting much tighter concentration than that predicted by Chebyshev's inequality. In light of this we applied the Chi-squared goodness of fit test at each iteration to determine the probability the 1000 samples were drawn from a truncated Gaussian distribution (i.e., $0 \leq C_K \leq 1$). For 84% of the iterations the test was significant at the
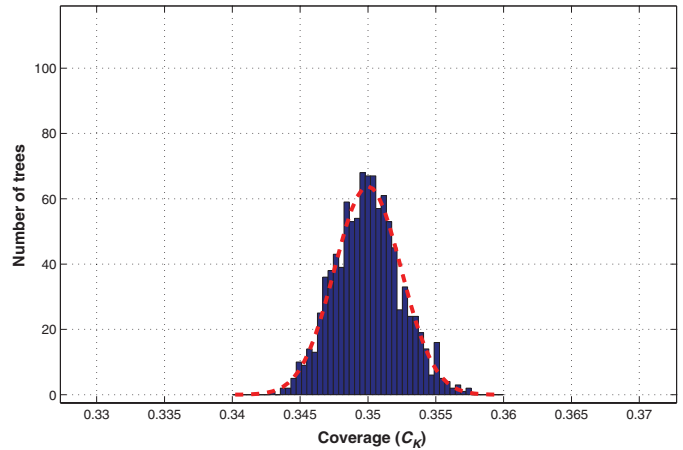

Fig. 11. (Colour online) A snapshot of the distribution of the coverage for 1000 trees. The red dashed line is the best fit truncated Gaussian distribution (scaled).
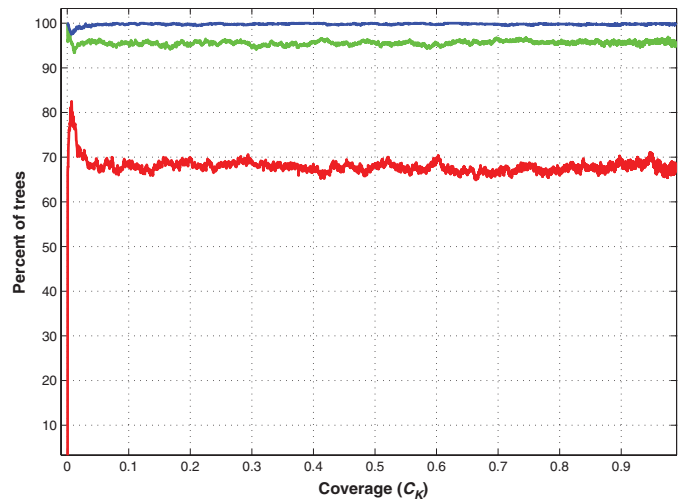

Fig. 12. (Colour online) Concentration about the mean. The percentage of 1000 trees with coverage within 1 (red), 2 (green), and 3 (blue) standard deviations of the mean as a function of coverage.

0.05 level, suggesting that the distribution of coverage is likely Gaussian. Most of the iterations that did not meet this criteria occurred when $C_K < 0.03$ or $C_K > 0.99$. We obtained similar results for problems with 2–6-dimensional state spaces and across a variety of step sizes. These findings lead to the conjecture that inequality 4.3 can be tightened to the well-know relation for Gaussian distributions,

$$Pr\big(|C_K - E(C_K)| \leq m \cdot \sqrt{V(C_K)}\big) = erf(m/\sqrt{2}).$$

In addition, consider that the ratio of the standard deviation to the expected value steadily decreases (see Eqs. (3.10) and (3.13) and Fig. 13), suggesting that, in absolute terms, the distribution becomes more heavily concentrated around the expected value as $C_K \to 1$.

5. Additional Examples

While the coverage model is designed for holonomic systems operating in obstacle-free workspaces, few problems of
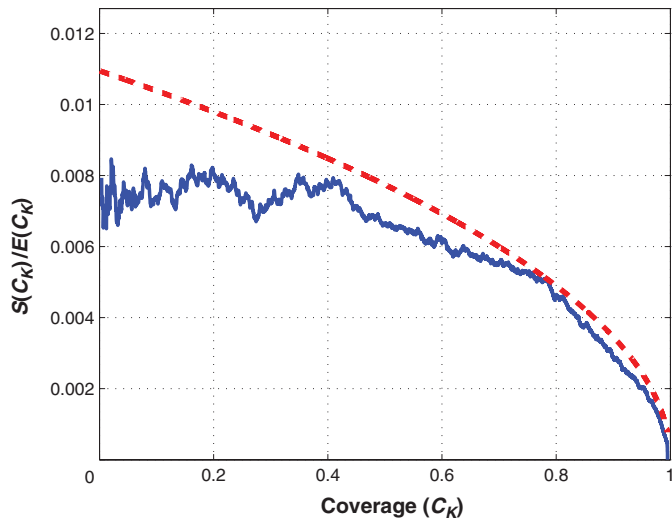
Fig. 13. (Colour online) The ratio of the standard deviation of coverage to the mean decreases as coverage approaches 1. Blue solid line is experimental measurement; and red dashed line is a predicted value.

interest fall into that class. Here we examine the model's applicability to other classes of systems.

5.1. Example 1: Expansive spaces with obstacles

Figure 14 depicts a unit cube state space (N = 3), whose middle section (width 0.2) is populated by randomly placed spherical obstacles of radius 0.025 (inspired by ref. [12]). But the proportion of the midsection covered by the obstacles varied between 0% and 50%. As the density increases, passing from the left half of $\mathcal{X}$ to the right may require traveling through passages as narrow as 0.025. However, because there are many such passages the workspace remains expansive.

Figure 14 shows experimental coverage averaged over 50 trials versus the predicted coverage (red dashed line). We used $\Delta t \cdot u^{\max} = 0.05$ and $\delta x = 0.005$. Note that to a certain extent the model is able to account for the presence of obstacles by reducing the denominator in Eq. (3.1). Because $\Delta t \cdot u^{\max}$ is held constant but the volume of the free space is reduced, $\hat{G}$ is effectively increased when more obstacles are present – hastening coverage. Apparently the predictive power of the model is not significantly diminished compared with the 0% (no obstacles) example. Also, note that the histograms (not shown) are virtually identical to the ones shown in Fig. 7 – implying Assumptions 2 and 3 are not adversely affected.

5.2. Example 2: Poorly expansive spaces

Problems with a single long narrow passages are much more likely to weaken Assumption 3. Consider the problem in Fig. 15 (top), where two large open regions within the unit cube are connected by a very long narrow tube of diameter, GapWidth, and length, GapLength = 0.6 (also inspired by ref. [12]).

Figure 15 (bottom) shows the coverage as we vary GapWidth. First note that we saw no appreciable change in expected coverage when 0.16 ≤GapWidth≤ 1.0. When
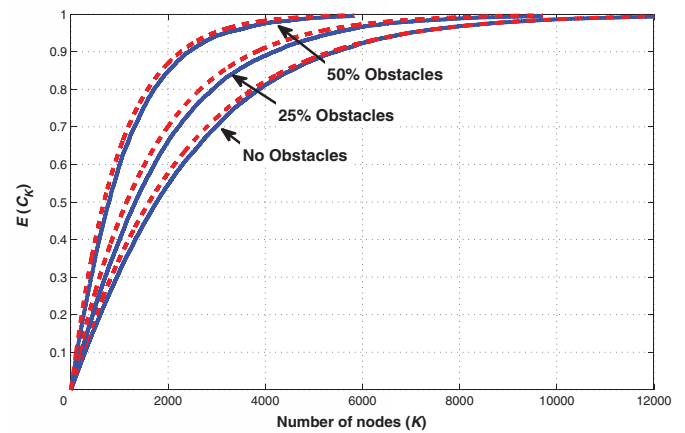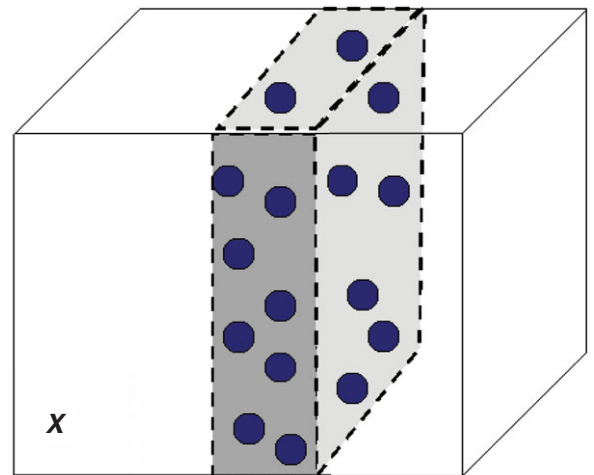

Fig. 14. (Colour online) (Top) An expansive workspace filled with randomly generated spherical obstacles. (Bottom) Actual coverage (blue line) vs. predicted coverage (red dashed) for obstacle densities of 0, 25, and 50%.

0.04 ≤GapWidth≤ 0.16, the model fit is still very high, with $r^2 \approx 0.998$, however the growth rate, $\hat{G}$, is effectively reduced by up to 15%. This is explained by examining the histogram of $G_k|x^{\mathrm{rand}} \in \mathcal{X} - \mathcal{E}$ (Fig. 16), which is now bi-modal. It resembles a mixture of Fig. 7 (top and bottom), with an effectively reduced mean. Clearly, the last term in Eq. (3.7) is no longer negligible, suggesting that the variance is much higher than originally predicted.

When GapWidth≤ 0.04, the fit to the model is qualitatively poor ($r^2 \leq 0.8$). Rather than looking at the expected coverage, Fig. 17 shows coverage for 20 individual trials, for the extreme example of GapWidth= $\delta x = 0.005$. The growth actually takes place in two separate phases. The first phase corresponds to the tree filling the free space to the left of the passage, where the initial condition is located (see Fig. 18, top). The number of nodes required to find the passage varies considerably among the trials – accounting for the variation in the length of the "tail" of the first phase. The second growth phase corresponds to the filling of the free space to the right of the passage. In the best case scenarios, when the passage is found quickly, the original coverage model is still remarkably accurate (upper red dashed line,
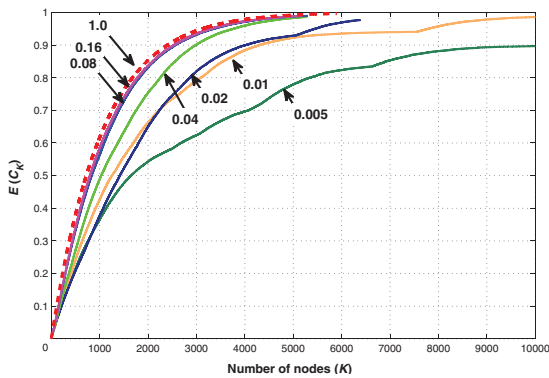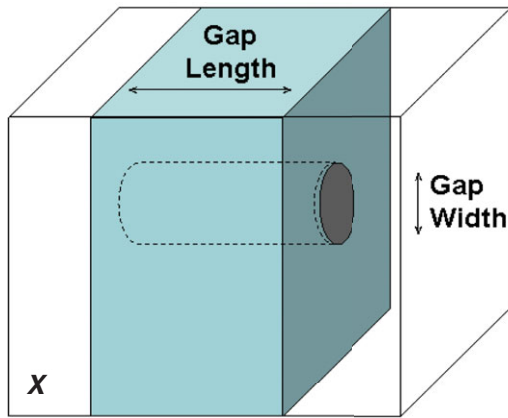
FIg. 15. (Colour online) (Top) A non-expansive workspace. (Bottom) Actual mean coverage (solid) for a variety of GapWidths vs. predicted coverage (red dashed line).
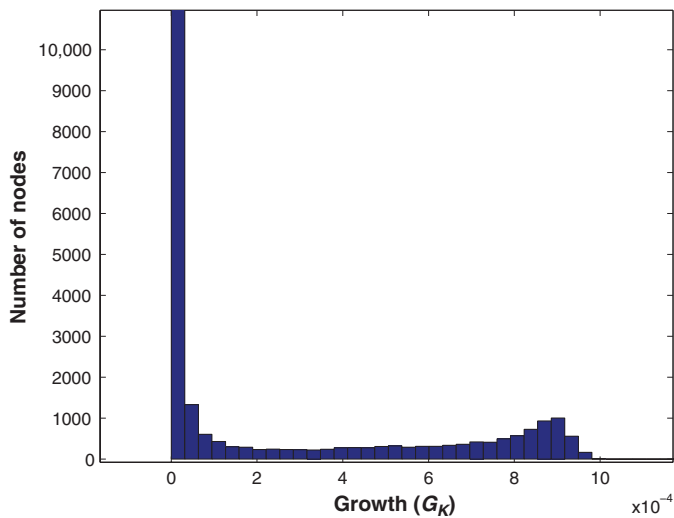

Fig. 16. (Colour online) A histogram of $G_k|x^{\text{rand}} \in \mathcal{X} - \mathcal{E}$ for a GapWidth = 0.005.

Fig. 17). However, in the worst case scenarios, an extended model is needed.

The lower dashed green line in Fig. 17 is the coverage predicted by two sequential applications of Eq. (3.10). Let $\mathcal{X}_L$ denote the expansive portion of the workspace to the left of the passage, where the tree is rooted; and $\mathcal{X}_R$ denote the expansive portion to the right. Using Eq. (3.12), compute K_L – the number of nodes required to achieve sufficient
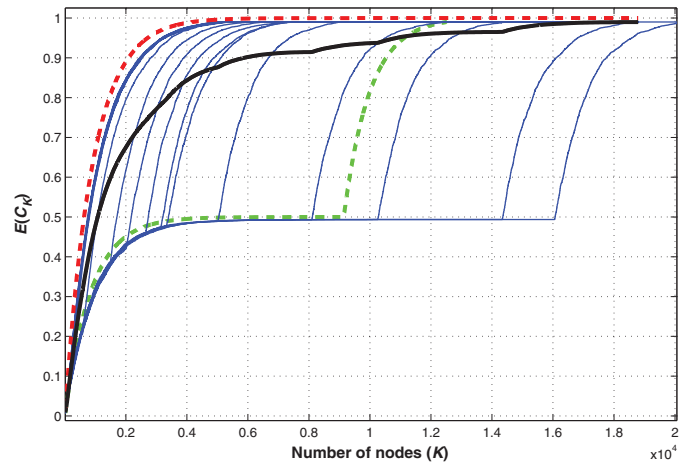

Fig. 17. (Colour online) Blue lines are coverage as a function of number of nodes for 20 individual RRTs, applied to the problem in Fig. 15, with GapWidth = 0.005. Mean coverage is shown in black. Red dashed line is the expected value predicted by Eq. (3.10). Green dashed line is coverage predicted by sequential application of Eq. (3.10) in each expansive component of the workspace.

coverage of X_L so that the expected number of nodes in the lookout set of $\mathcal{X}_R$ is 1. Then ignoring the volume of the narrow passage,

$$E(C_K) =$$
$$\begin{cases} \mu(\mathcal{X}_L) \cdot \left(1 - \left(1 - \frac{\hat{G}}{\mu(\mathcal{X}_L)}\right)^K\right) & \text{if } K < K_L \\ \mu(\mathcal{X}_L) \cdot \left(1 - \left(1 - \frac{\hat{G}}{\mu(\mathcal{X}_L)}\right)^K\right) + \mu(\mathcal{X}_R) \cdot \\ \left(1 - (1 - \frac{\hat{G}}{\mu(\mathcal{X}_R)})^{K-K_L}\right) & \text{if } K \geq K_L. \end{cases}$$

This extension of the model is naive in many respects. For example, it ignores the shape and length of the passage; and it is not clear how to extend it to the case where there are more than two expansive components connected in a cyclic fashion. However, Fig. 17 shows that this model does qualitatively reproduce the twice exponential coverage trend in many of the trials. We chose not to develop this further because models that require in-depth knowledge of solution parameters, such as passage length, diameter, or connectivity, are antithetical to our goal of finding a simple coverage formula of practical significance. However, it is a topic of ongoing work.

5.3. Example 3: Non-holonomic

Consider the classic unicycle model of a non-holonmic mobile robot where $\dot{x_1} = u_1 \cos x_3$, $\dot{x_2} = u_1 \sin x_3$, and $\dot{x_3} = u_2$ and $u_1 \in [-1, 1]$ and $u_2 \in [-\bar{\omega}, \bar{\omega}]$. For this example we used the (imperfect) metric

$$\rho(x^A, x^B) = \sqrt{\left(x_1^A - x_1^B\right)^2 + \left(x_2^B - x_2^B\right)^2}$$
$$+ \frac{1}{\pi} \cdot \min\left(|x_3^A - x_3^B|, 2\pi - |x_3^A - x_3^B|\right),$$
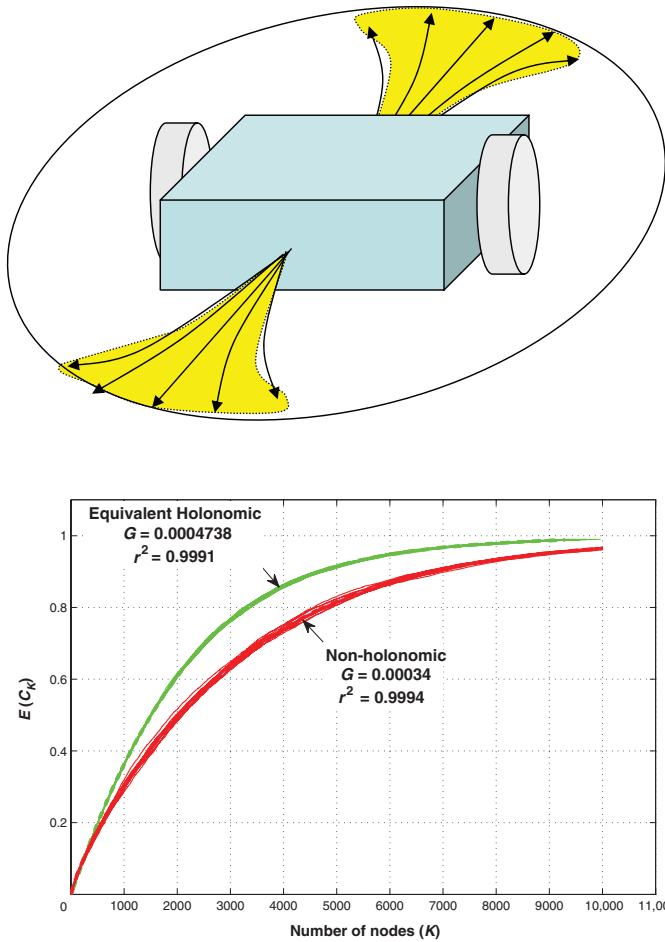
$$(5.1)$$

Fig.18. (Colour online) (Top) A differential drive example with bounded angular velocity. (Bottom) 3-dimensional non-holonomic comparison.

and use an arc search algorithm[27] as the local planner to connect x^{near} and x^{rand}.

We applied our coverage model to the non-holonomic system and compared it to a holonomic system whose connection sets have identical volume. On a positive note, we found that the coverage model remains qualitatively valid for a wide range of $\bar{\omega}$ values. Consider Fig. 18 (bottom), leaving $\hat{G}$ as a free parameter, the fit of Eq. (3.10) to the non-holonomic coverage data (red lines) is quite good ($r^2 > 0.99$).

However, the growth rate is reduced from the equivalent holonomic system (green lines) by about 25%. Note that the connection set is bowtie-shaped as shown in Fig. 18 (top; vs. circular in Fig. 3). A theoretical result from the study of *stochastic coverage processes*[9] suggests that the shape and orientation of the connection sets should not impact the expected value of coverage; however, our initial results indicate this is not true for the RRT. An investigation of the histogram of growth reveals a situation similar to the moderately narrow passage examples in Fig. 16 – a bimodal histogram.

6. Conclusions and Future Work

In this paper we provided a stochastic difference equation which models the fractional volume of the configuration space covered by the forward search RRT algorithm as a function of the number of nodes in the tree. We model the growth at each iteration as a random variable drawn from one of the two distinct distributions – termed as the CDG model. Under two simplifying assumptions, the model permits simple closed-form approximations of the expected value and variance of the coverage. The expected value of coverage follows an exponential convergence pattern, while the variance peaks when 2/3 of the state space has been explored then decays to zero. The model only depends on a single rate constant; and we provided a formula and table to compute it for systems with single integrator dynamics, using the time-step and a Lipschitz constant – two easily obtainable parameters. The model recovers many of the properties observed by users of the RRT such as probabilistic completeness, and rapid initial exploration followed by a slower exploration phase. Our results are summarized in Table IV.

When applied to holonomic systems operating in expansive configuration spaces, the computational experiments and model exhibit strong agreement across a variety of dimensions and parameter choices. The presence of obstacles alone does not appear to invalidate the model. Furthermore, we presented experimental evidence that the probability distribution of coverage is highly concentrated around its expected value for this class of problems. If many trees are grown, each with an identical number of nodes, the distribution of their coverage is highly Gaussian, suggesting that one could place a tight confidence interval on the expected coverage.

When applied to an example with moderately narrow passages, the model for the expected value of coverage is a good representation; however, its predictive power gradually degrades as the passageway becomes *extremely* narrow. For

Table IV. Summary of results.

Quantity	Function	Comment
$E(G_K)$	$g \cdot C_K + \hat{G}(1 - C_K)$	From Assumptions 2 and 3
$V(G_K)$	$(\hat{G} - g)C_K \cdot (1 - C_K)$	Assuming conditional variances are negligible.
$E(C_K)$	$\frac{\hat{G}}{\hat{G}-g}\left(1 - (1 - \hat{G} - g)^K\right)$	Where $g \to 0$ as $C \to 1$, simplify by assuming $g \equiv 0$.
$V(C_K)$	$\frac{(\hat{G}-g)}{2}C_K^2 \cdot (1 - C_K)$	Assuming conditional variances are negligible.
$E(K)$	$\frac{\ln(\frac{\hat{G}}{\hat{G}-g} - C)}{\ln\left(\frac{\hat{G}}{\hat{G}-g}(1-\hat{G}-g)\right)}$	Assume $g = 0$ if $C \approx 1$.
$Pr[|C_K - E(C_K)| \le m \cdot \sqrt{V(C_K)}]$	$erf(m/\sqrt{2})$	Chi-squared test suggests coverage is Gaussian.

poorly expansive configuration spaces the expected value behaves in a qualitatively similar way; but the growth rate is reduced and the variance of coverage becomes quite high as the histogram of growth per iteration becomes bi-modal. When the model was applied to a non-holonomic system, similar behavior was detected. These examples shed light on a previous observation[28] that the variance of coverage (and runtime) can be quite high for poorly expansive spaces. Our initial results suggest that it is possible to extend the model to account for poorly expansive workspaces (the model can be sequentially applied to each expansive component of the workspace); but it requires *a priori* knowledge of several workspace parameters and is therefore of limited practical value.

While the problem class of expansive, holonomic systems is admittedly quite limited, it has been argued that most workspaces of practical interest are indeed expansive.[14] In addition, we argue that this result is important on several grounds. First, it addresses some previously open foundational questions posed in ref. [25]. Second, we feel that one of the strengths of the model lies in its computational and informational simplicity. Estimating the coverage only requires a handful of floating point operations. In addition, all the parameters of the model, when applicable, are easily computable, and do not require *a priori* knowledge of the solution. Third, from a practitioner's perspective, the formula can be used to provide estimates of the number of nodes required to fill a specified percentage of the space (ex. 1500 nodes for 99.9% coverage with 95% confidence). To our knowledge this is the first RRT termination criteria developed using a formal approach. Fourth, many adaptive variants of the RRT algorithm have been introduced in the literature (e.g., refs. [1, 5, 19, 21]) for the purpose of addressing difficult problems that may contain narrow passages. Our simple formula could be used as a baseline metric of just *how difficult* a problem is, by examining the lag in predicted versus actual coverage. For example, if after 1500 nodes the model predicts 99.9% coverage, but only 50% of the space has been covered, it is highly unlikely the configuration space is expansive.

Possible directions of future work include formalizing the extended model of coverage (Eq. 5.2) for systems with narrow passages – in particular to model the variance; estimating the nature of the distributions seen in Fig. 7 for the purpose of developing a mixture model for the bi-modal distribution in Fig. 16; and developing an analogous model for Probabilistic Road Map planners.

References

1. T. Simeon, A. Yershova, L. Jaillet and S. M. LaValle, "Dynamic-Domain RRTs: Efficient Exploration by Controlling the Sampling Domain," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, Barcelona, Spain (2005) pp. 3856–3861.
2. N. M. Amato and Y. Wu, "A Randomized Roadmap Method for Path and Manipulation Planning," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis, MN, USA (May 1996) pp. 113–120.
3. R. Bohlin and L. E. Kavraki, "Path Planning Using Lazy PRM," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, CA (2000) pp. 521–528.
4. M. S. Branicky, S. M. LaValle, K. Olson and L. Yang, "Quasi-Randomized Path Planning," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, Seoul, Korea (May 2001) pp. 1481–1487.
5. B. Burns and O. Brock, "Single-Query Motion Planning with Utility-Guided Random Trees," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy (2007) pp. 3307–3312.
6. J. F. Canny, *The Complexity of Robot Motion Planning* (MIT Press, Cambridge, Massachusetts, 1988).
7. P. Cheng and S. M. LaValle, "Resolution complete rapidly-exploring random trees," **In**: *Proceedings IEEE International Conference on Robotics and Automation*, Washington DC, USA (2002) pp. 267–272.
8. E. Frazzoli, M. A. Dahleh and E. Feron, "Real-time motion planning for agile autonomus vehicles," *J. Guid. Control Dyn.* **25**(1), 116–129 (Jan–Feb, 2002).
9. P. Hall, *Introduction to the Theory of Coverage Processes* (Wiley, New York, 1988).
10. J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numer. Math.* **2**(1), 84–90 (1960).
11. J. M. Hammersley, "Monte-Carlo methods for solving multivariable problems," *Ann. New York Acad. Sci.* **86**, 844–874 (1960).
12. D. Hsu, "Randomized Single-Query Motion Planning in Expansive Spaces," *Ph.D. thesis* (Stanford University, California, 2000).
13. D. Hsu, L. E. Kavraki, J. C. Latombe, R. Motwani and S. Sorkin, "On Finding Narrow Passages with Probabalistic Roadmap Planners," **In**: *Proceedings of the 3rd Workshop on the Algorithmic Foundations of Robotics (WAFR)*, Houston, TX (1998) pp. 141–153.
14. D. Hsu, J.-C. Latombe and H. Kurniawati, "On the probabilistic foundations of probabilistic roadmap planning," *International Syposium on Robotics Research*, San Francisco, CA, USA (2005).
15. S. Karaman and E. Frazzoli, "Sampling-based algorithms for optimal motion planning," *Int. J. Robot. Res.* **30**(7), 846–894 (June 2011).
16. L. E. Kavraki, M. N. Kolountzakis and J. C. Latombe, "Analysis of probabilistic roadmaps for path planning," *IEEE Trans. Robot. Autom.* **14**(1), 166–171 (1998).
17. L. E. Kavraki, P. Svestka, J. C. Latombe and M. H. Overmars, "Probabilistic roadmaps for path planning in high dimensional configuration space," *IEEE Trans. Robot. Autom.* **12**, 566–580 (1996).
18. J. Kim, J. Keller and V. Kumar, "Design and Verification of Controllers for Airships," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV, USA (Oct. 2003).
19. J. W. Kim, J. M. Esposito and V. Kumar, "RRT enhancements," *Int. J. Robot. Res.* **15**(12), 1257–1272 (2006).
20. A. Ladd and L. E. Kavraki, "Measure theoretic analysis of probabilistic path planning," *IEEE Trans. Robot.* **20**(2), 229–242 (2004).
21. A. M. Ladd and L. E. Kavraki, "Fast Tree-Based Exploration of State Space for Robots with Dynamics," **In**: *Proceedings of the Workshop on the Algorithmic Foundations of Robotics*, Utrecht, Netherlands (2005) pp. 297–312.
22. S. M. LaValle, M. S. Branicky and S. R. Lindemann, "On the relationship between classical grid search and probabilistic roadmaps," *Int. J. Robot. Res.* **23**, 673–692 (2004).
23. S. M. LaValle, *Planning Algorithms* (Cambridge University Press, Cambridge, UK, 2006).
24. S. M. LaValle and J. J. Kuffner, "Randomized kinodynamic planning," *Int. J. Robot. Res.* **20**(5), 378–400 (2001).
25. S. M. LaValle and J. J. Kuffner, "Rapidly exploring random trees: Progress and prospects," **In**: *Algorithmic and Computational Robotics: New Directions* (B. Donald, K. Lynch and D. Rus, eds.) (A.K. Peters, Wellesley, MA, 2001) pp. 293–308.

26. S. R. Lindemann and S. M. LaValle, "Incrementally Reducing Dispersion by Increasing Voronoi Bias in RRTs," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, New Orleans, LA, USA (2004) pp. 3251–3257.

27. F. von Hundelshausen, M. Himmelsbach, F. Hecker, A. Mueller and H. Wuensche, "Driving with tentacles: Integral structures for sensing and motion," *J. Field Robot*. **29**(9), 640–665 (2008).

28. N. A. Wedge and M. S. Branicky, "On Heavy-Tailed Runtimes and Restarts in Rapidly Exploring Random Trees," *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, Chicago, Illinois (2008).