Resolution-Optimal Motion Planning for Steerable Needles

Mengyu Fu¹, Kiril Solovey², Oren Salzman², and Ron Alterovitz¹

Abstract—Medical steerable needles can follow 3D curvilinear trajectories inside body tissue, enabling them to move around critical anatomical structures and precisely reach clinically significant targets in a minimally invasive way. Automating needle steering, with motion planning as a key component, has the potential to maximize the accuracy, precision, speed, and safety of steerable needle procedures. In this paper, we introduce the first resolution-optimal motion planner for steerable needles that offers excellent practical performance in terms of runtime while simultaneously providing strong theoretical guarantees on completeness and the global optimality of the motion plan in finite time. Compared to state-of-the-art steerable needle motion planners, simulation experiments on realistic scenarios of lung biopsy demonstrate that our proposed planner is faster in generating higher-quality plans while incorporating clinically relevant cost functions. This indicates that the theoretical guarantees of the proposed planner have a practical impact on the motion plan quality, which is valuable for computing motion plans that minimize patient trauma.

I. INTRODUCTION

Medical steerable needles have the potential to improve patient care in diagnostic and therapeutic procedures including biopsy, localized drug delivery, and radioactive seed implantation for cancer treatment [2]. Steerable needles have a small diameter and are made of a highly flexible material, which allows them to follow 3D curvilinear trajectories inside the tissue. These properties enable steerable needles to move around critical anatomical structures to reduce patient trauma and reach sites previously unreachable with traditional straight needles [3–6].

Automating needle steering can improve the accuracy, precision, speed, and safety of steerable needle procedures. Automating these procedures can also facilitate their broad use, since manual control of a steerable needle is challenging due to the nonholonomic constraints on the needle’s motion and the high level of precision required to operate it. A key component of automating steerable needle procedures is motion planning: computing feasible, obstacle-avoiding trajectories through the tissue to reach a target. The trajectory of the needle through tissue should also maximize patient safety, which can be quantified using metrics such as minimizing trajectory length [7], maximizing clearance from obstacles [8–11], and minimizing damage to sensitive tissue [12]. An example is shown in Fig. 1.

A motion planner for steerable needles should ideally offer guarantees on both finite-time completeness (i.e., compute in finite time a motion plan or indicate that none exists) and optimality (i.e., return a globally optimal motion plan with respect to a chosen cost metric). Most prior motion planners for steerable needles lack one or both of these criteria. For instance, some methods for steerable needle motion planning lack completeness guarantees [7, 13–18], and so may fail to find a motion plan when one exists. Some methods do aim to optimize motion plan cost but they lack global optimality guarantees [7, 18, 19].

Some sampling-based planners are known to be both complete and optimal, albeit those properties are usually proven only for an asymptotic regime where the number of samples tends to infinity [20–27]. Thus, it is unclear what should be the number of samples necessary to achieve those guarantees in practice. Recent work has developed optimality guarantees for finite sampling, although those results cannot be currently applied to steerable needles as they deal with holonomic systems [28, 29].

In this paper, we introduce the first motion planner for steerable needles that offers excellent practical performance in terms of runtime while providing strong theoretical guarantees on completeness and the cost of the motion plan in finite time. In particular, we consider a specific type of optimality in relation to the motion plan cost—resolution...
optimal. Generally speaking, a resolution characterizes the discretization of some space (e.g., state space, configuration space, action space, and time). An algorithm is resolution complete if there exists a fine-enough resolution with which the algorithm finds a motion plan in finite time when a qualified motion plan exists, and otherwise correctly returns that no such plan exists [30]. An algorithm is resolution optimal if it is resolution complete and if, when it does return a motion plan, the plan’s cost is guaranteed to be within a desired approximation factor of the cost of a globally optimal qualified motion plan.

Our new motion planner builds on Resolution-Complete Search (RCS) [30], which is a resolution-complete but not resolution-optimal motion planner for steerable needles. If a motion plan exists, RCS would find a motion plan in finite time assuming that the parameter resolution is fine enough, but it provides no guarantees on the motion plan cost. To achieve resolution optimality, we enhance RCS with cost-aware duplicate pruning while incorporating motion plan cost tracking and a heuristic function to improve efficiency. We provide a proof sketch to show the resolution optimality of our method with a careful discussion of assumptions and required conditions. We also demonstrate experimentally on a realistic lung biopsy scenario that our new method, RCS*, outperforms the state-of-the-art in terms of runtime and plan quality.

II. RELATED WORK

A. Motion planning for steerable needles

A variety of approaches have been proposed for the motion planning of steerable needles. Duindam et al. [13] proposed a planner based on inverse kinematics but provided no theoretical guarantees. Liu et al. [18] developed the Adaptive Fractal Tree (AFT) for needle steering. Their method iteratively refines the lowest-cost plan from the previous iteration, but refining the best plan of a coarse resolution does not necessarily lead to the best plan in a finer resolution. Pinzi et al. [19] extended it to account for goal orientation constraints.

Some planners adapt sampling-based methods such as Rapidly-exploring Random Tree (RRT) [20] for steerable needles. Xu et al. [31] used an RRT variant for needle steering but showed low time efficiency. Patil et al. [15] developed an RRT-based needle planner that samples in the 3D workspace rather than the configuration space. Sampling in a lower-dimension space and their customized distance function made the planner work efficiently in many practical cases, but this also invalidates the probabilistic-completeness guarantee of RRT [20, 23]. To avoid dealing with curvature constraints directly, Favaro et al. [7] proposed a hybrid method to combine sampling and smoothing. First, a tree embedded in the 3D workspace is built with RRT* [24], then candidate plans found by the tree are smoothed to further account for the curvature constraint. However, such a decoupling invalidates the asymptotic optimality guarantee [24, 25]. Sun et al. [27] proposed a needle planner by building multiple RRTs, which is asymptotically optimal when the number of trees tends to infinity. Other methods focus on accounting for the uncertainty during needle insertion without providing formal guarantees [14, 16, 17].

B. Resolution-optimal motion planners

Although resolution completeness has been frequently mentioned [32–36], resolution optimality earned little attention, possibly due to being rather complex to analyze mathematically, particularly for nonholonomic systems. Consequently, many planners developed for nonholonomic systems focus on asymptotic optimality instead [22, 26, 37, 38].

Barraquand et al. [33] proposed a resolution-complete planner for single or multiple robots with nonholonomic constraints. Their method is also optimal with respect to the number of reverse maneuvers in the plan. Pivtoraiko et al. [39] proposed the idea of motion planning using state lattices for field robots. Their state lattices planner is resolution optimal since the search is optimal for a graph of some resolution and the discrete state grid approximates the continuous space as resolution increases. Ljungqvist et al. [40] later extended [39] for a general two-trailer system in 2D. However, these methods are designed for large-scale workspaces, making them unsuitable for tasks where a high level of precision is required, such as for steerable needles.

III. PROBLEM DEFINITION

We consider bevel-tip flexible steerable needles [3–6] controlled by insertion and rotation at their base. A bevel-tip steerable needle is made of a highly flexible material such that, when being inserted through tissue, asymmetric forces applied by the bevel cause the needle to take a curved trajectory. The maximum curvature of the needle’s path in the tissue, κ_{max}, is influenced by the mechanical design of the needle and the tissue that the needle moves through. Additionally, axially rotating the needle at its base changes the direction the bevel is facing, enabling us to control the steering direction. The figure to the right illustrates the needle’s kinematics, where the needle can be inserted by length ℓ and axially rotated at its base by θ.

We make the common assumption that the steerable needle is sufficiently flexible so the needle shaft moves along the trajectory created by its tip while the lateral motions are negligible. Thus, the configuration space of the steerable needle is defined by the pose of its tip, X ⊂ SE(3), where a given configuration x = (p, q) ∈ X, specifies the position and orientation components p ∈ R^3 and q ∈ SO(3), respectively. We assume that X is compact. We denote the 3D workspace by W ⊂ R^3, a subset of which is occupied by obstacles W_{obs} ⊂ W. A configuration x = (p, q) is collision free if and only if p /∈ W_{obs}. We define X_{free} as the union of all collision-free configurations.

A motion plan of the needle is a trajectory σ : [0, ℓ] → X, where ℓ is the length of the trajectory. We also use the notation ℓ_σ to denote the length of σ. A motion plan (or trajectory) σ is collision free if ∀s ∈ [0, ℓ], σ(s) ∈ X_{free}. To evaluate the quality of a motion plan, we consider a configuration-based cost function c : X → R. We require c to be well behaved (formally defined in Sec V), which includes being Lipshitz continuous and bounded within [c_{min}, c_{max}].
We define the cost of a motion plan as the integral of the configuration-based cost along a given trajectory $\sigma$, i.e., $\mathcal{C}(\sigma) = \int_0^1 c(\sigma(s)) ds$. This definition captures a variety of cost functions, including trajectory length and integrating over a cost map derived from medical images.

A motion plan is (kinematically) feasible if the curvatures along the trajectory never exceed $\kappa_{\text{max}}$. A motion plan is valid if it is collision free and feasible. We are now ready to state the steerable needle motion planning problem.

**Problem 1.** An optimal steerable needle motion planning problem is defined as the tuple $\Delta = (X, \mathcal{W}_{\text{obs}}, x_{\text{start}}, p_{\text{goal}}, \tau, \ell_{\text{max}}, \kappa_{\text{max}}, \mathcal{C})$, where $\mathcal{W}_{\text{obs}}$ is the obstacle set, $x_{\text{start}}$ is the start configuration, $p_{\text{goal}} \in \mathcal{W}$ is the goal point, $\tau > 0$ is the goal tolerance, $\ell_{\text{max}}$ is the maximum insertion length, $\kappa_{\text{max}}$ is the maximum curvature, and $\mathcal{C}$ is a cost function. The problem calls for computing an optimal valid motion plan $\sigma^* = \arg\min_{\sigma \in \mathcal{C}(\sigma)} \mathcal{C}(\sigma)$ subject to:

- $\sigma$ is valid,
- $\sigma(0) = x_{\text{start}}$,
- $\ell_{\sigma} \leq \ell_{\text{max}}$,
- $\|\text{Proj}(\sigma(\ell_{\sigma})) - p_{\text{goal}}\|_2 \leq \tau$,

where $\text{Proj}(x) = p$ for $x = (p, q)$.

As we show in Sec. V, for any given instance of Problem 1, under some mild assumptions, there exists a fine-enough cutoff resolution $R_{\text{min}} = \{\delta\ell_{\text{min}}, \delta\theta_{\text{min}}\}$ (corresponding to the needle’s insertion and axial rotation, respectively) for which our planner is guaranteed to return a motion plan with a cost to be within a desired approximation factor of a globally optimal qualified motion plan in finite time (if any qualified motion plan exists), or indicate that no qualified motion plan exists. Similar to [30], we assume there exist minimal motions that are precisely achievable by the hardware system in tissues, informing the cutoff resolution. In our specific case, the minimal motions are the minimal insertion and minimal rotation the needle tip can precisely perform, determined by a lower-level needle controller.

**IV. THE RCS**\(^*\) **ALGORITHM**

We describe our RCS\(^*\) algorithm for resolution-optimal motion planning. We also highlight the differences between our new algorithm and our previous method RCS [30]. After presenting RCS\(^*\), we discuss an additional procedure to further improve its performance.

**A. Algorithm description**

The core idea of RCS\(^*\) (and RCS) is to build a search tree with predefined motions of multiple resolutions. Specifically, RCS\(^*\) constructs a search tree $T = (V, E)$ embedded in the configuration space, where each node $v \in V$ is associated with a configuration $x_v \in \mathcal{X}$ and each edge $e = (u, v) \in E$ represents a transition from $x_u$ to $x_v$. The search tree is rooted at the given start configuration $x_{\text{start}}$ and explores valid motion plans when expanded in the configuration space. See the pseudocode of RCS\(^*\) in Alg. 1.

To generate new nodes, RCS\(^*\) expands existing nodes with predefined (kinematically) feasible motion primitives [41]. In RCS\(^*\), a motion primitive defines a local circular trajectory with some constant curvature $\kappa \leq \kappa_{\text{max}}$, some length $\delta \ell > 0$, and some curving direction $\delta \theta \in [0, 2\pi]$. That is, a motion primitive is defined as a tuple $M = (\kappa, \delta \ell, \delta \theta)$. We denote the operation of applying a motion primitive $M$ to configuration $x_v$ as $x_{\u} = x_v \oplus M$, where $x_u$ is the resultant configuration. RCS\(^*\) uses a fixed set of curvatures $\{0, \kappa_{\text{max}}\}$ and defines the resolution of a motion primitive as a function of $\delta \ell$ and $\delta \theta$; since any curvature $\kappa \in [0, \kappa_{\text{max}}]$ can be well approximated by interleaving curvature 0 and $\kappa_{\text{max}}$ [30]. Generally speaking, the finer a resolution is, the finer the intervals $[0, \delta \ell_{\text{max}}]$ and $[0, \delta \theta_{\text{max}}]$ are discretized. We mention that the coarsest resolution is set with a user-defined $\delta \ell_{\text{max}}$ and $\delta \theta_{\text{max}} = \pi/2$ (4 initial steering orientations as shown in line 1).

In each iteration of RCS\(^*\), an expansion of existing nodes is performed in an $A^*$ fashion. In particular, nodes are iteratively extracted from the OPEN list (line 5), wherein nodes are ordered according to their rank and a secondary metric $f(\cdot)$. We define the rank of a node as a function of the node’s depth in the tree and the resolution of the motion primitives leading to the node. The deeper a node is in the tree and the finer resolution the motion primitives leading to the node, the higher rank a node has (see formal definition of rank in [30]). The secondary metric $f(v) = C(v) + h(v)$ has $C(v)$ denoting the cost of the trajectory from the root of $T$ to $v$ with respect to $C$ and $h(v)$ being a heuristic function estimating the cost of the trajectory from the node $v$ to the goal point. For example, in the case where $C$ is trajectory length, we have $h(v)$ be the length of the Dubins curve [42] on the plane spanned by $x_v$ and $p_{\text{goal}}$. Unlike in RCS, where nodes with lower rank are always extracted first, RCS\(^*\) relaxes this ordering by introducing a look-ahead parameter denoted as $n_{\text{la}} \in \mathbb{N}$ (similar to the idea in [35]). At any time during the search, we denote the minimum rank of nodes in the OPEN list as $r_{\text{open}}$. Then we order all nodes with rank $r \leq r_{\text{open}} + n_{\text{la}}$ according to a secondary metric $f(\cdot)$. This is done to prioritize searching nodes from a coarser resolution, which speeds up finding an initial motion plan.

Given an extracted node $v$, we first check if it is valid (line 6) using the conditions described for RCS, which

<table>
<thead>
<tr>
<th>Algorithm 1 RCS(^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> $W_{\text{obs}}, x_{\text{start}}, p_{\text{goal}}, \tau, \kappa_{\text{max}}, \ell_{\text{max}}, \delta \ell_{\text{max}}, R_{\text{min}}$</td>
</tr>
<tr>
<td><strong>Output:</strong> bestPlan</td>
</tr>
<tr>
<td><strong>1:</strong> $\Theta \leftarrow {0, \frac{\pi}{2}, \frac{3\pi}{2}}$, $K \leftarrow {0, \kappa_{\text{max}}}$</td>
</tr>
<tr>
<td><strong>2:</strong> root $\leftarrow (x_{\text{start}}, 0, 0)$ // The root has rank 0 and cost 0</td>
</tr>
<tr>
<td><strong>3:</strong> OPEN $\leftarrow {\text{root}}$, CLOSED $\leftarrow \emptyset$, bestPlan $\leftarrow \text{NULL}$</td>
</tr>
<tr>
<td><strong>4:</strong> while not OPEN.empty() do</td>
</tr>
<tr>
<td><strong>5:</strong> $v \leftarrow$ OPEN.extract()</td>
</tr>
<tr>
<td><strong>6:</strong> if Valid($v, W_{\text{obs}}, p_{\text{goal}}, \ell_{\text{max}}$) then</td>
</tr>
<tr>
<td><strong>7:</strong> if not CLOSED.existDuplicate($v$) then</td>
</tr>
<tr>
<td><strong>8:</strong> if GoalReached($v, p_{\text{goal}}, \tau$) then</td>
</tr>
<tr>
<td><strong>9:</strong> bestPlan.update($v$)</td>
</tr>
<tr>
<td><strong>10:</strong> for $M \in \text{Primitives}(K, \delta \ell_{\text{max}}, \Theta)$ do</td>
</tr>
<tr>
<td><strong>11:</strong> OPEN.insert($v \oplus M$)</td>
</tr>
<tr>
<td><strong>12:</strong> CLOSED.insert($v$)</td>
</tr>
<tr>
<td><strong>13:</strong> if $v! = \text{root}$ then</td>
</tr>
<tr>
<td><strong>14:</strong> for $M \in \text{RefinedPrimitives}(M_v)$ do</td>
</tr>
<tr>
<td><strong>15:</strong> if ValidResolution($M, R_{\text{min}}$) then</td>
</tr>
<tr>
<td><strong>16:</strong> OPEN.insert($v. parent \oplus M$)</td>
</tr>
<tr>
<td><strong>17:</strong> return bestPlan</td>
</tr>
</tbody>
</table>
ensure that (i) the insertion length would not exceed $\ell_{\text{max}}$, (ii) the goal region is still reachable after getting to $v$, (iii) the trajectory from the root to $v$ is not identical to another node that only needs coarser motion primitives to get to, and (iv) that the edge leading to $v$ is collision free [30]. In addition, RCS* checks that the cost $C(v)$ is smaller than the cost of the best plan reaching the goal region found so far. If the heuristic function $h(v)$ is admissible, we use $f(v)$ instead of $C(v)$ in the last condition, as $f$ provides a better estimate of the node cost and hence allows to prune more vertices.

To further boost efficiency, the algorithm avoids expanding nodes that are overly similar to existing nodes in terms of the induced configuration and cost by performing duplicate detection (line 7). A node $v$ is determined as a duplicate if there exists a node $u$ in the CLOSED list that satisfies (i') $\rho(x_u, x_v) < d_{\text{sim}}$ and (ii') $C(u) \leq C(v)$, where $\rho$ is a distance function defined on $\mathcal{X}$ and $d_{\text{sim}}$ is a similarity parameter. We use $\rho(x_u, x_v) = \|p_u - p_v\|_2 + \alpha \cdot \text{dist}_c(q_u, q_v)$, where $\alpha > 0$ is a weighting parameter and $\text{dist}_c()$ is the angular distance between two orientations. Sec. V specifies the value of $d_{\text{sim}}$. Condition (i') is shared between RCS and RCS*, while condition (ii') is important for keeping RCS* resolution optimal as it prevents nodes with lower cost from being pruned away by nodes with higher cost.

RCS* uses a set of motion primitives of different resolutions, but instead of applying all motion primitives together, only the coarsest motion primitives are used when a node is initially expanded (line 10). The resolution of a motion primitive (with respect to $\delta \ell$ and $\delta \theta$) is refined when a node with a coarser motion primitive is processed (line 14). More specifically, a finer motion primitive is obtained by changing $\delta \ell$ or $\delta \theta$ by a small value that corresponds to a finer resolution (see [30] for mathematical expressions). Note that resolution refinement is done even if a node is invalid since finer resolutions might be valid.

Resolution refinement will be cut off when reaching a finer resolution, with respect to $\delta \ell$ or $\delta \theta$, than the predefined cutoff resolution $R_{\text{min}}$ (line 15). Here the cutoff resolution is determined by the minimal motions of the needle tip that are precisely achievable with the hardware system.

The algorithm terminates when the OPEN list is exhausted (line 4), and the best plan is returned (if any is found). RCS* is guaranteed to terminate in finite time due to the cutoff resolution.

### B. Domain-specific optimization

We describe an additional procedure to further improve RCS*’s performance. We incorporate the concept of inevitable collisions [42] to eliminate potential nodes that would lead to collisions as they are expanded. In particular, for a given vertex $v$ and the goal point, region growing is performed from $x_v$ within an approximated reachable workspace, considering the existence of obstacles. This region is defined as the intersection of the kinematically forward-reachable workspace and the olive-shaped feasible workspace defined by $x_{v_0}, p_{\text{goal}}$, and tolerance $\tau$ (see Fig. 2).

We mention that due to (i) maximum curvature constraint, (ii) maximum turning angle constraint (the needle would shear or buckle when turning over $\pi/2$), and (iii) maximum insertion length constraint, the kinematically forward-reachable workspace for a given needle configuration is a trumpet-shaped volume (see Fig. 2 left). In the case that the goal is not reached by the grown region, $v$ is discarded. For additional optimizations applicable to RCS*, see [43].

### V. Resolution Optimality of RCS*

We study the theoretical properties of RCS* and provide a proof sketch for the algorithm being resolution optimal. Informally, resolution optimality implies that RCS* is guaranteed to find a plan whose cost is as close as desired to the cost of the globally optimal qualified motion plan $\sigma^*$, assuming that the cutoff resolution $R_{\text{min}} = \{\delta \ell_{\text{min}}, \delta \theta_{\text{min}}\}$ is fine enough. Thm. 1 given below, states our main theoretical contribution relating to the resolution optimality of RCS*.

Before reaching Thm. 1, we introduce the notions of well-behaved cost, as well as robust and decomposable trajectories, which will be used as assumptions on $C$ and $\sigma^*$ to prove our result. It may not be possible to approximate $C$ or $\sigma^*$ using motion plans with a finite number of transitions without additional constraints on the cost function $C$ or if the plan $\sigma^*$ has no clearance from obstacles.

The following definition states that close-by configurations have similar costs and that there are bounds on the values that the cost can attain.

**Definition 1** (Well-behaved cost). A configuration-based cost function $c$ is well-behaved if (i) it is Lipschitz continuous, i.e., $\forall x_1, x_2 \in X_{\text{free}}, |c(x_1) - c(x_2)| \leq L_c \cdot \rho(x_1, x_2)$ for $L_c \in \mathbb{R}$, and (ii) $\forall x \in X_{\text{free}}, c(x) \in [c_{\text{min}}, c_{\text{max}}]$, where $c_{\text{min}}, c_{\text{max}} \in \mathbb{R}$ and $c_{\text{min}} > 0$. In such a case, we also say that the trajectory-based cost function $C(\sigma) = \int_0^1 c(\sigma(s)) \, ds$ is well-behaved.

We assume $c_{\text{max}}$ is not infinitely large since such configurations can be removed from $X_{\text{free}}$. We also require a well-behaved cost to satisfy $c_{\text{min}} > 0$ since in the case of needle steering, there is always a cost associated with puncturing tissue, so it makes no sense to allow cost-free regions.

Next, we provide two definitions that are used to characterize motion plans that RCS* can approximate. The first definition (borrowed from [30]) is concerned with trajectories that are induced by a finite set of motion primitives (not necessarily the ones used by RCS*). The second definition is concerned with trajectories that are robust with respect to clearance. A motion plan is considered qualified if it satisfies both definitions.

**Definition 2** (Decomposable trajectory). A trajectory $\sigma : [0, l] \rightarrow X$ is decomposable if it can be decomposed into a finite set of motion primitives. Namely, there exist primitives $M_\sigma = \{M_1, \ldots, M_n\} \subset A$ such that $\sigma = \sigma(0) \oplus M_\sigma$,.
where \( x \oplus M \) denotes the resultant trajectory obtained by sequentially applying elements in \( M \) to \( x \).

**Definition 3** (Robust trajectory). A trajectory \( \sigma : [0, l] \rightarrow \mathcal{X} \) is \( \delta \)-robust, for some \( \delta > 0 \), if it has \( \delta \) clearance from (i) obstacles \( \min_{x \in [0, l], x \in \mathcal{X}_{\text{obs}}} \rho(\sigma(x), x) \geq \delta \), and (ii) the goal region boundary, \( \| \text{Proj}(\sigma(l)) - p_{\text{goal}} \| \leq \tau - \delta \).

We are ready to state our main theoretical result concerning the resolution optimality of RCS*.

**Theorem 1** (Resolution optimality). Let \( \Delta = (\mathcal{X}, \mathcal{V}_{\text{clust}}, x_{\text{start}}, p_{\text{goal}}, \ell_{\max}, \kappa_{\max}, C) \) be an optimal steerable needle motion-planning problem, \( \varepsilon \in (0, \infty) \) is an approximation factor, and \( \sigma^* \) be a trajectory. Also, suppose that the following conditions are satisfied:

1. **C1** The steerable-needle system is Lipschitz continuous.
2. **C2** The cost function \( C \) is well-behaved and characterized with \( L_c, \epsilon_{\min}, \epsilon_{\max} \). Denote \( k = \frac{L_c \epsilon_{\max}}{\epsilon_{\min}} \).
3. **C3** \( \sigma^* \) is decomposable and \( \delta \)-robust with \( \delta = \min\left\{ \frac{\varepsilon}{2}, \frac{\tau}{2} \right\} \).
4. **C4** The radius \( d_{\text{sim}} \) used to reject similar nodes satisfies

\[
d_{\text{sim}} \leq \min\left\{ \frac{2 \kappa_{\max} \epsilon_{\min} \delta(L_{\max} - 1)}{2(L_{\max} - 1)} \right\}
\]

Then RCS* is resolution optimal, i.e., for fine-enough cutoff resolution \( R_{\min} = \{ \delta_{\ell_{\min}}, \delta_{\theta_{\min}} \} \), RCS* will find a motion plan that satisfies \( C(\sigma) \leq (1 + \varepsilon) \cdot C(\sigma^*) \).

**A. Proof sketch for Thm. 1**

We provide a sketch of the proof for Thm. 1, which consists of two main steps. We first show that the plan \( \sigma^* \) can be approximated by another plan \( \sigma_{\text{obs}}^* \) that is composed solely of the motion primitives used by RCS*. Then, we show that even though RCS* might not be able to exactly find \( \sigma_{\text{obs}}^* \) due to pruning, it will be able to recover another plan \( \tilde{\sigma}_{\text{obs}}^* \) whose cost is similar to that of \( \sigma^* \) and \( \sigma_{\text{obs}}^* \). The full proof is in the extended version of our paper [43].

We assume that the conditions in Thm. 1 are met for a plan \( \sigma^* \) and approximation factor \( \varepsilon > 0 \). As a first step, we show that there exists a resolution \( R_{\min} \) and a trajectory \( \sigma_{\text{obs}}^* \) that approximates \( \sigma^* \) and is constructed solely from the motion primitives of RCS*. In particular, \( \sigma_{\text{obs}}^* \) is a *piecewise \( \beta \)-approximation* of \( \sigma^* \) which is defined as follows: the two trajectories \( \sigma_{\text{approx}}^* \) and \( \sigma_{\text{approx}}^* \) can be partitioned into a sequence of trajectories \( \sigma_1, \ldots, \sigma_n \) and \( \sigma'_1, \ldots, \sigma'_n \), respectively, for some positive and finite integer \( n \), such that for any \( 1 \leq i \leq n \) (i) the Hausdorff distance between \( \sigma_i \) and \( \sigma'_i \) at any some value \( \beta \) > 0, and (ii) the cost of the two trajectories satisfies \( C(\sigma_i^*) \leq (1 + \beta)C(\sigma_i) \) (see formal definition in [42]). This step follows from refining our previous proof that showed the existence of such an approximation, albeit only for the length cost [30, Lemma 2], and relies on the assumptions C1 and C2, which characterize the system and cost function, and assumption C3, which states that \( \sigma^* \) is decomposable.

The value \( \beta \) is chosen to guarantee that \( C(\sigma_{\text{approx}}^*) \leq (1 + \varepsilon)C(\sigma^*) \). Moreover, the choice of \( \beta \), and the fact that \( \sigma_{\text{approx}}^* \) is a \( \beta \)-approximation of \( \sigma^* \) also ensure that \( \sigma_{\text{approx}}^* \) is collision free and satisfies the goal tolerance according to assumption C3. Note that the above properties would still hold even if we replace the constant \( \beta \) with a slightly larger value \( \beta' > \beta \). However, we use the more conservative value \( \beta > 0 \) to compensate for the fact that RCS* prunes the tree. In particular, due to pruning, RCS* might eliminate some of the vertices induced by the trajectory \( \sigma_{\text{approx}}^* \), and thus, we cannot guarantee that \( \sigma_{\text{approx}}^* \) would be returned as a solution. However, we can show that even in the presence of pruning RCS* will compute a valid plan \( \sigma_{\text{approx}}^* \) that tightly bounds the cost of \( \sigma_{\text{approx}}^* \) (and thus tightly bounds the cost of \( \sigma^* \)).

We now elaborate on this. Denote the sequence of motion primitives that define \( \sigma_{\text{approx}}^* \) as \( M_{\sigma^*} = \{ M_1, \ldots, M_n \} \). When the motions in \( M_{\sigma^*} \) are sequentially applied to \( x_{\text{start}} \), we obtain a sequence of configurations \( \{ x_0, x_1, \ldots, x_n \} \), where \( x_0 = x_{\text{start}}, x_i = x_{i-1} + M_i, i \in [1, n] \), some of which may be pruned. By carefully bounding \( d_{\text{sim}} \) (according to C4), we guarantee that in the worst case the trajectory obtained by applying the same sequence of motion primitives to pruned nodes stays within a collision-free tunnel around \( \sigma_{\text{approx}}^* \), ensuring that it remains valid. The choice of \( d_{\text{sim}} \) also takes care of goal tolerance, guaranteeing that the trajectory still satisfies \( \gamma \) goal tolerance. Finally, because (i) pruning is allowed only when there exists a node with equal or smaller cost, and (ii) the subtrajectories obtained from applying the same motion primitive to close-by configurations are strict approximations, the plan \( \tilde{\sigma}_{\text{obs}}^* \) tightly bounds the cost of \( \sigma^* \).

**VI. RESULTS**

We focus on the medical procedure of lung biopsy for evaluation. Lung and bronchus cancer has the highest death rate among all types of cancer, killing over 130,000 Americans each year [44]. Lung biopsy enables definitive diagnosis of suspicious lung nodules at an early stage, which is important to increase the survival rate. One potential approach to safely and accurately access lung nodules for biopsy and localized treatment is deploying a steerable needle trans-orally through a bronchoscope to avoid transthoracic access which could cause severe side effects [45–48]. In this approach, a physician deploys a bronchoscope through a patient’s bronchial tubes and then the steerable needle is deployed through the bronchoscope, exits out of the bronchial tube, and steers in the lung parenchyma to reach the nodule. Our motion planner focuses on the final stage of automatically steering the needle through the lung parenchyma to the nodule, while avoiding anatomical obstacles including large blood vessels, bronchial tubes, and the lung boundary.

We used the method developed in [12] to reconstruct the anatomical workspace from a chest CT scan with the above-mentioned obstacles. We created 100 test cases. For each test case, we randomly sampled a start pose close to the bronchial tubes and a target in the lung parenchyma. We finished creating test cases when the number of test cases reached 100 after rejecting (i) impossible scenarios where the start pose has inevitable collision and (ii) trivial scenarios where the start pose can be connected directly to the goal point with a collision-free arc. The simulated needle has \( \kappa_{\max} = (50 \text{mm})^{-1} \), \( \ell_{\max} = 100 \text{mm} \), and a diameter of 2mm. We set the goal tolerance \( \tau = 1 \text{mm} \).

We compared RCS* in simulation with several planners:

- **RRT**: The RRT-based needle planner [10, 15] with 5% goal biasing and 100% goal connecting ratio.
- **AO-RRT**: AO-RRT [22, 23] adapted for steerable needles, with maximum rotation control \( 2 \pi \) and a maximum insertion control 20mm. We follow the guidelines in
for three iterations, as suggested in [18]. For each method, except performance over time, and only reported the final cost after used a CPU implementation of AFT. Additionally, for all cost functions, we set \( \epsilon \) except for \( \text{AFT} \) used the same basic setup as \( \text{RCS}^* \) with Motion Planning Templates (MPT) [49].

(iii) \text{AFT}: The AFT-based needle planner [18, 19], with setup following [19]. AFT internally uses a hybrid cost function; we use \( C_{\text{hybrid}}(\sigma) = \omega \cdot C(\sigma) + \|\sigma(\ell_\sigma) - p_{\text{goal}}\|_2/\tau \), where \( \omega \) is a weighting parameter depending on the scale of \( C \). Note that \( C_{\text{hybrid}} \) is only used internally in AFT while \( C \) is always used for performance comparison across different planners.

(iv) \text{RCS}: The RCS needle planner [30] with cutoff resolution \( R_{\text{min}} = \{0.125\text{mm}, 0.157\text{rad}\} \), \( \delta \ell_{\text{max}} = 20\text{mm} \). \( \text{RCS}^* \) used the same basic setup as RCS and \( R_{\text{hit}} = 3 \). Additionally, for all cost functions, we set \( \varepsilon = 0.1 \). Since we used a CPU implementation of AFT, we did not compare its performance over time, and only reported the final cost after three iterations, as suggested in [18]. For each method, except for AFT, the timeout was set to 100 seconds. All methods except for AFT achieved 100% success rate when timed out while AFT achieved 87% success rate. When reporting the relative cost of AFT, we only consider the test cases that are successfully solved by AFT. As \( \text{RRT} \) and RCS are designed to find a plan instead of optimizing a plan, we modified them to keep running after the first plan is found, and always return the best plan found when timed out. All experiments were run on a dual 2.1GHz 16-core Intel Xeon Silver 4216 CPU and 100GB of RAM. All parallelizations were implemented with Motion Planning Templates (MPT) [49].

We used two well-behaved cost functions for which \( \text{RCS}^* \) is resolution optimal: (i) trajectory length, and (ii) a cost function informed by a cost map derived from medical images [12], where each voxel in the 3D cost map is associated with a cost value that represents tissue damage. We used trilinear interpolation to smooth out the voxelized cost map and forced \( c_{\text{min}} = 0.01 \).

Fig. 3 shows results for the different planners for the cost map and trajectory length cost functions. For all cost functions, the cost of a plan may vary significantly between test cases. For example, trajectory length is affected by how far away the target lies relative to the start pose and cost map values are much higher when the needle is steering in a vessel-cluttered region. To account for the large variation between test cases, we first computed relative cost within one test case. Specifically, we took the first-found plan (no matter which method returns it as long as it is returned the fastest), and computed relative cost to this plan for all other plans found for the same test case. We did this for all the test cases, and then averaged over the 100 test cases (see the first two plots in Fig. 3). Such relative cost decreased as the result plans are gradually optimized. AFT, which is omitted from those plots, achieved a final cost of 0.574 and 0.991 for cost map and trajectory length, respectively.

For both cost functions, \( \text{RCS}^* \) achieved the best final costs. For the cost map, which is more clinically relevant, \( \text{RCS}^* \) outperformed other methods after 1.5 seconds with a final cost of 0.277, which is 50% lower than AFT (0.574), 35% lower than \( \text{RRT} \) (0.423) and \( \text{AO-RRT} \) (0.427), and 7% lower than RCS (0.299). This indicates that the final plan produced by \( \text{RCS}^* \) successfully avoids more small blood vessels than the other methods. It is worth mentioning that unlike trajectory length, small perturbations may lead to very different costs when we use the cost map. As for trajectory length, all methods generated trajectories with roughly similar lengths, although \( \text{RCS}^* \) computed shorter trajectories than all other methods.

In Fig. 3, we report the cumulative distribution function across the 100 test cases of the relative final cost of RCS with respect to \( \text{RCS}^* \) for the cost-map setting. Although RCS achieved comparable costs with \( \text{RCS}^* \) in most cases, for about 20% of the test cases RCS achieved a motion plan that is at least 10% more costly than \( \text{RCS}^* \). In the extreme case, the final cost of RCS was 3.67 times that of \( \text{RCS}^* \).

Finally, we mention that the number of nodes \( \text{RCS}^* \) explored is only 53% (for cost map) and 30% (for trajectory length) of that RCS explored. This indicates that although \( \text{RCS}^* \) spent more time choosing the right node to explore, it explores much fewer nodes than RCS to get a plan with even higher quality. The experimental results demonstrate that \( \text{RCS}^* \) is faster than competing methods and that the theoretical guarantees of \( \text{RCS}^* \) have a practical impact on the quality of result plans, which can reduce patient trauma.

VII. Conclusion

In this paper, we introduce the first resolution-optimal motion planner for steerable needles. In particular, our method returns in finite time a motion plan whose cost can be as close as desired to the globally optimal qualified motion plan, assuming the given resolution is fine enough. We also provide a proof sketch to show the resolution optimality of our method with a careful discussion of assumptions and required conditions. We evaluate our proposed planner with simulation experiments and show it can efficiently compute high-quality motion plans considering clinically relevant cost functions. In the future, we plan to investigate speedup techniques for vertex validation and pruning to reduce vertex expansion time. In addition, we plan to develop explicit expressions for the resolution \( R_{\text{min}} \) necessary to achieve the desired level of approximation quality for a given problem instance. We will also experimentally evaluate the planner with steerable needles in ex-vivo tissues.
REFERENCES


