

A PROBABILISTIC METHOD FOR RIGID BODY MOTION PLANNING
USING SAMPLING FROM THE MEDIAL AXIS OF THE FREE SPACE

A Dissertation

by

STEVEN ALBERT WILMARTH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 1999

Major Subject: Mathematics

A PROBABILISTIC METHOD FOR RIGID BODY MOTION PLANNING
USING SAMPLING FROM THE MEDIAL AXIS OF THE FREE SPACE

A Dissertation

by

STEVEN ALBERT WILMARTH

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Peter F. Stiller
(Co-Chair of Committee)

Nancy M. Amato
(Co-Chair of Committee)

Emil J. Straube
(Member)

Bruce H. McCormick
(Member)

William Rundell
(Head of Department)

December 1999

Major Subject: Mathematics

ABSTRACT

A Probabilistic Method for Rigid Body Motion Planning

Using Sampling from the Medial Axis of the Free Space. (December 1999)

Steven Albert Wilmarth, B.S., Tufts University;

M.S., Rutgers University

Co-Chairs of Advisory Committee: Dr. Peter F. Stiller
Dr. Nancy M. Amato

Motion planning in the presence of obstacles is an important problem in robotics with numerous applications in other areas. While complete motion planning algorithms do exist, they are rarely used in practice since they are computationally infeasible in all but the simplest cases. For this reason, many recent efforts have focused on probabilistic methods, which sacrifice completeness in favor of computational feasibility and applicability. In particular, several algorithms, known as *probabilistic roadmap planners*, have been shown to perform well in a number of practical situations; however, their performance degrades when paths are required to pass through narrow passages in the free space. In this dissertation we present a method of sampling the configuration space of a rigid body moving in three dimensions in which randomly generated configurations are retracted onto the medial axis of the free space. We develop some theory of the medial axis on the configuration space $SE(3)$ and present algorithms that perform the retraction while avoiding explicit computation of the medial axis. Finally, we give some preliminary experimental results demonstrating the performance of the algorithm.

ACKNOWLEDGMENTS

First, I would like my advisors, Peter Stiller and Nancy Amato, for their help and guidance – scientific, technical, practical, political, motivational, and otherwise. Getting this degree was a great experience for me and this is due mainly to the many talents and skills of these two individuals.

I would also like to thank the faculty and student members of the robotics group at Texas A&M: Prof. Jeff Trinkle, Burchan Bayazit, Lucia Dale, Chris Jones, Wookho Son, Guang Song, Daniel Vallejo, and particularly Li Han for many helpful ideas and discussions. Also, thanks to Juan Manuel Ahuactzin for his helpful advice.

I would like to thank the GAT group in the Department of Mathematics and the National Science Foundation for financial support. The Institute for Scientific Computation generously provided computing resources.

I would like to thank my family, particularly my parents, for their unwavering support of my education. It really does help if you never hear the words: “What, more school...?”

Finally, I would like to thank my wife Jenny for her tolerance, moral support, technical advice, and whatever else I happened to need at the time. You know, I thought I might look into this program at UMD... Joke. Honest.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
	A. Motion planning	2
	B. Probabilistic roadmap methods	3
	C. The medial axis	5
	D. Our research	7
II	PRELIMINARIES	10
	A. Metric spaces	10
	B. Retraction	12
	C. Curves in the plane	16
	1. The Frenet frame	18
	2. Curvature	19
	D. Regions in the plane	24
	E. Riemannian manifolds	26
	F. Motion planning for a rigid body	29
	G. $SE(3)$	31
	H. Probabilistic roadmap methods	32
III	MAPRM IN THE PLANE	36
	A. Definition	36
	B. The canonical retraction map	42
	C. The extended retraction map	50
	D. Regions with C^2 boundary	53
	E. Regions with piecewise C^2 boundary	59
	F. MAPRM in the plane	62
	1. Rudimentary analysis	64
	2. Examples	66
IV	THE MEDIAL AXIS ON $SE(3)$	69
	A. The definition of the medial axis	70
	B. The choice of Riemannian metric	71
	C. The medial axis in \mathbb{R}^3	77
	D. The retraction maps	79

CHAPTER	Page
E. Continuity of the canonical retraction map	83
F. The medial axis of a free space of polyhedra	85
V THE MAPRM ALGORITHM	89
A. Finding the nearest contact configuration	89
B. Implementation and experimental results	95
1. Narrow corridor example	96
2. Wider corridor example	98
VI CONCLUSIONS AND FURTHER RESEARCH	100
REFERENCES	102
VITA	108

LIST OF TABLES

TABLE		Page
I	Experimental results for cubical workpiece.	97
II	Experimental results for a smaller workpiece.	98

LIST OF FIGURES

FIGURE		Page
1	Retraction examples.	14
2	SDR for motion planning.	15
3	Intersections of normal lines approximate the center of the oscu- lating circle.	22
4	The maximal ball at a point.	37
5	Some regions and their medial axes.	39
6	The image of a point x under the canonical retraction map.	43
7	The medial axis of F may have points in common with $\text{MA}(\overline{\mathbb{R}^2 \setminus F})$	47
8	F' for regions (a) and (b) from Figure 5.	48
9	Computing the canonical retraction map.	50
10	The canonical and extended retraction maps.	51
11	The area added by the extended retraction map.	65
12	Uniform sampling (a) vs. MAPRM sampling (b).	66
13	Another example: uniform sampling (a) vs. MAPRM sampling (b).	67
14	Motivation in 2-D.	93
15	Rigid body example.	96

CHAPTER I

INTRODUCTION

Motion planning in the presence of obstacles is an important problem in robotics with numerous applications in other areas. While complete motion planning algorithms do exist, they are rarely used in practice since they are computationally infeasible in all but the simplest cases. For this reason, recent attention has focused on probabilistic methods, which sacrifice completeness in favor of computational feasibility and applicability. In particular, several algorithms, known collectively as *probabilistic roadmap planners* or PRMs, have been shown to perform well in a number of practical situations, see, e.g., [25]. The idea behind these methods is to create a graph of randomly generated collision-free configurations with connections between these nodes made by a simple and fast local planning method. Actual global planning is then carried out on this graph. These methods run quickly and are easy to implement; unfortunately there are simple situations in which they perform poorly, in particular situations in which paths are required to pass through narrow passages in configuration space.

The *medial axis* or *generalized Voronoi diagram* has a long history of use in motion planning, see [7, 26]. This stems from the fact the medial axis $\text{MA}(F)$ of the free space F , the set of all collision-free configurations of the moving workpiece, has lower dimension than F but is still a complete representation for motion planning purposes. In particular, in two dimensions the medial axis is a one-dimensional graph-like structure which can be used as a roadmap. Paths on the medial axis also have other appealing properties such as large clearance from obstacles. However, the medial axis is difficult and expensive to compute explicitly, particularly in higher dimensions.

This dissertation follows the style and format of SIAM Review.

In this work, we combine these approaches to develop a method for rigid body motion planning which retracts sampled nodes onto the medial axis of the free space in a computationally efficient manner. The central idea is that *most sampled configurations, free or not, can be retracted onto the medial axis of the free space without having to compute the medial axis explicitly*. Sampling and retracting in this way will be shown to give improved performance on problems requiring traversal of narrow passages.

A. Motion planning

Roughly speaking, motion planning consists of finding a continuous sequence of movements to take an object from a given initial state to a given goal state. There are often restrictions placed on the allowed motions: avoiding obstacles, known or unknown, maintaining practical velocities and accelerations, etc. Also, the “movers” themselves may be inherently complex, such as complicated geometric shapes, robot arms or hands with multiple links, or mobile robots with restricted steering mechanisms. There are also many problems outside of traditional motion planning that can be made to fit into this framework, such as assembly planning [44] and problems from structural chemistry and drug design [17].

A simplified version of the problem of planning motions for a single rigid body results if we assume no constraints on the motion and just consider a purely geometric version of the problem: what sequence of motions will move the body from here to there without hitting anything? As it happens, such simplified geometric versions of motion planning have turned out to be extremely difficult to solve in practice. Such problems are still at the center of motion planning research and form the topic of this dissertation. See the text [26] for a survey of motion planning.

Motion planning problems are generally worded in terms of the *configuration space* of the moving object, the set of all possible “states” of the workpiece. For a rigid body in three dimensions, this is the set of all possible positions and orientations of the body. The *free space* is the set of all configurations that do not place the moving object in collision with an obstacle. In this terminology, the problem is: given start and goal configurations in the free space, find a path connecting them that stays entirely in the free space.

B. Probabilistic roadmap methods

Much of the existing motion planning literature focuses on theoretical considerations such as the computational complexity of various planning problems. Due to many factors, among them increased availability and use of robotic manipulators and CAD/CAM systems, the introduction and increased practicality of virtual reality systems, and the decreased expense of computers capable of rendering and manipulating three dimensional geometry, there has recently been a significant increase in the development of practical motion planning algorithms. Probabilistic roadmap methods are among those recent developments.

Probabilistic roadmap methods generally operate as follows, see, e.g., [21, 25, 38]. During a preprocessing phase, a set of configurations in the free space is generated by sampling configurations at random and removing those that put the workpiece in collision with an obstacle, i.e., those that do not land in the free space. These free configurations are then connected into a graph or *roadmap* by inserting an edge in the graph between a pair of configurations if they can be connected in the free space by a simple and fast local planning method. This roadmap can then be queried by connecting the given start and goal configurations to nodes in the roadmap, again

using the local planner, and then searching for a path in the roadmap connecting these nodes. The algorithms are easy to implement, run quickly, and are applicable to a wide variety of robots.

In [25] this method is applied to an articulated robotic arm in the plane and an application to multiple robots appears in [39]. Applications of this idea to the problem of moving an elastic sheet appear in [23], and to the problem of closed kinematic chains in [27]. Various sampling schemes, metrics, and local planners are compared in [1] and [1]. Results using a parallel implementation are presented in [4]. The paper [40] employs a similar method adapted to situations involving only a single query pair. A technique for allowing the user to manually input additional nodes using a haptic input device is explained in [3] and [8]. Applications to industrial assembly problems drawn from the automotive industry are explored in [20].

Analysis of the performance of these methods has proven to be difficult. Some performance results in terms of geometric characteristics of the free space appear in [24, 20]. These results are aimed at giving an indication of what make a problem difficult from the perspective of probabilistic roadmap methods. The results in [22] provide some guidance about how many nodes need to be sampled to guarantee a given probability of success for a particular problem in terms of an assumed path clearance from the boundary.

The main shortcoming of these methods is their poor performance on problems requiring paths to pass through narrow passages in the free space. This is a direct consequence of how the nodes are sampled from F . For example, using uniform sampling over F , any corridor of sufficiently small volume is unlikely to contain any sampled nodes whatsoever: this can potentially cause the roadmap to have more connected components than the free space.

In [21], uniform random sampling is used to generate free nodes along with a

post-connection roadmap enhancement heuristic designed improve connectivity of the roadmap. Intuitively, narrow corridors in configuration space may be characterized by their large surface area to volume ratio. The methods in [19] have exploited this idea. In [1, 5], nodes are sampled from the *contact space*, the set of configurations for which the workpiece is in contact but not collision with an obstacle. In [19], preliminary configurations are generated by allowing the workpiece to penetrate the obstacles by a small amount. The areas near these nodes are then re-sampled to find nearby collision-free configurations. The idea is that the allowed penetration dilates the free space by a small amount (albeit not uniformly), and the sampling in a narrow corridor is increased roughly in proportion to the surface area. As the authors point out, dilating the free space may alter its topology, opening corridors where none existed. In practice, the amount of dilation must be carefully regulated to mitigate this effect. A Gaussian sampling strategy which clusters nodes around the obstacles is presented in [10].

Our work represents another approach to dealing with this narrow passage problem.

C. The medial axis

The medial axis has been used extensively in motion planning, see [26]. A typical approach is to compute the medial axis of the free space, which generally has lower dimension, and to carry out the planning there instead. This is valid because $\text{MA}(F)$ is generally a *strong deformation retract* (SDR) of F , meaning that F can be continuously deformed onto $\text{MA}(F)$ while maintaining its topological structure. In particular, in two dimensions the medial axis is a one-dimensional graph-like structure which can be used as a roadmap of the free space. Paths on the medial axis also

have other appealing properties such as large clearance from obstacles. However, the medial axis is difficult and expensive to compute explicitly, particularly in higher dimensions, so its practical use has mostly been confined to two-dimensional problems. The medial axis and its relatives have also been proven to be of theoretical value in improving complexity bounds, see [13].

Motion planning is only one of many fields in which the medial axis has found practical application. In computational solid geometry, the medial axis is used as a representation of solid shapes, see for example [18, 41]. In image processing, the medial axis has found extensive use as a descriptor of shape and as a thinning method, see [36, 35]. One of original papers on the medial axis, [9], used it as a descriptor of biological shape. Most of this work has concentrated on the practical computation of the medial axis for various classes of objects and not on the theoretical foundations.

Mathematically, the medial axis has connections with the cut locus concept from Riemannian geometry, see [45, 37].

Although the dominant use of the medial axis has been in two-dimensional Euclidean space, even in this case it is still not mathematically well understood. The book [36] provides significant results and shows where current results fall short. Work of a fundamental nature has appeared recently in the mathematics literature, [14] for example.

Our interest in the medial axis centers on the details of retracting a free space onto its medial axis, on the circumstances under which the medial axis is a strong deformation retract under a particular naturally-arising retraction map, and on computing the images of this map in practice. We should note that the existing theoretical work in this area generally addresses the circumstances under which the medial axis is a strong deformation retract under *any* retraction map.

D. Our research

In this dissertation, we present a modified PRM sampling algorithm for rigid body motion in 3-D, which retracts sampled nodes onto the medial axis of the free space prior to their connection to form a roadmap. We refer to the algorithm as “MAPRM”: Medial Axis Probabilistic Roadmap Method. We first motivate MAPRM by presenting it for a point moving among obstacles in the plane, and then generalize to the case of a rigid body moving in 3-D.

As mentioned above, a typical approach using the medial axis in motion planning is to compute the medial axis of the free space and to carry out the planning there instead. Again, this is valid because in general F can be continuously deformed onto $MA(F)$ while maintaining its topological structure. In fact, as we will show, almost the entire configuration space, free and collision configurations alike, can be retracted onto $MA(F)$. Although a complete representation of the medial axis of the free space is difficult and costly to compute, the final retracted image on $MA(F)$ of a given free configuration can be computed efficiently without such a representation. We exploit this fact by sampling nodes from the full configuration space and retracting them onto $MA(F)$. These nodes, now all in the free space, can be connected in the usual way to form a roadmap. We will show that this has the effect of increasing sampling in narrow corridors in a way that is independent of the volume of the corridor.

Our main results are:

1. In the case of a polyhedron moving among polyhedral obstacles, under certain assumptions on the configuration space metric being used, the medial axis of the free space is a strong deformation retract of nearly the entire configuration space under an easily defined retraction mapping.
2. It is frequently possible to efficiently compute images under this retraction map-

ping without having to compute the medial axis explicitly.

3. Sampling and retracting in this manner increases the number of nodes found in narrow (small volume) corridors in a way that is independent of the volume of the corridor and depends only on the characteristics of the obstacles bounding it.
4. This improves performance on problems requiring traversal of such corridors.

Note that it is important to know that the medial axis is a strong deformation retract under the retraction mapping being used in order that it be a complete representation for motion planning purposes.

Our theoretical work on the medial axis gives conditions under which the medial axis of a region in the plane is a strong deformation retract under our canonical retraction map, namely that the region have piecewise C^2 boundary. A proof that the medial axis is a strong deformation retract for such regions appears in [45], but our algorithm requires proof of the continuity of our particular retraction map. (Note that our results and results from [45] contrast with a motivational example given in [14].) We also show how the retraction can be extended continuously to points outside the region; this forms the key to our algorithm. To generalize our algorithm to apply to the rigid body case, we extend some of our results to apply to the configuration space manifold $SE(3)$ in which the free space resides. In particular, by assuming a Riemannian metric on $SE(3)$ with particular properties, we give a condition under which the medial axis of the free space (on $SE(3)$) will be a strong deformation retract. We then show that this condition, namely that the medial axis is a closed subset of the free space, is satisfied for a polyhedral workpiece moving among polyhedral obstacles.

Our theoretical work on the medial axis in the plane includes some modifications, and alternate proofs of results that are mostly known in that case. We present these

primarily as motivation for the work in $SE(3)$. We note that there have been some recent (incorrect) examples given in the literature which contradict our results. Our theoretical results for the medial axis on $SE(3)$, our medial axis decomposition result, and the proof of the continuity of the retraction map are new.

The dissertation is organized as follows. After a brief presentation of a few preliminaries, we present in Chapter III a development of the entire algorithm for the case of a point moving in the plane. We then proceed to the development for rigid body motion in 3-D followed by a presentation of some preliminary experimental results.

Earlier versions of this work have appeared in [42] and [43].

CHAPTER II

PRELIMINARIES

In this chapter we outline the required mathematical and motion planning background material.

A. Metric spaces

We first collect a few facts about metric spaces in general.

Let X be a set and $d : X \times X \rightarrow [0, \infty)$. We say (X, d) is a *metric space* if:

1. For $x, y \in X$, $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$ for all $x, y \in X$, and
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

A metric space is, in particular, a topological space, so we have the usual notions of continuity of functions, open and closed sets, etc. In particular, the function d is continuous. See [32].

For $r \geq 0$, define the open and closed metric balls:

$$\overline{B}(x, r) = \{y \in X \mid d(x, y) \leq r\},$$

$$B(x, r) = \{y \in X \mid d(x, y) < r\}.$$

Note that for $r > 0$, $\overline{B}(x, r)$ is the closure of $B(x, r)$ and $B(x, r)$ is the interior of $\overline{B}(x, r)$, but $\overline{B}(x, 0) = \{x\}$ and $B(x, 0) = \emptyset$.

If S is a non-empty subset of X and x is a point in X , we define the distance from x to S by:

$$\text{dist}(x, S) = \inf_{y \in S} d(x, y).$$

If T is also a non-empty subset of X , we define the distance from S to T by:

$$\text{dist}(S, T) = \inf_{x \in S} \text{dist}(x, T).$$

Note that $x \mapsto \text{dist}(x, S)$ is a continuous function of x and that $\text{dist}(\{x\}, T) = \text{dist}(x, T)$ and $\text{dist}(\{x\}, \{y\}) = d(x, y)$ for $x, y \in X$. Also, in a general metric space, a closest point of a set S to a point x (that is, a point $y \in S$ with $d(x, y) = \text{dist}(x, S)$) need not exist or be unique, even if we assume that S is a closed set.

The primary metric spaces of interest to us, the plane and other complete Riemannian manifolds, have the property that the closed and bounded sets are always compact: in the plane this is known as the Heine-Borel Theorem, on a complete Riemannian manifold, it is a consequence of the Hopf-Rinow Theorem (see [16]). The next proposition shows that for a point in a metric space with this property, a nearest point on a closed set does always exist. Also, we show that the map which carries a point x to its closest point in a set S is continuous provided it is well-defined.

Proposition II.1. *Let (X, d) be a metric space in which the closed and bounded sets are compact, and let S be a non-empty closed subset of X .*

1. *For any x in X , there is at least one point $y \in S$ with $d(x, y) = \text{dist}(x, S)$.*
2. *Suppose that for each point x of some subset T of X there is a unique point $s(x) \in S$ with $d(x, s(x)) = \text{dist}(x, S)$. Then the map $x \mapsto s(x)$ is continuous on T .*

Proof. 1. Let $r = \text{dist}(x, S)$. Let $\{a_j\}_{j=1}^{\infty}$ be any sequence in S such that $d(x, a_j) \rightarrow r$ as $j \rightarrow \infty$. For j large enough, $d(x, a_j) < r + 1$, so we have the tail of the sequence contained in the compact set $\overline{B}(x, r + 1) \cap S$. But then $\{a_j\}_{j=1}^{\infty}$ must have a subsequence $\{a_{j_k}\}_{k=1}^{\infty}$ converging to a point $y \in \overline{B}(x, r + 1) \cap S$, and $d(x, y) = r$.

2. Take any sequence $\{x_j\}_{j=1}^{\infty}$ in T converging to $x \in T$. We have

$$\begin{aligned} \text{dist}(x, S) &= \lim_{j \rightarrow \infty} \text{dist}(x_j, S) \\ &= \lim_{j \rightarrow \infty} d(x_j, s(x_j)) \\ &\leq \lim_{j \rightarrow \infty} (d(x_j, x) + d(x, s(x_j))) \\ &= \lim_{j \rightarrow \infty} d(x, s(x_j)). \end{aligned}$$

So for $R = \text{dist}(x, S) + 1$, the point $s(x_j)$ will be inside $\overline{B}(x, R)$ when j is large enough. By compactness of $\overline{B}(x, R)$, the sequence $s(x_j)$ must have a limit point: let $y \in \overline{B}(x, R)$ be any such limit point. But then $\text{dist}(x, S) = d(x, y)$, so by the uniqueness of the point $s(x)$, we must have $s(x) = y$. Therefore, the $s(x_j)$ converge to $s(x)$. \square

Our interest in the nearest point map in part (2) of Proposition II.1 is actually in the particular map taking points inside a set to their nearest points on the boundary of the set. This map is not usually well-defined over the entire set – there are generally some points that will have more than one nearest boundary point. In the spaces of interest to us, these points are always contained in the medial axis of the set, which we will define in the next chapter. The role of the set T in the proposition will be played by the original set minus its medial axis and we will obtain a well-defined and continuous nearest boundary point map.

B. Retraction

Roughly speaking, a *strong deformation retract* of a metric space X is a subset A of X onto which X can be continuously deformed while leaving all points of A fixed throughout the deformation. This concept is useful in motion planning because in

order to solve the motion planning problem on X (where X is viewed as a free space for some planning problem), it is sufficient to know how to solve it on A . It is sometimes possible to choose an A that is simpler than X in some sense. For example, A might be chosen to be some set that is lower dimensional than X , or A might be the union of some simple shapes, splines for instance. We will show later that in certain situations the medial axis of the free space is a strong deformation retract. This is the primary motivation for the algorithm we will present: if the medial axis is a strong deformation retract, there is no harm in choosing the roadmap nodes to lie on the medial axis of the free space.

Let X be a metric space and $A \subseteq X$. The set A is called a *retract* of X if there is a continuous function $r : X \rightarrow A$ such that $r(a) = a$ for all $a \in A$. The set A is called a *strong deformation retract* (SDR) of X if there is a continuous function $h : [0, 1] \times X \rightarrow X$ such that:

1. $h(0, x) = x$ for all $x \in X$,
2. $h(1, x) \in A$ for all $x \in X$,
3. $h(t, a) = a$ for all $a \in A$ and all $t \in [0, 1]$.

The functions h and r are called *retraction maps*.

Any SDR of X is clearly a retract of X by $h(1, \cdot) : X \rightarrow A$. Figure 1 shows an example of a planar region X , an SDR of X , and a retract of X that is not an SDR. Because the points of a retract are fixed points of a continuous function, we can see immediately that a retract must be relatively closed:

Proposition II.2. *Let X be a metric space, A a retract of X . Then A is closed in X .*

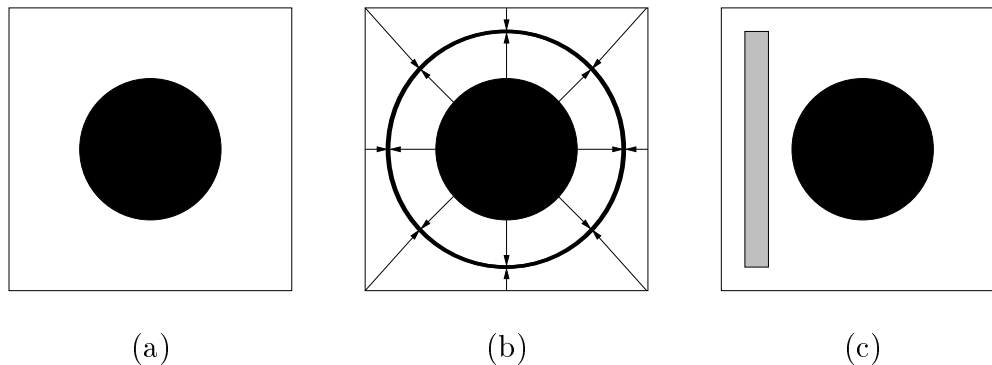


Fig. 1. Retraction examples. Part (a) shows a region X (the square minus the filled circle), part (b) shows an SDR of X (the circle), and part (c) shows a retract of X (the shaded rectangle) that is not an SDR.

Proof. Let $r : X \rightarrow A$ be a retraction map, and let $\{a_j\}_{j=1}^{\infty}$ be a sequence in A converging to $x \in X$. Then $r(x) = r(\lim_{j \rightarrow \infty} a_j) = \lim_{j \rightarrow \infty} r(a_j) = \lim_{j \rightarrow \infty} a_j = x$. Because $r(x) \in A$, we have $x \in A$. \square

Frequently a retract may be shown to be an SDR by using a “straight line” homotopy: suppose $r : X \rightarrow A \subseteq X$ is a retract of a planar region X . If for each $x \in X$, the line segment connecting x to $r(x)$ is contained in X , then we can define $h : [0, 1] \times X \rightarrow X$ by:

$$h(t, x) = (1 - t)x + tr(x).$$

As t increases from 0 to 1, each point of X moves along a line segment connecting it to its final destination on A . The maps we will use to show that (under certain conditions) the medial axis of a region is an SDR are straight line homotopies. Also, this same idea may sometimes be applied on Riemannian manifolds using homotopy along geodesics.

From the topological point of view, the importance of an SDR A of X is that A is

homotopy equivalent to X and so retains the topological structure of X ; in particular it has the same homotopy and homology groups, see [11].

From the point of view of motion planning, we are primarily interested in paths between points: we are generally given start and goal points or configurations x_0 and x_1 in X and asked to find a path between them in X (the free space). See Figure 2.

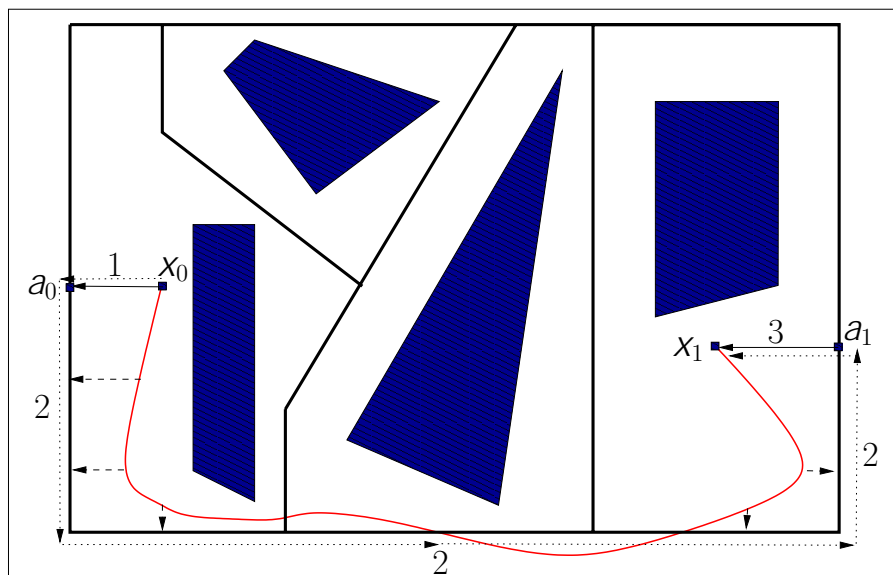


Fig. 2. SDR for motion planning.

Now suppose we know how to solve the corresponding problem on some SDR A of X (the heavy lines in the figure), i.e., given two points a_0 and a_1 of A , we have some method of computing a path in A connecting a_0 and a_1 , provided such a path exists. Then, in order to solve the original problem on X , we can first use the retraction map $h : [0, 1] \times X \rightarrow X$ to retract the start and goal points x_0 and x_1 onto A as shown in Figure 2. The start point x_0 retracts to $a_0 = h(1, x_0)$ and the function $t \mapsto h(t, x_0)$

for $t \in [0, 1]$ is a path in X connecting x_0 to a_0 . Similarly we can retract the goal point x_1 to a point a_1 on A . We now apply our motion planning method for A to obtain a path $\gamma : [0, 1] \rightarrow A$ connecting $a_0 = \gamma(0)$ to $a_1 = \gamma(1)$ on A . The final path from x_0 to x_1 consists of the three pieces: $h(t, x_0), t \in [0, 1]$ (labeled (1) in the figure) taking us from x_0 to a_0 , $\gamma(t), t \in [0, 1]$ taking us from a_0 to a_1 (labeled (2) in the figure), and $h(1 - t, x_1), t \in [0, 1]$ taking us from a_1 to x_1 (labeled (3) in the figure).

One detail has been left out, namely, assuming there is a path in X connecting the points x_0 and x_1 , how do we know that there is always a path in A between the points a_0 and a_1 ? Again using the retraction, any such path in X can be deformed onto a path in A connecting the points a_0 and a_1 . See Figure 2.

The Voronoi diagram is a well-known example of an SDR; it is defined as follows. Given a finite set of points in the plane, the *Voronoi cell* of each site is defined to be the set of all points in the plane that are closer to that site than to any other. The Voronoi diagram is defined to be the boundaries of all of the cells (see [7]). The Voronoi diagram turns out to be an SDR of the plane minus all of the point sites. The SDR that will be of interest to us, the medial axis, is a generalization of the Voronoi diagram.

C. Curves in the plane

The structure of the medial axis of a region in the plane turns out to be intimately related to the properties of the boundary of the region. In this section we briefly review a few points about curves in the plane, see [15] for full details.

If U is an open subset of \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ is called C^k , where k is a positive integer, if f has continuous partial derivatives of order k at each point of U . If f is continuous on U , we say that f is C^0 . If V is any subset of \mathbb{R}^n , we say $f : V \rightarrow \mathbb{R}$

is C^k , $k \geq 0$, if f extends to a C^k function on some open set in \mathbb{R}^n containing V . If $f : V \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, we say f is C^k if each component function of f is C^k on V .

For $k \geq 0$, a C^k curve in the plane is a C^k function $\gamma : [a, b] \rightarrow \mathbb{R}^2$, where $a < b$. If γ is injective, it is called a *simple* curve. If $k \geq 1$ and $\gamma'(t) \neq 0$ for all $t \in [a, b]$, then γ is called a *regular* curve. The image $\gamma([a, b])$ is often called the *trace* of γ . We sometimes use the term “ C^k curve” to refer to a set that is the trace of some C^k curve.

We will mostly be concerned here with C^2 curves in \mathbb{R}^2 ; for the remainder of this section, let $\gamma : [a, b] \rightarrow \mathbb{R}^2$ be a C^2 simple regular curve in the plane.

For each $t \in [a, b]$, the vector $T(t) = \gamma'(t)$ is tangent to the curve at $\gamma(t)$. The number $|T(t)|$ is called the *speed* of γ at t ; we say γ has *unit speed* or *is parametrized by length* if $|T(t)| \equiv 1$.

It is easy to show that any simple regular C^2 curve γ can be reparametrized by length; i.e., there is a C^2 bijection $\tau : [0, L] \rightarrow [a, b]$, $L > 0$, with $\tau(0) = a$, and $\tau(L) = b$ such that $\gamma \circ \tau : [0, L] \rightarrow \mathbb{R}^2$ has unit speed. Also, the curve $\eta : [a, b] \rightarrow \mathbb{R}^2$ given by $\eta(t) = \gamma((b+a) - t)$ has the same image as γ but the image is traced out in the opposite direction. We say that γ and η differ by a *change of orientation*. The parametrization by length of a curve is essentially unique in the sense that if two curves γ_1 and γ_2 are parametrized by length and have the same trace, then either $\gamma_1 \equiv \gamma_2$ or γ_1 and γ_2 differ by change of orientation.

A common example of a C^k curve is the graph of a C^k function $f : [a, b] \rightarrow \mathbb{R}$. The graph of f is the curve $\eta : [a, b] \rightarrow \mathbb{R}^2$ given by $\eta(t) = (t, f(t))$. Note that such a curve is always simple and regular. In fact, if $k \geq 1$, the implicit function theorem tells us that any C^k curve η is given locally as a graph of a C^k function, either by writing y as a function of x or x as a function of y .

1. The Frenet frame

In the next chapter we will analyze the medial axis of a region in terms of the curves bounding the region. It will be convenient to use a specific coordinate system in which curves have a particularly simple form: given a point of interest on a boundary curve, we would like to choose coordinates so that the point appears at the origin of our coordinate system, the curve is tangent to the x -axis at the origin, and the y -axis points into the interior of the region. Also, the new coordinate system should be related to the original coordinates by a rigid transformation of the plane so that the geometry in the new coordinates is the same as the original geometry. While it is reasonable to expect that such a frame should exist we spell out the details here. Borrowing the terminology usually applied to curves in \mathbb{R}^3 , we will call such a coordinate system a *Frenet frame*. See [15] for details of the \mathbb{R}^3 case.

Let $\gamma : [0, L] \rightarrow \mathbb{R}^2$ be a simple regular C^2 curve parametrized by length. The implicit function theorem tells us that near any point on γ we can write γ as the graph of a C^2 function, either by writing y as a function of x or x as a function of y . We can construct the Frenet frame for γ at a point $\gamma(t_0)$ as follows. We have the tangent vector $T = (T_1, T_2)^T : [0, L] \rightarrow \mathbb{R}^2$ defined along γ ; define the *normal vector* $N = (N_1, N_2)^T : [0, L] \rightarrow \mathbb{R}^2$ along γ by $N_1(t) = -T_2(t)$ and $N_2(t) = T_1(t)$. Note that $N \cdot T \equiv 0$, $|N| \equiv |T| \equiv 1$, and that $\det(T \ N) \equiv 1$. The Frenet frame is given by choosing the unit vectors $T(t_0)$ and $N(t_0)$ to be respectively the x and y axes of the Frenet frame at $\gamma(t_0)$. Note that the determinant of the matrix $(T(t_0) \ N(t_0))$ is 1, so the Frenet frame will have the same orientation as the original coordinate frame on \mathbb{R}^2 . Given x and y coordinates of a point in the original coordinate frame, the

coordinates in the Frenet frame at $\gamma(t_0)$ are:

$$\begin{pmatrix} T(t_0) & N(t_0) \end{pmatrix}^T \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \gamma(t_0) \right)$$

It is easy to see that the point $\gamma(t_0)$ appears at the origin in this new system; from the definition, it is clear that the x axis in the Frenet frame is tangent to γ at the frame origin. The implicit function theorem then tells us that in some small neighborhood of the frame origin, the curve γ is given in the Frenet frame by the graph of a C^2 function $f : [-\delta, \delta] \rightarrow \mathbb{R}$, and that $f(0) = f'(0) = 0$. Furthermore, the point $\gamma(t)$ moves from left to right in the Frenet system as t increases.

If we were to change the orientation of γ , the tangent vector $T(t)$ would reverse direction and normal $N(t)$ would “flip over” and point in the opposite direction. So for a given point on a C^2 curve, there are actually two Frenet frames depending on which direction the curve is traversed, but when a particular direction is specified the Frenet frame is uniquely determined. Below we will establish the convention that for boundary curves of a region, we always choose the frame with the y -axis pointing into the interior of the region.

Also, note that by definition, a C^2 curve $\gamma : [a, b] \rightarrow \mathbb{R}^2$ is actually a C^2 function on a slightly larger interval $(a - \epsilon, b + \epsilon)$, $\epsilon > 0$, so the definitions of the tangent, normal, Frenet frame, etc., make sense at the endpoints $\gamma(a)$ and $\gamma(b)$.

2. Curvature

In this section we briefly review the definitions of the curvature and the osculating circle for a plane curve. We also show that the center of the osculating circle at a point can be approximated by the intersections of lines normal to the curve near that point.

The curvature of a plane curve at a point describes how fast the curve is pulling away from its tangent line at that point. Because we assume γ is parametrized by length, i.e., that $|T(t)| = 1$, any change in the tangent vector $T(t)$ represents a change in direction, not a change in length. More precisely, we have $T(t) \cdot T(t) = 1$ for all $t \in [0, L]$. Differentiating this equality, we see that $2 T'(t) \cdot T(t) = 0$ for all $t \in [0, L]$, i.e., $T'(t)$ is perpendicular to $T(t)$. But this means that $T'(t)$ must always be parallel to $N(t)$. This permits us to define the *signed curvature* $\kappa(t)$ by:

$$T'(t) = \kappa(t)N(t).$$

Note that $|T'(t)| = |\kappa(t)|$. The sign of $\kappa(t)$ does depends on the direction in which the curve γ is traversed: reversing the orientation of γ reverses the sign of $\kappa(t)$ because reversing the orientation reverses the sign of N but leaves the sign of T' unchanged. But also note that the signed curvature at a point on the curve will not be changed by an orientation-preserving rigid transformation of the curve. Also, it is clear that because γ is C^2 , the curvature is a continuous function.

For a curve given as the graph of a C^2 function $f : [-a, a] \rightarrow \mathbb{R}$, (i.e., $\gamma(t) = (t, f(t))$) a simple calculation shows that

$$\kappa(t) = \frac{f''(t)}{(1 + f'(t)^2)^{3/2}}.$$

If the curve passes through the origin and is tangent to the x -axis at the origin (as in the Frenet frame for a curve), then

$$\kappa(0) = f''(0).$$

Direct calculation also shows that the curvature of a circle of radius R is equal to $1/R$ at each point. Another interpretation of the curvature is that its absolute value is the reciprocal of the radius of the circle that best approximates the curve at that

point. This best approximating circle is known as the *osculating circle* to the curve at that point. To define this explicitly, suppose we have a curve given as the graph of a C^2 function $f : [-a, a] \rightarrow \mathbb{R}$, with $f(0) = f'(0) = 0$, so that the signed curvature at the origin is $f''(0)$. In order that the osculating circle at least be tangent to the curve at the origin, it must have center on the y -axis: it remains to determine the radius of the circle. Suppose that $f''(0) > 0$. Near the origin, the circle of radius $R > 0$ with center at $(0, R)$ is given by the graph of the function

$$c_R(x) = R - \sqrt{R^2 - x^2}$$

where $x \in [-R, R]$. Note that $c_R(0) = c'_R(0) = 0$ and $c''_R(0) = 1/R$. If we set $1/R = f''(0)$, then the circle at the curve will agree to second order at the origin. This is the *osculating circle* to the curve at the origin. If $f''(0) < 0$, then we will need to have the center of the osculating circle on the negative y -axis instead, with radius $1/R = -f''(0)$. If $f''(0) = 0$, the osculating “circle” is a straight line which we can think of as a circle with center at infinity. In general for a C^2 curve γ , the center of the osculating circle at $\gamma(t)$ is $\gamma(t) + \kappa(t)N(t)$, and the radius is $|\kappa(t)|$. This means that although the sign of $\kappa(t)$ depends on the parametrization, the center and the radius of the osculating circle do not. The osculating circle depends only on the geometry of the curve and not on our choice of parametrization.

The following lemma shows that the center of the osculating circle to a curve at a point can be approximated by the intersections of normal lines to the curve near that point. See Figure 3.

Lemma II.3. *Let $f : (-a, a) \subset \mathbb{R} \rightarrow \mathbb{R}$, $a > 0$ be a C^2 function with $f(0) = f'(0) = 0$. For $x \in (-a, a)$, let $\ell_x(s) = (x, f(x)) + s(-f'(x), 1)$ a parametric equation for the normal line to the graph of f at $(x, f(x))$.*

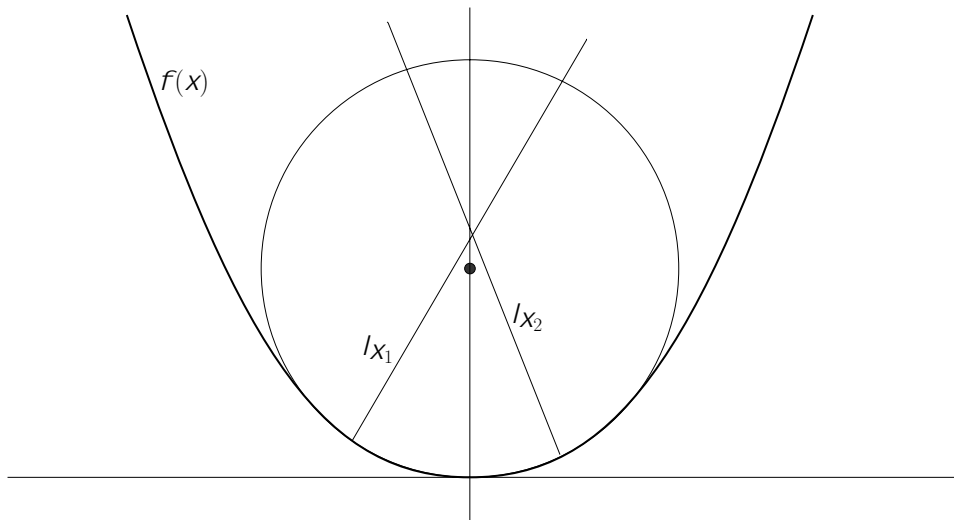


Fig. 3. Intersections of normal lines approximate the center of the osculating circle.

1. If $f''(0) \neq 0$, then given $\epsilon > 0$, there is a $\delta > 0$ such that for any $x_1, x_2 \in (-\delta, \delta)$ with $x_1 \neq x_2$, the normal lines ℓ_{x_1} and ℓ_{x_2} intersect within $B(c, \epsilon)$, where $c = (0, 1/f''(0)) \in \mathbb{R}^2$ is the center of the osculating circle to the graph of f at the origin.
2. If $f''(0) = 0$, then for any $R > 0$ there is a $\delta > 0$ such that for any $x_1, x_2 \in (-\delta, \delta)$ the normal lines ℓ_{x_1} and ℓ_{x_2} intersect outside $B((0, 0), R)$ if they intersect at all.

Proof. Let $x_1, x_2 \in (-a, a)$ with $x_1 \neq x_2$. If the normal lines ℓ_{x_1} and ℓ_{x_2} intersect, then there must be real numbers $s_{x_1x_2}$ and $t_{x_1x_2}$ such that $\ell_{x_1}(s_{x_1x_2}) = \ell_{x_2}(t_{x_1x_2})$, i.e., so that

$$(x_1, f(x_1)) + s_{x_1x_2}(-f'(x_1), 1) = (x_2, f(x_2)) + t_{x_1x_2}(-f'(x_2), 1).$$

Using Cramer's rule, we find that if this system has a unique solution (i.e., if there is

a unique intersection point), then

$$s_{x_1 x_2} = \frac{(x_1 - x_2) - f'(x_2)(f(x_2) - f(x_1))}{f'(x_1) - f'(x_2)}.$$

We take $\sigma(x_1, x_2)$ to be the reciprocal of this expression (which may be infinite):

$$\begin{aligned} \sigma(x_1, x_2) &= \frac{f'(x_1) - f'(x_2)}{(x_1 - x_2) - f'(x_2)(f(x_2) - f(x_1))} \\ &= \frac{\frac{f'(x_1) - f'(x_2)}{x_1 - x_2}}{1 + f'(x_2) \left(\frac{f(x_1) - f(x_2)}{x_1 - x_2} \right)} \end{aligned}$$

Observe that the denominator of this expression converges to $1 + f'(0)^2 = 1$ as x_1 and x_2 approach 0. In particular, we may take $\delta > 0$ small enough that the denominator is non-zero for all $x_1, x_2 \in (-\delta, \delta)$, with $x_1 \neq x_2$. This makes σ a well-defined continuous function mapping $\{(x_1, x_2) \in (-\delta, \delta)^2 \mid x_1 \neq x_2\}$ into \mathbb{R} . But observe that $\sigma(x_1, x_2)$ has a well-defined limit along the diagonal $x_1 = x_2$, i.e., as $x_2 \rightarrow x_1$ for fixed x_1 , namely

$$\frac{f''(x_1)}{1 + f'(x_1)^2}.$$

Because f is C^2 , σ is then a continuous function on all of $(-\delta, \delta) \times (-\delta, \delta)$. In particular, $\sigma(x_1, x_2)$ converges to

$$\frac{f''(0)}{1 + f'(0)^2} = f''(0)$$

as x_1 and x_2 approach 0.

Suppose now that $f''(0) \neq 0$. For $\delta' < \delta$ small enough, $\sigma(x_1, x_2)$ will be nonzero for $x_1, x_2 \in (-\delta', \delta')$. In particular, we will have a well-defined value for $s_{x_1 x_2} = 1/\sigma(x_1, x_2)$, and this approaches $1/f''(0)$ as x_1, x_2 approach 0. The intersection point

$\ell_{x_1}(s_{x_1x_2})$ is then:

$$(x_1, f(x_1)) + s_{x_1x_2}(-f'(x_1), 1)$$

which converges to $(0, 1/f''(0))$, the center of the osculating circle at 0, as x_1 and x_2 approach 0. Therefore, given $\epsilon > 0$ we may take $\delta'' < \delta'$ small enough so that the intersection point is in the ball $B((0, 1/f''(0)), \epsilon)$.

Now suppose that $f''(0) = 0$. For particular $x_1, x_2 \in (-\delta, \delta)$ we have $\sigma(x_1, x_2) \neq 0$ if and only if the normal lines intersect. If $\sigma(x_1, x_2) \neq 0$, then the intersection point is:

$$(x_1, f(x_1)) + \left(\frac{1}{\sigma(x_1, x_2)} \right) (-f'(x_1), 1).$$

But $\sigma(x_1, x_2) \rightarrow f''(0) = 0$ as x_1 and x_2 approach zero. In particular, we may take δ small enough that any such intersection point is outside the ball $B((0, 0), R)$ for a given R . \square

D. Regions in the plane

Our results about the medial axis in the plane will rely on fairly standard assumptions about the smoothness of the boundary of the region. In this section we introduce these assumptions.

Definition II.4. A set $F \subseteq \mathbb{R}^2$ is said to have C^k boundary at $x \in \partial F$, $k \geq 0$, if there is a open neighborhood U of x in \mathbb{R}^2 such that part of the boundary of F inside U is given by a simple regular C^k curve; i.e., there is a simple regular C^k curve $\gamma : [a, b] \rightarrow \mathbb{R}^2$ such that $\partial F \cap U = \gamma((a, b))$. A set $F \subseteq \mathbb{R}^2$ is said to have C^k boundary if it has C^k boundary at each $x \in \partial F$.

Our interest is in the medial axis of a closed planar region; in order to prove results about the medial axis, we will eventually assume that we actually have a set

that is the closure of an open set. Note that F is the closure of an open set if and only if $F = \overline{F^\circ}$.

Suppose $F = \overline{F^\circ}$ has C^2 boundary at $p \in \partial F$, and let $\gamma : [a, b] \rightarrow \mathbb{R}^2$ and U be as in the above definition. We remarked above there are only two essentially distinct unit-speed parametrizations of γ , which differ by a change of orientation. The directions of tangent and normal vector fields and the sign of the signed curvature κ all depend on this choice of orientation. Having a region to which to refer allows us to avoid this complication by selecting an orientation convention for boundary curves. We will always choose the orientation for boundary curves in which the normal to the curve points into the interior of the region F . Another way to say this is that the region is on the left as we walk along the given boundary curve.

To see that this makes sense, shrink U to U' so that γ is given as a graph of a C^2 function $f : [-\delta, \delta] \rightarrow \mathbb{R}$ in the Frenet frame at p , i.e., so that we have

$$\{(x, f(x)) \mid x \in (-\delta, \delta)\} = \partial F \cap U'.$$

Note that $W^+ = \{(x, y) \in U' \mid y > f(x)\}$, and $W^- = \{(x, y) \in U' \mid y < f(x)\}$ are connected open sets that do not meet the boundary of F , i.e., $W^-, W^+ \subseteq F^\circ \cup (\mathbb{R}^2 \setminus F)$. But then W^- is either contained entirely in either F° or $\mathbb{R}^2 \setminus F$, and similarly for W^+ . Because F is closed, W^+ and W^- cannot both be contained in F , otherwise their closure, U' would be contained entirely in F . Because F is the closure of an open set, W^- and W^+ cannot both be contained in $\mathbb{R}^2 \setminus F$. We simply arrange the orientation of the curve so that W^+ is in F and W^- is in $\mathbb{R}^2 \setminus F$.

Having made this fixed choice of orientation for the boundary of a set F that is the closure of an open set, we can then speak of *the* (signed) curvature of the boundary at a boundary point p , which we will write $\kappa(p)$, without reference to a particular parametrization. Similarly, *the* normal vector, by which we mean the inward normal

vector, and *the* tangent vector are well-defined, hence we also have *the* Frenet frame at a boundary point uniquely defined.

Definition II.5. A set $F \subseteq \mathbb{R}^2$ is said to have *piecewise C^2 boundary* if for each point $x \in \partial F$ either:

1. F has C^2 boundary at x , or
2. there is an open neighborhood $U \subset \mathbb{R}^2$ of x such that the part of the boundary of F inside U is given by two simple regular C^2 curves joined at x ; i.e., there are two simple regular C^2 curves $\gamma_1 : [a, b] \rightarrow \mathbb{R}^2$ and $\gamma_2 : [b, c] \rightarrow \mathbb{R}^2$, $a < b < c$, such that
 - (a) $\gamma_1(b) = \gamma_2(b) = x$,
 - (b) $\gamma_1([a, b]) \cap \gamma_2([b, c]) = \{x\}$, and
 - (c) $\gamma_1((a, b)) \cup \gamma_2([b, c)) = \partial F \cap U$.

A boundary point of a region F with piecewise C^2 boundary is called a *knot point* if it satisfies (2) but not (1) in the above definition. It is obvious from the definition that such knot points must be isolated. We always assume that the curves are oriented as described above for the C^2 case.

E. Riemannian manifolds

In this section we summarize a few facts from basic Riemannian geometry that will be needed. See [31] for an introductory treatment or [16] for more details.

- Roughly speaking, a *smooth manifold* M is a higher dimensional analog of a smooth surface in 3-D. It is defined by the property that locally it looks like Euclidean space of some fixed dimension: this defines the dimension of the manifold.

- The definition of a smooth manifold provides a way to define what we mean by smooth (C^∞ differentiable) functions on the manifold and maps between manifolds. Because of this, many results from multi-variable calculus carry over to smooth manifolds.
- There is a well defined notion of the *tangent space* $T_p M$ at any point $p \in M$ which agrees with the usual notion of the tangent plane or line when M is a curve or surface.
- A *Riemannian metric* on a M is a smooth assignment of an inner product on the vector space of tangent vectors at a point; i.e., for a fixed $p \in M$, the Riemannian metric gives an inner product denoted by $\langle v_1, v_2 \rangle_p$. A smooth manifold together with a Riemannian metric is called a *Riemannian manifold*.
- The Riemannian metric provides a way of measuring the length of curves, namely by integrating the length of the tangent vector to the curve (according to the Riemannian metric) along the entire curve. If $\gamma : [a, b] \rightarrow M$ is a C^1 curve, then the length of γ is:

$$\begin{aligned} \text{length}(\gamma) &= \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)}} dt \\ &= \int_a^b |\gamma'(t)| dt \end{aligned}$$

- The ability to measure the length of a curve provides a way to define a distance metric: the distance $d(p, q)$ between two points $p, q \in M$ is the infimum of the lengths of all piecewise C^1 curves connecting them. In general there may be any number of curves, including none at all, that realize that length.

- The intrinsic definition of a *geodesic* requires some additional development which we omit here, but roughly speaking geodesics are curves whose second derivative is zero when viewed from the manifold. Geodesics have the property that they minimize the distance between two points on them provided that the points are close enough together. The geodesics play roughly the same role played by lines in Euclidean geometry.
- A *minimizing geodesic* between two points $p, q \in M$ is a curve $\gamma : [a, b] \rightarrow M$ with $\gamma(a) = p$, $\gamma(b) = q$, having the property that $\text{length}(\gamma) = d(p, q)$.
- A Riemannian manifold with the property that any pair of points $p, q \in M$ have a minimizing geodesic between them is called *complete*.
- The Hopf-Rinow Theorem is usually worded in terms of the *exponential map* on a manifold, but an equivalent statement says that a Riemannian manifold M is complete provided every closed and bounded set in M is compact. See [16] for details.

The plane with its usual Euclidean geometry is an example of a complete Riemannian manifold: the geodesics are lines and there is always a line connecting any pair of points. Note that \mathbb{R}^2 with the origin removed is no longer complete: there is no minimizing geodesic connecting the points $(-1, 0)$ and $(1, 0)$. It turns out that most of the interesting properties of the medial axis fail to be true in such a setting. The manifold we are ultimately interested in is $\text{SE}(3)$ which will be a complete Riemannian manifold with our choice of Riemannian metric.

The sphere $S^2 = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$ in \mathbb{R}^3 can be made into a Riemannian manifold in which the geodesics are the great circles, i.e., the intersections of the sphere with a plane through the origin. This manifold is complete, but observe that

through any pair of points there is always a geodesic that is *not* minimizing.

F. Motion planning for a rigid body

In this section we state precisely the version of the motion planning problem for a rigid body that we will study; see [26] for a complete discussion of motion planning in general.

The version of the motion planning problem for a rigid body that we will study is the following. We are given a nonempty set U of points in \mathbb{R}^3 , the *workpiece*, which we think of as the moving object, and a (possibly empty) set of points V , the *obstacle*, in \mathbb{R}^3 which we think of as fixed. The problem is to move U rigidly, that is preserving distances between all its points, from a given starting position and orientation to a given goal position and orientation, with the restriction that at no time during the motion do points of U overlap points of V . Of course there are examples in which this is not possible: a complete algorithm to solve this problem would find a path if one exists, or else state that no such path exists. In the absence of any obstacle, the workpiece is unconstrained and is allowed to assume any position or orientation.

The configuration space for a rigid body U in \mathbb{R}^3 describes all possible positions and orientations of U ignoring any obstacles that may be present. A particular configuration of a rigid body may be described by specifying the position and orientation of a moving coordinate system attached to U , the *body frame*, with respect to a particular fixed system, the *world frame*. Such coordinate systems are related by a rotation matrix in

$$\text{SO}(3) = \{R \in \mathbb{R}^{3 \times 3} \mid RR^T = I \text{ and } \det(R) = 1\}$$

giving the orientation of the body frame with respect to the world frame, together with a vector in \mathbb{R}^3 specifying the location of the origin of the body frame with respect

to the world frame. (See [6] for a proof of this fact.) We denote the set of all such pairs by $\text{SE}(3) = \text{SO}(3) \times \mathbb{R}^3$. Given a particular configuration $(R, p) \in \text{SE}(3)$, this is interpreted as operating on the body frame coordinates q of a point to produce the world frame coordinates $(R, p) \cdot q = Rq + p$ of that same point when U is in configuration (R, p) . If $g = (R, p)$, we write $g \cdot U$ or $(R, p) \cdot U$ to mean the coordinates of all points of U with respect to the world frame when U is in configuration g , i.e.,

$$g \cdot U = \{g \cdot x \mid x \in U\}$$

Stated precisely, the version of the problem we will study is the following:

- $W = \mathbb{R}^3$ is called the *workspace*.
- The *workpiece* $U \subset \mathbb{R}^3$ is assumed to be a closed and bounded subset of W .
- Possibly by taking the union of all obstacles in the workspace, we assume there is a single obstacle V that is a closed subset of W .
- C denotes the configuration space, $\text{SE}(3)$.
- We define $C(U, V)$ to be the set of all configurations of U in C that cause U to meet V , i.e.,

$$C(U, V) = \{g \in C \mid (g \cdot U) \cap V \neq \emptyset\}.$$

$C(U, V)$ is known as the C -obstacle of V .

- The *blocked space* B is the subset $C(U, V)$ of the configuration space.
- The *free space* of F is the closure of $C \setminus B$.
- Configurations in ∂F are called *contact configurations*: they put U in contact but not overlap with V .

The motion planning problem in this setting is: given g_0 and g_1 in C find explicitly a continuous path $\gamma : [0, 1] \rightarrow F$ such that $\gamma(0) = g_0$ and $\gamma(1) = g_1$, or determine that such a path does not exist.

Note that what we seek here is not simply a declaration that a path exists or not, but an explicit representation of the path, so that $\gamma(t)$ can be evaluated for each $t \in [0, 1]$.

Also, note that with our definitions, contact configurations are always the limit of non-contact free configurations, i.e., configurations in the interior of F .

G. SE(3)

As explained above, the set of all orientation-preserving rigid transformations of \mathbb{R}^3 is given by SE(3). In this section we summarize a few facts about SE(3). See [33] for an excellent introduction.

- SE(3) = SO(3) \times \mathbb{R}^3 , being the product of smooth manifolds is itself a smooth manifold.
- SE(3) also has a multiplicative group structure given by composition of transformations:

$$(R_1, p_1) \cdot (R_2, p_2) = (R_1 R_2, R_1 p_2 + p_1).$$

Algebraically, this gives it the structure of a *semi-direct product*.

- The group structure of SE(3) is compatible with its smooth manifold structure making it a *Lie group*.

Some sort of a notion of distance on the configuration space is frequently needed in motion planning algorithms and our situation will be no different. There are many different distance metrics possible on SE(3) and the choice of metric can have

significant consequences for the performance of the algorithm, see [2]. Defining and using the medial axis on $SE(3)$ will actually require more structure than a simple distance metric: this is where the Riemannian geometry will enter, but there are even many possible choices of Riemannian metric.

One standard method of selecting a Riemannian metric unfortunately does not apply to $SE(3)$, namely coordinate frame invariance. Note that in immediately dealing with our workpiece and obstacles in \mathbb{R}^3 , rather than coordinate-free Euclidean space itself, we have already made a choice of coordinate system. In practice this is necessary for the very presentation of the problem, but such a choice is arbitrary. Coordinate frame invariance mandates that the distance between any two configurations, that is physical positions and orientations, should not depend on how we assigned coordinates on \mathbb{R}^3 . In the terminology of Lie groups, this would be a *bi-invariant* Riemannian metric. However it has been shown in [29] that no such Riemannian metric exists, so coordinate frame invariance is no help here.

Our approach will be to make some assumptions about the Riemannian metric being used and to show that our algorithm works for any metric in that class. See [34, 46] for more detail on such issues in general.

H. Probabilistic roadmap methods

As remarked above, many motion planning problems, including those with many degrees of freedom, can be viewed as planning motions of a point in some higher dimensional configuration space C .

The key motivations behind Probabilistic Roadmap Methods (PRMs) in general are the following:

1. For many problems it is easy to randomly select configurations from the free

space by first selecting a random node from the configuration space and then checking to see it places the workpiece in collision with an obstacle. Collision detection is an active research area, and there are many well-developed algorithms available, see [28] for a survey.

2. For many problems it is possible to produce a simple and efficient “local” planning algorithm that is capable of connecting pairs of nodes in the free space that are not too far apart, but is not capable of solving the general, global problem. In the context of PRMs, such a planner is called the *local planner*.

The idea in PRMs is to take a random sampling of nodes from the free space and attempt to connect them in pairs using a local planner, forming a connected network or *roadmap* of nodes in the configuration space. If enough nodes are sampled, the roadmap has a good chance of reflecting the connectivity of the free space. Given a start and goal configuration in the free space, we can use the local planner to try to connect them to nodes in the roadmap. If this is successful, we then perform a graph search on the roadmap to try to find a path connecting these roadmap nodes. See [25] for more details. See also [2] for a comparative study of various local planning methods.

A summary of the basic PRM is given in Algorithm II.1.

In practice it is generally possible to construct a local planner from a collision detection algorithm: we simply declare that a certain set of paths will be used for the local planner and use the collision detector to check the path for collisions at a series of points along it.

As mentioned in the introduction, it is not difficult to concoct simple examples, even for moving a point in the plane, for which the basic PRM algorithm performs poorly. One need simply arrange to have a very low volume region through which

Algorithm II.1 The basic PRM algorithm

Sampling:

Input. N , the number of nodes to generate.

Output. N nodes in the free space F

- 1: **repeat**
- 2: Generate a uniformly random point p in C .
- 3: Using a collision detection algorithm determine if p is in the free space; if so output it.
- 4: **until** N nodes have been generated

Connection:

Input. N nodes in the free space F

Output. The nodes connected into a roadmap.

- 1: For each pair of nodes, use the local planner to check if they can be connected by a path entirely in the free space.
- 2: If so, output an edge between the two.

Query:

Input. Start and end configurations in F .

Output. A free path connecting the configurations or **failure**.

- 1: **repeat**
 - 2: Try connecting the start configuration to a node in the roadmap using the local planner.
 - 3: **until** Connection is successful or all nodes have been tried.
 - 4: If start configuration was not connected successfully, output **failure**.
 - 5: **repeat**
 - 6: Try connecting the goal configuration to a node in the roadmap using the local planner.
 - 7: **until** Connection is successful or all nodes have been tried.
 - 8: If goal configuration was not connected successfully, output **failure**.
 - 9: Perform a graph search to try to connect the starting node (the node to which the start configuration was connected) to the end node (the node to which the start configuration was connected).
 - 10: If the search is successful output the full path, otherwise output **failure**.
-

a solution path would be required to pass. Then assuming uniform sampling with respect to volume, the algorithm will have a very small chance of finding a node in that region. This might produce a roadmap with more connected components than the free space. For this reason, most PRM methods include some sort of re-sampling step: after the roadmap has been constructed, some heuristic is generally applied to try to improve the connectivity of the roadmap. See [21] for example.

The algorithms presented in [1] and [19], discussed in the introduction, are specifically aimed at modifying the sampling step in order to find additional nodes in low volume corridors. MAPRM, the method we present in this dissertation, is another modification to the sampling strategy. Rather than discarding sampled configurations that do not land in the free space, MAPRM keeps all samples and retracts them onto the medial axis of the free space. It turns out that this improves performance on low volume corridor problems.

CHAPTER III

MAPRM IN THE PLANE

In this chapter we define the medial axis of a region in the plane and investigate conditions under which the medial axis is an SDR of the original region under a simply defined *canonical retraction map*. We investigate the details of this retraction map and show how it may be extended continuously to nearly the entire plane. Most significantly, the details of the retraction map will make it clear how to compute the retracted image of a point without having to first compute the medial axis. In the context of motion planning, this will allow us to retract sampled points onto the medial axis of the free space without computing the medial axis explicitly.

A. Definition

For any proper¹ subset S of \mathbb{R}^2 we observe that $\partial S \neq \emptyset$. We define the *boundary distance* function $\rho_S : S \rightarrow \mathbb{R}$ by

$$\rho_S(x) = \text{dist}(x, \partial S) \text{ for } x \in S.$$

Let F be a proper closed subset of \mathbb{R}^2 . Note that for $x \in F$, the largest closed ball centered at x that is still contained in F is $\overline{B}(x, \rho_F(x))$. We will use the notation $B_F(x) = \overline{B}(x, \rho_F(x))$, and we refer to this as the *maximal (closed) ball in F centered at x* . See Figure 4.

The *medial axis* $\text{MA}(F)$ of F in \mathbb{R}^2 is defined as follows: a point x is on the medial axis if there is no point y in F whose associated maximal ball properly contains the

¹By *proper* we mean that the set and its complement are non-empty.

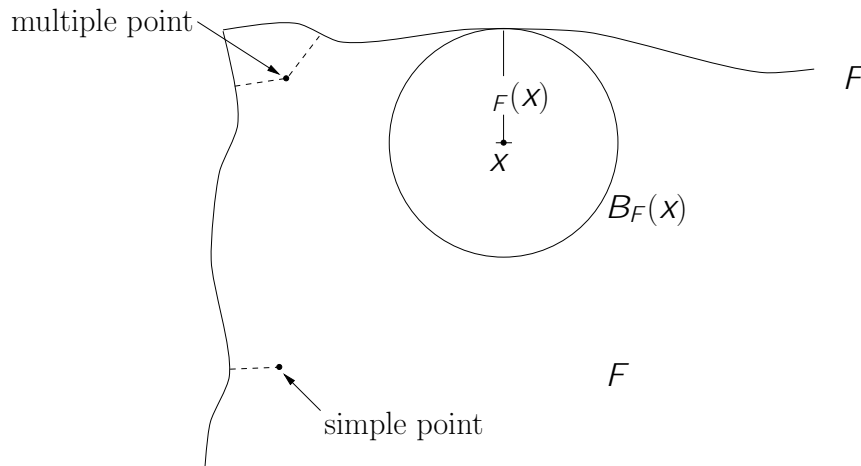


Fig. 4. The maximal ball at a point.

maximal ball centered at x ; i.e.,

$$\text{MA}(F) = \{x \in F \mid \nexists y \in F \text{ with } B_F(x) \subsetneq B_F(y)\}.$$

The medial axis is sometimes called the “generalized Voronoi diagram” or “skeleton” of the set F .

Another way to say this is that the medial axis of F in \mathbb{R}^2 is the set of all points x in F for which associated $B_F(x)$ is maximal with respect to containment among all $B_F(y)$ for $y \in F$. For $x \in \text{MA}(F)$, we call $B_F(x)$ a *maximal ball in F* .

We refer to points $y \in \partial F$ satisfying $\rho_F(x) = d(x, y)$ as *nearest boundary points of F to x* . Observe that because ∂F is closed, any $x \in F$ always has at least one nearest boundary point. A point $x \in F$ is called a *simple point* if x has a unique nearest boundary point, i.e., if $B_F(x)$ meets the boundary of F in a single point. Otherwise, x is called a *multiple point*. See Figure 4.

Example III.1. Figure 5 shows several examples of the medial axis of a planar region. Figure 5(a) shows a polygon and its medial axis. Figure 5(b) shows an example of a connected set with disconnected medial axis.

The region in Figure 5(c) has infinitely many triangular “teeth” which causes the medial axis to have infinitely many “prongs” sticking up. If we take x, y coordinates as shown, our region F is the union of the following sets: two discs, $B((0, 0), 1)$ and $B((0, 2), 1)$, a rectangle $[0, 2] \times [-1, 1]$, and one triangle for each $n = 0, 1, 2, \dots$ having vertices: $(2^{-(n-1)}, 1)$, $(2^{-n}, 1)$, and $((2^{-(n-1)} + 2^{-n})/2, 5/4)$. There are then infinitely many medial axis segments reaching to the highest vertex on each triangle. Note that the only medial axis point on the y -axis is the origin, so the medial axis is not closed in this example. Therefore $\text{MA}(F)$ cannot be an SDR of the region by Proposition II.2.

In Figure 5(d), F is the region above the graph of a parabola. Note that the endpoint of the medial axis is a simple point in the interior of the region. That point is the center of the osculating circle to the boundary at the single contact point, the vertex of the parabola.

There are also other commonly used definitions of the medial axis. In [36], an open region is assumed and maximal open balls are used to define the medial axis. This avoids certain pathologies as the authors explain. For example, Figure 5(b) shows that with our definition the medial axis of a connected set is not necessarily connected. However, our eventual interest is to extend the retraction maps through the boundary of the region and into most of the plane; for this reason we follow [14] and deal with closed sets from the outset. Our approach will be to eliminate the pathologies by deleting particular points from the region.

There is a simple relationship between the definitions using closed and open sets:

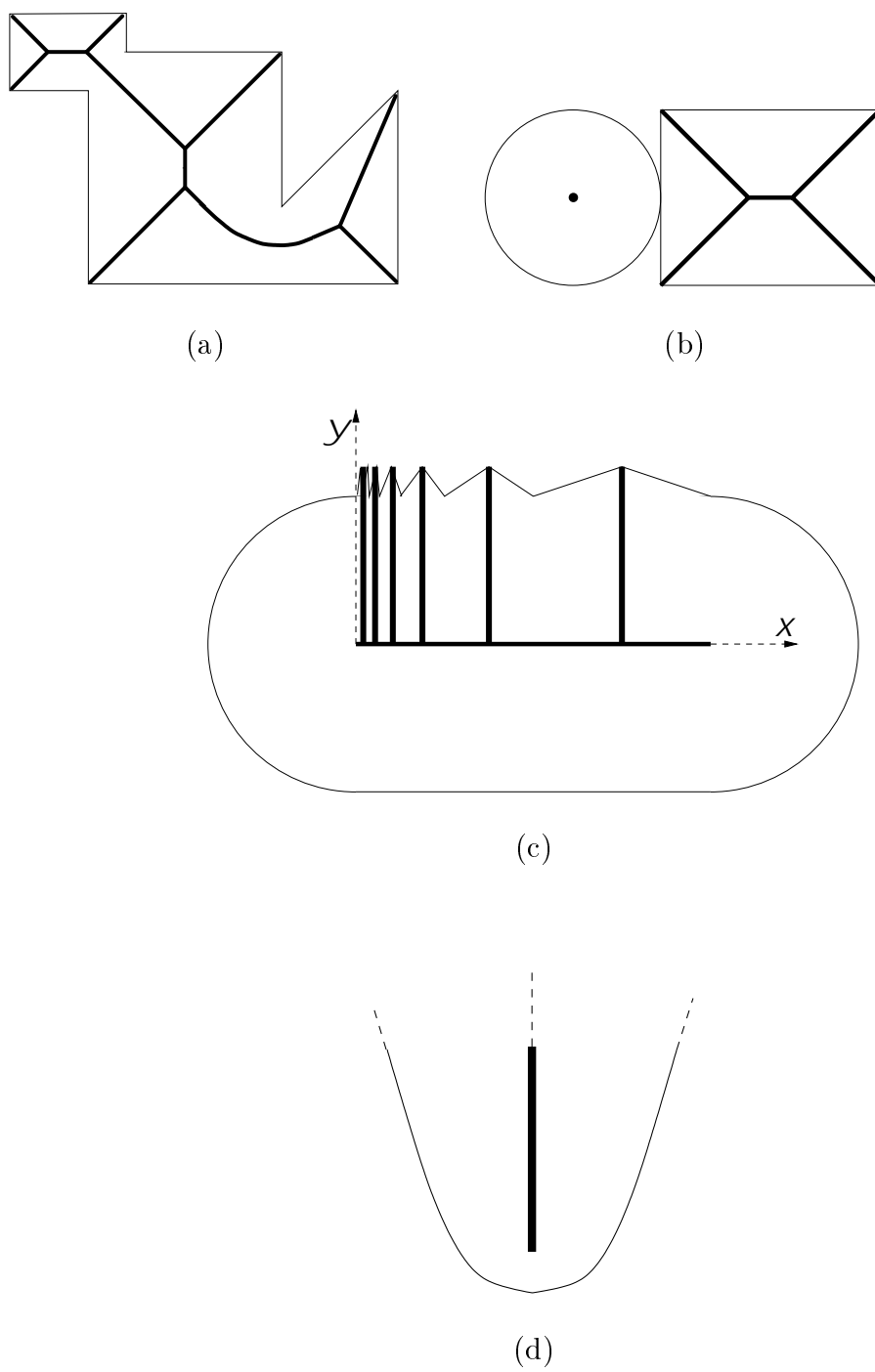


Fig. 5. Some regions and their medial axes.

for a proper open subset E of the plane, the maximal open ball centered at a point $x \in E$ is just $B(x, \rho_E(x))$, so using open balls we define:

$$\text{MA}^\circ(E) = \{x \in E \mid \nexists y \in E \text{ with } B(x, \rho_E(x)) \subsetneq B(y, \rho_E(y))\}.$$

Proposition III.2. *Let F be a proper closed subset of \mathbb{R}^2 . Then*

$$\text{MA}(F) \cap F^\circ = \text{MA}^\circ(F^\circ).$$

Proof. It is easy to show that $\rho_F \equiv \rho_{F^\circ}$ on F° . But then for any $x \in F^\circ$, we have $\overline{B}(x, \rho_F(x)) \subseteq F$ if and only if $B(x, \rho_F(x)) \subseteq F^\circ$. It follows that $\text{MA}(F) \cap F^\circ = \text{MA}^\circ(F^\circ)$. \square

Another commonly used definition is to take the medial axis to be the set of all multiple points. This definition is motivated by the usual definition of the Voronoi diagram (see [7]) for a finite number of point sites as the set of points equidistant from two or more sites. We define:

$$\text{Vor}(F) = \{x \in F \mid x \text{ is a multiple point of } F\}.$$

In the polygonal case this agrees with our definition of the medial axis, at least on the interior of the polygon. However, as Figure 5(d) shows, the medial axis may well contain simple points in F° . Also, if F has no interior, then there are no multiple points at all. It is shown in [36] that if E is a bounded open set, then in fact $\text{Vor}(E)$ is dense in $\text{MA}^\circ(E)$.

For any proper closed subset of \mathbb{R}^2 , $\text{Vor}(F)$ is always contained in $\text{MA}(F)$:

Lemma III.3. *Let $x_1, x_2 \in \mathbb{R}^2$ and $r_2 \geq r_1 > 0$ with $B(x_1, r_1) \subseteq B(x_2, r_2)$. Suppose that there is a point y in $\partial B(x_1, r_1) \cap \partial B(x_2, r_2)$. Then x_1, x_2 , and y are collinear.*

Proof. The circles $\partial B(x_1, r_1)$ and $\partial B(x_2, r_2)$ must be tangent at the point y , hence

x_1, x_2 , and y are collinear. □

Proposition III.4. *Let F be a proper closed subset of \mathbb{R}^2 . Multiple points of F are contained in $\text{MA}(F)$, i.e., $\text{Vor}(F) \subseteq \text{MA}(F)$.*

Proof. Let $x \in F$ be a multiple point with distinct nearest boundary points $x', x'' \in \partial F$ and suppose that $B_F(x) \subseteq B_F(y)$ for some $y \in F$. By Lemma III.3, y is on the lines $\overset{\leftarrow}{x'}x$ and $\overset{\leftarrow}{x''}x$. If these lines only intersect at x , then $y = x$. Otherwise, these lines coincide, i.e., x, x' , and x'' are collinear and y is also on this line. Because $x', x'' \in \partial B_F(y)$, we have $d(y, x') = d(y, x'')$, so again $y = x$. □

The following is a useful rewording of Lemma III.3:

Lemma III.5. *Let F be a proper closed subset of \mathbb{R}^2 and let $x, y \in F$. Then $B_F(x) \subseteq B_F(y)$ if and only if $\rho_F(y) = \rho_F(x) + d(x, y)$.*

Proof. Let x' be a nearest boundary point of x .

Suppose $\rho_F(y) = \rho_F(x) + d(x, y)$. Then for $a \in B_F(x)$,

$$\begin{aligned} d(a, y) &\leq d(a, x) + d(x, y) \\ &\leq d(x', x) + d(x, y) \\ &= \rho_F(x) + d(x, y) \\ &= \rho_F(y) \end{aligned}$$

which implies that $a \in B_F(y)$.

Now suppose $B_F(x) \subseteq B_F(y)$; by Lemma III.3, x, x' , and y are collinear. But

then:

$$\begin{aligned}\rho_F(y) &= d(y, x') \\ &= d(x', x) + d(x, y) \\ &= \rho_F(x) + d(x, y).\end{aligned}$$

□

B. The canonical retraction map

To see how to retract a region F , such as the polygon in Figure 5(a), onto its medial axis, we observe that for most points x in the region there is a unique $a_x \in \text{MA}(F)$ such that $B_F(a_x)$ contains the maximal disc $B_F(x)$ at x ; we would like to define a map taking x to a_x . In this section we explain this map in detail and investigate the exact circumstances under which it gives a well-defined continuous retraction map.

Let F be a proper closed subset of \mathbb{R}^2 , and assume for the moment that F is bounded. Because $\text{MA}(F)$ contains all multiple points of F , any interior point x of F not on the medial axis must have a unique nearest point x' on the boundary of F . Any ball $B_F(y)$ containing $B_F(x)$ must contain x' on its boundary, so by Lemma III.3, y must be on the line joining x and x' . By starting at x and moving away from x' along this line, we will eventually reach a point a_x on the medial axis with $B_F(x) \subseteq B_F(a_x)$; in this manner we can retract interior points onto the medial axis. See Figure 6. This is the idea behind the *canonical retraction map*. Note that if this map is continuous, then because metric balls in the plane are convex, we can use a straight line homotopy to show that this is actually a strong deformation retraction.

There are two potential problems in defining this map for general proper closed sets $F \subset \mathbb{R}^2$. First, if F is not bounded, then it is possible that for some $x \in F$

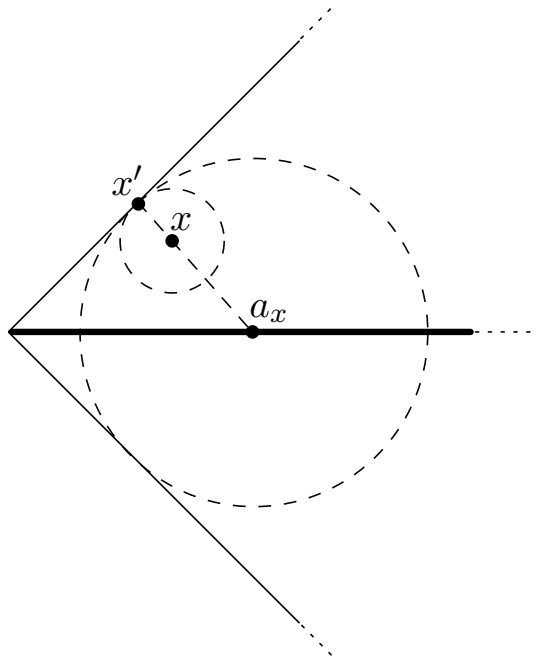


Fig. 6. The image of a point x under the canonical retraction map.

there is no maximal disc containing $B_F(x)$, but just an ever-increasing chain of discs. Second, we would like to map all of F , not just its interior, onto the medial axis using retraction along lines, but this is not necessarily well-defined at boundary points without a well-defined tangent. After resolving these issues, we will show that this retraction along lines is a continuous function if and only if the medial axis of the region turns out to be a closed set; in particular, this shows that the medial axis is an SDR of F whenever the medial axis is closed. (Under slightly different definitions, results in [36] actually characterize the points of discontinuity of this map for general regions: these are precisely the points in the closure of the medial axis that are not in the medial axis itself.)

Let F be any proper closed subset of \mathbb{R}^2 . As we have said, for $x \in F$, there is not necessarily a point $a_x \in \text{MA}(F)$ with $B_F(x) \subseteq B_F(a_x)$. For example, if we take F to

be the closed upper half plane, then for any point $x \in F$ there are arbitrarily large balls containing $B_F(x)$, so there is no point a_x in $\text{MA}(F)$ with $B_F(x) \subseteq B_F(a_x)$. There are, in fact, no maximal balls whatsoever, so the medial axis is empty. To avoid this complication, we follow [36] and simply exclude regions containing entire closed half planes. Because we are primarily interested in motion planning applications where the free space is generally a bounded set, this will be sufficient for our purposes.

Proposition III.6. *Let F be a proper closed subset of \mathbb{R}^2 not containing a half plane. Then for any x , there is an $a_x \in \text{MA}(F)$ such that $B_F(x) \subseteq B_F(a_x)$.*

Proof. Suppose for some $x \in F$ there is no $a_x \in \text{MA}(F)$ with $B_F(x) \subseteq B_F(a_x)$. Let x' be a closest boundary point to x . $B_F(x)$ is not maximal, so there must be a $y \in F$ with $B_F(x) \subsetneq B_F(y)$. Because there is no maximal disc containing $B_F(x)$, we have $B_F(x) \subseteq B_F(a)$ for all $a \in \overrightarrow{xy}$. The union of these $B_F(a)$ is a closed half plane in F . \square

We will also need the following related statement:

Proposition III.7. *Let F be a proper closed subset of \mathbb{R}^2 . Then F contains a half plane if and only if there is a point of F contained in arbitrarily large metric balls contained in F .*

Proof. Suppose F contains a half plane H . Any point on the boundary of H will be contained in arbitrarily large metric balls contained in H . More precisely, if $H = \{p \in \mathbb{R}^2 \mid (p - x_0) \cdot v \geq 0\}$, where v is a unit vector, then any ball $\overline{B}(x_0 + s v, s)$ will be contained in H .

Now, suppose that there is an $x \in F$ contained in arbitrarily large metric balls contained in F : for $j = 1, 2, \dots$ let c_j be the center of a ball of radius at least j containing x . Without loss of generality, we may assume that $c_j \neq x$ for all j . The unit

direction vectors $v_j = (c_j - x)/|c_j - x|$ in S^1 must have an accumulation point $v \in S^1$. Let $\epsilon > 0$ and let $a = x + \epsilon v$. We claim that the half plane $\{y \in \mathbb{R}^2 \mid (y - a) \cdot v \geq 0\}$ is contained in F .

It suffices to show that given any $R > 0$, the disc $B(a + Rv, R)$ is contained in F . Given an $R > 0$, because $v_j \rightarrow v$, there must exist some $j > R + \epsilon$ such that $d(x + (R + \epsilon)v_j, x + (R + \epsilon)v) < \epsilon$. If $y \in B(a + Rv, R)$, then

$$\begin{aligned} d(y, x + (R + \epsilon)v_j) &\leq d(y, x + (R + \epsilon)v) + d(x + (R + \epsilon)v, x + (R + \epsilon)v_j) \\ &= d(y, (x + \epsilon v) + Rv) + d(x + (R + \epsilon)v, x + (R + \epsilon)v_j) \\ &= d(y, a + Rv) + d(x + (R + \epsilon)v, x + (R + \epsilon)v_j) \\ &\leq R + \epsilon, \end{aligned}$$

i.e., y is in $B(a + (R + \epsilon)v_j, R + \epsilon) \subseteq B(c_j, j) \subseteq F$, therefore $B(a + Rv, R) \subseteq F$. \square

We now know that provided F does not contain a half plane, for $x \in F$ there always exists an a_x in $\text{MA}(F)$ such that $B_F(x) \subseteq B_F(a_x)$. We note in passing that any such F may be reconstructed from its medial axis and the values of ρ_F on $\text{MA}(F)$:

$$F = \bigcup_{y \in \text{MA}(F)} B_F(y).$$

The pair $(\text{MA}(F), \rho_F|_{\text{MA}(F)})$ is sometimes referred to as the *medial axis transform* of F . This concept has been used in computer aided geometric design, particularly in three dimensions, as a means of creating and representing solid objects, see, e.g., [18, 41]. In particular, the operations of dilation and erosion are simplified in this representation, see [36].

We now turn to the question of the uniqueness of the destination point a_x . For points x on the interior of F , there is always a unique point $a_x \in \text{MA}(F)$ such that $B_F(x) \subseteq B_F(a_x)$. We can see this as follows (see Figure 6). An interior point x always

has at least one nearest boundary point $x' \neq x$. If we are to have $a_x \in \text{MA}(F)$ with $B_F(x) \subseteq B_F(a_x)$, then clearly $x' \in \partial B_F(a_x)$. But then Lemma III.3 tells us that x' , x , and a_x are collinear, so the point a_x must be unique.

Uniqueness problems occur at boundary points not already on the medial axis that are included in discs with centers in varying directions. For example, the non-convex corners of a polygon are such points, see Figure 5(a). In the case of a polygon, such corners are convex corners of the complement of the polygon, so they are actually on the medial axis of the complement. (This is sometimes called the *exo-skeleton* of the set.) This is always true of points at which such uniqueness problems occur:

Proposition III.8. *Let F be a proper closed subset of \mathbb{R}^2 , $x \in \partial F$. Suppose that there are distinct points a_1 and a_2 in $F \setminus \{x\}$ such that $x \in B_F(a_1)$ and $x \in B_F(a_2)$ and that a_1 , a_2 , and x are not collinear. Then $x \in \text{MA}(\overline{\mathbb{R}^2 \setminus F})$.*

Proof. Let $A = \overline{\mathbb{R}^2 \setminus F}$ and suppose that there is a point $b \in A$ with $x \in \partial B_A(b)$. Then $\partial B_A(b)$ must be tangent to $B_F(a_1)$ and $B_F(a_2)$ at x , so $b \in \overleftrightarrow{xa_1} \cap \overleftrightarrow{xa_2}$, which implies that $b = x$. Therefore $x \in \text{MA}(A)$. \square

We would like to exclude such points, i.e., points x for which there is not a unique point $a_x \in \text{MA}(F)$ with $B_F(x) \subseteq B_F(a_x)$, from the domain of the retraction map. The previous proposition shows that it would certainly suffice to exclude points that are in the medial axis of the closure of the complement of F , but this could unnecessarily remove some points of $\text{MA}(F)$ as shown in Figure 7.

Adding these points back in, we define:

$$F' = \left(F \setminus \text{MA}(\overline{\mathbb{R}^2 \setminus F}) \right) \cup \text{MA}(F).$$

Figure 8 shows F' for two of the domains from Figure 5.

Putting this all together, assume F is a proper closed subset of \mathbb{R}^2 not containing

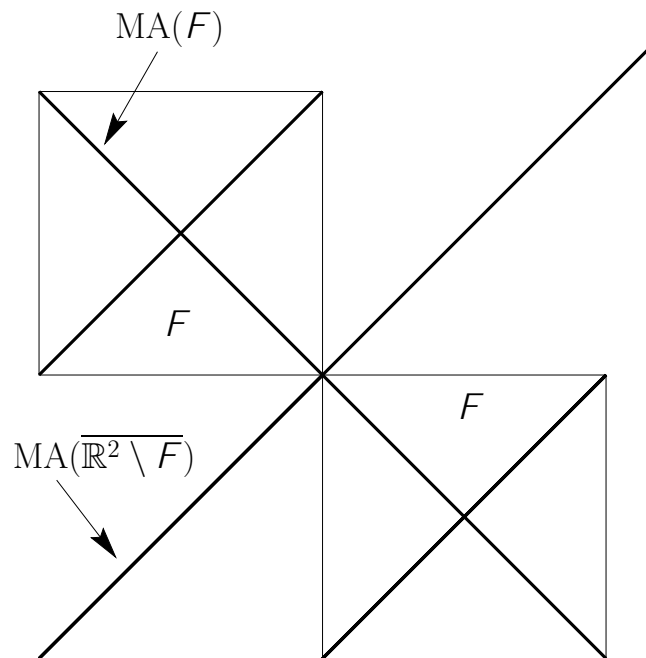


Fig. 7. The medial axis of F may have points in common with $\text{MA}(\overline{\mathbb{R}^2 \setminus F})$.

any half planes. Propositions III.6 and III.8 allow us to define the *canonical retraction map* $r_F : F' \rightarrow \text{MA}(F)$ by mapping each point $x \in F'$ to the unique point $r_F(x) \in \text{MA}(F)$ satisfying $B_F(x) \subseteq B_F(r_F(x))$, i.e., we have:

Proposition III.9. *Let F be a proper closed subset of \mathbb{R}^2 not containing any half planes. For each point $x \in F'$, there is a unique point $r_F(x) \in \text{MA}(F)$ such that $B_F(x) \subseteq B_F(r_F(x))$.*

We will now show that this map is continuous if and only if $\text{MA}(F)$ is closed. With this result, it is then easy to see that if $\text{MA}(F)$ is closed, $\text{MA}(F)$ is a strong deformation retract of F' .

Theorem III.10. *Let F be a proper closed subset of \mathbb{R}^2 not containing a half plane. The canonical retraction map $r_F : F' \rightarrow \text{MA}(F)$ is continuous if and only if $\text{MA}(F)$*

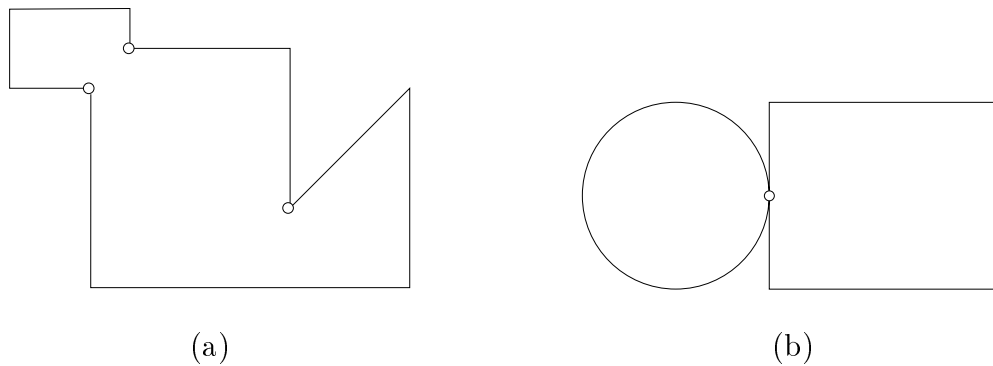


Fig. 8. F' for regions (a) and (b) from Figure 5.

is closed in \mathbb{R}^2 .

Proof. Suppose the map r_F is continuous, and let $\{x_j\}_{j=1}^{\infty}$ be a sequence in $\text{MA}(F)$ converging to $x \in F$. If $x \in F'$, then x is the limit of fixed points of r_F and so must be fixed by r_F . But points of F' are fixed by r_F if and only if they are in $\text{MA}(F)$, hence $x \in \text{MA}(F)$. This shows that $\text{MA}(F)$ is relatively closed in F' . Now suppose $x \notin F'$. Then x must be in ∂F but not in $\text{MA}(F)$, so there is at least one maximal disc $B_F(a)$, $a \in \text{MA}(F)$, that contains x . By definition of r_F , all points of the segment \overline{xa} map to a under r_F . But there is also a sequence of fixed points x_j of r_F converging to x ; this contradicts the continuity of r_F .

Now assume $\text{MA}(F)$ is closed and let $\{x_j\}_{j=1}^{\infty}$ be a sequence in F' converging to x in F' . We claim that $\{r_F(x_j)\}_{j=1}^{\infty}$ is bounded. By Lemma III.5, we have $\rho_F(r_F(x_j)) = \rho_F(x_j) + d(x_j, r_F(x_j))$ each j , and $x_j \rightarrow x$, so if $\{r_F(x_j)\}_{j=1}^{\infty}$ were unbounded, then $\{\rho_F(r_F(x_j))\}_{j=1}^{\infty}$ would be unbounded. But then for j large enough, we would have $x \in B_F(x_j)$, so this would imply that x is contained in arbitrarily large metric balls, contradicting Proposition III.7.

Take $y \in F$ to be any limit point of $\{r_F(x_j)\}_{j=1}^{\infty}$, say $r_F(x_{j_k}) \rightarrow y$, as $k \rightarrow \infty$,

$k = 1, 2, \dots$. Because $\text{MA}(F)$ is closed, we have $y \in \text{MA}(F)$. Then

$$\rho_F(r_F(x_{j_k})) = \rho_F(x_{j_k}) + d(x_{j_k}, r_F(x_{j_k})),$$

so by continuity of ρ_F and d , we have:

$$\rho_F(y) = \rho_F(x) + d(x, y),$$

i.e., $B_F(x) \subseteq B_F(y)$. But $y \in \text{MA}(F)$, so $y = r_F(x)$, hence $r_F(x_j) \rightarrow r_F(x)$. \square

Corollary III.11. *Let F be a proper closed subset of \mathbb{R}^2 not containing a half plane. Suppose that $\text{MA}(F)$ is closed. Then $\text{MA}(F)$ is a strong deformation retract of F' under the homotopy $h : [0, 1] \times F' \rightarrow F'$ given by:*

$$h(t, x) = (1 - t)x + t r_F(x).$$

Proof. Note that for any $x \in F'$, the line segment $h([0, 1], x)$ is contained in $B_F(r_F(x)) \subseteq F'$. The function r_F is continuous by Theorem III.10. \square

We sometimes also refer to the homotopy map h above as the *canonical retraction map*.

Our goal is to be able to compute the image of any point under this retraction map without computing the medial axis explicitly. We briefly describe here how to compute images under the canonical retraction map, at least for interior points of F ; the algorithm will be given in detail later in this chapter. Take F to be a proper closed subset of \mathbb{R}^2 not containing a half plane, and let x be a point in the interior of F . If x has more than one nearest boundary point, then x is already on the medial axis, so $r_F(x) = x$. Otherwise, x must have a unique nearest point $x' \neq x$ on the boundary of F . See Figure 9. We know that x' , x , and $r_F(x)$ must be collinear. Observe that any point y on this line that is strictly between x' and $r_F(x)$ has x'

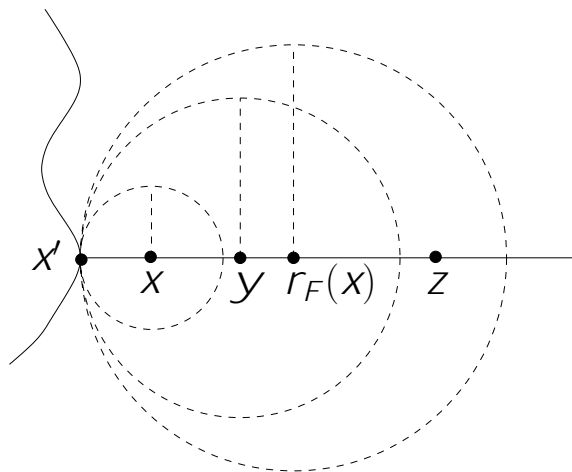


Fig. 9. Computing the canonical retraction map.

as its only nearest boundary point, and any point z “past” $r_F(x)$ (i.e., on the ray $\overrightarrow{x' r_F(x)}$ but not on the segment connecting x' and $r_F(x)$) cannot have x' as a nearest boundary point. This allows us to use a bisection method to compute $r_F(x)$: i.e., we can “squeeze” $r_F(x)$ between points on the line (like y) for which x' is a nearest boundary point and points on the line (like z) for which it is not. Note that $r_F(x)$ has x' as a nearest boundary point, but does not necessarily have any others.

C. The extended retraction map

Figure 10 shows the lines along which points move during the canonical retraction onto the medial axis. Except near the convex corners of the polygon, it is clear that the domain of the retraction may actually be extended into the complement slightly by just extending the lines of retraction. In fact, any point in the complement with a unique nearest point on the boundary of the polygon may be retracted along a line to that boundary point. If the boundary point is in the domain of the canonical retraction map, we may then retract onto the medial axis. This idea will allow us to

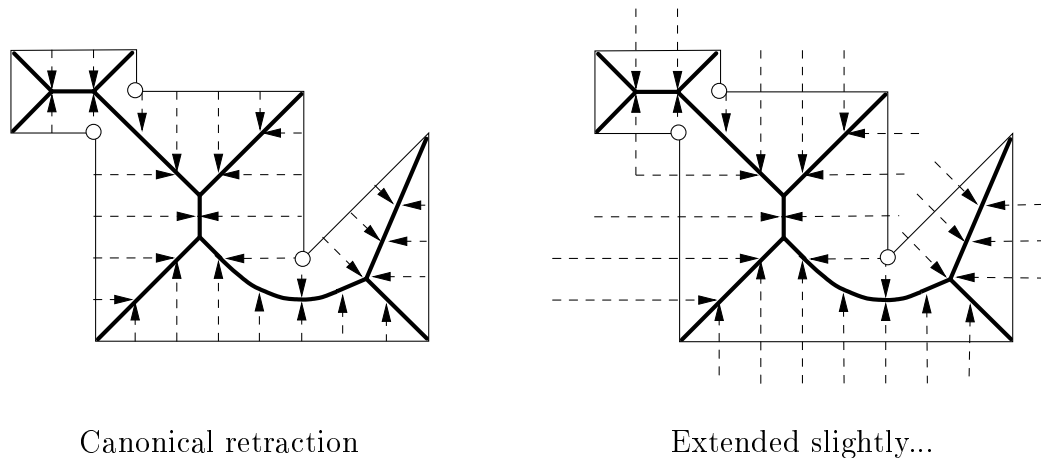


Fig. 10. The canonical and extended retraction maps.

extend the retraction map to nearly the entire plane.

Recall that for any proper closed subset S of \mathbb{R}^2 , $\text{Vor}(S)$ is the set of all multiple points of S . We can define a map $b_S : S \setminus \text{Vor}(S) \rightarrow \partial S$ by taking each point to its closest point on ∂S . Proposition II.1 tells us that this map is continuous.

Take F to be a proper closed subset of \mathbb{R}^2 not containing a half plane, so that the canonical retraction map is well-defined. For points in $B = \overline{\mathbb{R}^2 \setminus F}$, we would like to apply b_B to map them into $\partial B = \partial F$ and then use r_F to retract them onto $\text{MA}(F)$. The problem is that the image of b_B may contain points not in the domain of r_F . From the previous section, we know that such points are on the medial axis of B , so we would like to exclude from the domain of b_B any points that map into $\text{MA}(B)$.

Clearly any point of ∂B is fixed by b_B , but it is also clear that any point on $\text{MA}(B)$ that is also on the boundary cannot be a closest boundary point of any other point of B . We know that $\text{Vor}(B) \subseteq \text{MA}(B)$, so we have:

$$b_B(B \setminus \text{MA}(B)) \subseteq \partial B \setminus \text{MA}(B).$$

Therefore, we have the following composition:

$$B \setminus \text{MA}(B) \xrightarrow{b_B} \partial B \setminus \text{MA}(B) \xrightarrow{=} \partial F \setminus \text{MA}(B) \xrightarrow{r_F} \text{MA}(F)$$

Note that again we may have unnecessarily excluded points of $\text{MA}(F)$ that are also in $\text{MA}(B)$. We define the domain \tilde{F} by:

$$\tilde{F} = \text{MA}(F) \cup (\mathbb{R}^2 \setminus \overline{\text{MA}(\mathbb{R}^2 \setminus F)})$$

We can now define the *extended retraction map* $\tilde{r}_F : \tilde{F} \rightarrow \text{MA}(F)$ by:

$$\tilde{r}_F(x) = \begin{cases} r_F(x) & \text{if } x \in F^\circ, \\ r_F(b_B(x)) & \text{otherwise.} \end{cases}$$

Because points on the boundary ∂F are fixed by b_B it is clear that \tilde{r}_F is continuous whenever r_F is continuous.

As we would expect, a point x of $B \setminus \text{MA}(B)$ can be retracted onto the medial axis of F along a single line segment connecting x , $b_B(x)$, and $\tilde{r}_F(x)$:

Proposition III.12. *Let F be a proper closed subset of \mathbb{R}^2 not containing a half plane, let $B = \overline{\mathbb{R}^2 \setminus F}$, and let $\tilde{r}_F : \tilde{F} \rightarrow \text{MA}(F)$ be the extended retraction map. If $x \in B \setminus \text{MA}(B)$, then the points x , $b_B(x)$, and $\tilde{r}_F(x)$ are collinear.*

Proof. Assume the points x , $b_B(x)$, and $\tilde{r}_F(x)$ are distinct. The line segment $\overline{x \tilde{r}_F(x)}$ must contain some point $y \in \partial F$, so that $d(x, \tilde{r}_F(x)) = d(x, y) + d(y, \tilde{r}_F(x))$. If the points x , b_B , and $\tilde{r}_F(x)$ are not collinear, then $d(x, \tilde{r}_F(x)) < d(x, b_B(x)) + d(b_B(x), \tilde{r}_F(x))$, so we must have $d(x, y) < d(x, b_B(x))$, which would contradict the definition of b_B , or $d(y, \tilde{r}_F(x)) < d(b_B(x), \tilde{r}_F(x))$, which would contradict the fact that $b_B(x)$ is a nearest boundary point of $r_F(b_B(x)) = \tilde{r}_F(x)$. \square

This also tells us that for $x \in \tilde{F}$ the line segment $\overline{x \tilde{r}_F(x)}$ stays inside $\mathbb{R}^2 \setminus \text{MA}(B)$.

Provided r_F is continuous, the map \tilde{r}_F is again homotopic to the identity map using a straight line homotopy:

Theorem III.13. *Let F be a proper closed subset of \mathbb{R}^2 not containing a half plane. If $\text{MA}(F)$ is closed, then $\text{MA}(F)$ is a strong deformation retract of \tilde{F} under the homotopy: $h : [0, 1] \times \tilde{F} \rightarrow \tilde{F}$ given by:*

$$h(t, x) = (1 - t) x + t \tilde{r}_F(x).$$

Proof. Note that for any $x \in \tilde{F}$, the line segment $h([0, 1], x)$ is contained in S , and \tilde{r}_F is continuous by Theorem III.10. \square

We discussed in the previous section how the canonical retraction map can be evaluated for an interior point of F without actually computing the medial axis explicitly. To compute \tilde{r}_F for a point x that is outside F but in the domain of \tilde{r}_F , we note that x must have a unique nearest boundary point x' , and that x , x' , and $\tilde{r}_F(x)$ must be collinear. Assuming $x \neq x'$, we can again use a nearest boundary point primitive and a bisection method to compute $\tilde{r}_F(x)$.

If the point sits exactly on the boundary ∂F , then we would have $x' = x$, and the direction in which to search for the point $\tilde{r}_F(x)$ is unknown. In this case, simply having a primitive that computes the nearest boundary point $x' = x$ is not sufficient to allow us to compute $\tilde{r}_F(x)$. We will discuss this problem further in Section F. In practical terms, we will be using polygons and retracting randomly sampled points which will have probability zero of being on the boundary.

D. Regions with C^2 boundary

In this section, we prove that the medial axis of any region in the plane having C^2 boundary is closed. Using the results of the previous section, this will show that

the medial axis of any such region F is always an SDR of \tilde{F} under the canonical retraction map. For the remainder of this section, F is a proper closed subset of the plane such that $F = \overline{F^\circ}$. (A similar result under slightly different conditions showing that the medial axis is a strong deformation retract for regions with C^2 boundary appears in [45].)

An example showing that the medial axis of a region with C^1 boundary is not necessarily a strong deformation retract is given in [37]. The example uses the same idea as Figure 5(c) which has a comb-like medial axis, but with the triangles replaced by part of the graph of the C^1 function $x^4 \sin(1/x^2)$. The medial axis again will look roughly like a comb with the “last” tooth missing.

We collect some simple facts when some boundary smoothness is assumed.

Lemma III.14. *Let a be a point in F that is not on the boundary, and $a' \in \partial F$ be a nearest boundary point to a . Suppose that the boundary of F is C^2 at a' .*

1. *The boundary of $B_F(a)$ is tangent to the boundary of F at a' .*
2. *The point a is on the normal line to the boundary of F at a' .*
3. *The curvature $\kappa(a')$ of the boundary of F at a' is always less than or equal to $1/d(a, a')$, the curvature of the circle $\partial B_F(a)$.*
4. *If $B_F(a)$ is the osculating circle to ∂F at a' , then $a \in \text{MA}(F)$.*

Proof. (1) and (2) are clear.

Taking the Frenet system at a' , the boundary ∂F may be written locally as the graph of a C^2 function $f : [-x_0, x_0] \rightarrow \mathbb{R}$. The point a' is at the origin of this system and the point a is on the y -axis by (2). If a has coordinates $(0, a_y)$, then $\rho_F(a) = d(a, a') = a_y$.

The lower half of the circle of radius R centered at $(0, R)$ is given by the graph

of the function

$$c_R(x) = R - \sqrt{R^2 - x^2}$$

where $x \in [-R, R]$.

In order for the circle c_R to stay inside F it must at least stay above f near 0, i.e., we must have $c_R(x) - f(x) \geq 0$ for x near zero. Applying Taylor's theorem to $c_R(x) - f(x)$, we see that for each $x \in [0, x_0]$, there must be an $\eta_x \in [0, x]$ such that:

$$c_R(x) - f(x) = \frac{1}{2} \left(\frac{1}{R} - f''(\eta_x) \right) x^2$$

If $f''(0)$ were strictly greater than $1/R$, then by continuity of f'' we would have $c_R(x) - f(x) < 0$ for all sufficiently small x . Therefore, we must have $1/R \geq f''(0) = \kappa(a')$, which proves (3).

(3) shows that there can be no larger circle containing a' than the osculating circle at a' , which proves (4). \square

Suppose that $a \in F$ is a simple point with nearest boundary point a' . In light of the proof of (3) above, if a is not the center of the osculating circle to the boundary at a' , i.e., if $1/d(a, a')$ is strictly less than $\kappa(a')$, then by moving a slightly farther away from a' along the y -axis, say by $\Delta a = (0, \Delta a_y)$, the disc $B_F(a + \Delta a)$ still stays inside F near a' . For a small enough movement Δa of a , we don't expect the disc $B_F(a + \Delta a)$ to meet other parts of the boundary away from a' , so the disc $B_F(a + \Delta a)$ should remain entirely inside F . Of course, $B_F(a) \subsetneq B_F(a + \Delta a)$, so this would show that a is not a medial axis point. We could then conclude that a simple point on the medial axis must be the center of an osculating circle. The next two lemmas show that this is indeed the case.

We first show that the only contact with the boundary produced by moving the center by a small amount is near the existing contact.

Lemma III.15. *Let p be any point in F . Suppose U is a (possibly disconnected) open subset of \mathbb{R}^2 containing $\partial B_F(p) \cap \partial F$. Then there is an open neighborhood $N \subset F$ of p such that for any $q \in N$, $\partial B_F(q) \cap \partial F \subset U$.*

Proof. Let $B = B_F(p)$. Because $\partial B \setminus U$ and ∂F are disjoint closed sets, there is an open neighborhood V of $\partial B \setminus U$ that does not intersect ∂F . Then $\partial B \subset U \cup V$; choose $\epsilon > 0$ such that $\epsilon < \text{dist}(\partial B, \mathbb{R}^2 \setminus (U \cup V))$. Since ∂B is compact, we have $\epsilon > 0$. Let $N = B(p, \epsilon/2)$, and let $q \in N$. We claim that $\partial B_F(q) \subset U \cup V$.

First, we have

$$\rho_F(p) - \epsilon/2 \leq \rho_F(q) \leq \rho_F(p) + \epsilon/2.$$

For any $x \in \partial B_F(q)$, we have:

$$\begin{aligned} d(x, p) &\leq d(x, q) + d(p, q) \\ &\leq \rho_F(q) + \epsilon/2 \\ &\leq \rho_F(p) + \epsilon/2 + \epsilon/2 = \rho_F(p) + \epsilon. \end{aligned}$$

and

$$\begin{aligned} d(x, p) &\geq d(x, q) - d(p, q) \\ &\geq \rho_F(q) - \epsilon/2 \\ &\geq \rho_F(p) - \epsilon/2 - \epsilon/2 = \rho_F(p) - \epsilon. \end{aligned}$$

Therefore, x is in the annulus centered at p with inner radius $\rho_F(p) - \epsilon$ and outer radius $\rho_F(p) + \epsilon$. But this annulus is contained in $U \cup V$ so we have $B_F(q)$ contained in $U \cup V$. Because there are no points of ∂F inside V , $B_F(q)$ can only intersect ∂F inside U . \square

We can now prove that any simple point on the medial axis must be the center

of an osculating circle to the boundary.

Lemma III.16. *Suppose F has C^2 boundary. If a is in $\text{MA}(F)$ and is a simple point with nearest boundary point a' , then $B_F(a)$ is the osculating circle to the boundary of F at a' .*

Proof. If $a \neq a'$, the proof of Lemma III.14 showed that it was necessary to have $1/d(a, a') \geq \kappa(a')$ in order that the circle $\partial B_F(a)$ stay inside F near a' . Note that this is also sufficient to show that $\partial B_F(a)$ is inside F near a .

Take the Frenet system at the point a' . Now suppose a is not the center of the osculating circle to the boundary at a' , i.e., $1/d(a, a') > \kappa(a')$ or $a = a'$. By moving a a small amount Δa along the y -axis, we can still maintain $1/d(a + \Delta a, a') > \kappa(a')$, from which we can conclude that in some neighborhood U of a' , $B_F(a + \Delta a)$ only meets the boundary of F at a' . But then Lemma III.15 gives a neighborhood N of a such that the only points in $\partial B_F(q) \cap \partial F$ for $q \in N$ are in U . By taking Δa small enough that $a + \Delta a$ is in N , we get that a' is a nearest boundary point of $a + \Delta a$, hence $B_F(a) \subsetneq B_F(a + \Delta a)$. \square

We already know that any multiple point is on the medial axis, so to see that the medial axis of a region with C^2 boundary is closed, we need to focus on simple points $a \in F$ that are limit points of the medial axis. There are essentially two situations:

1. If a is a limit of simple points on the medial axis, i.e., centers of osculating circles to the boundary, we can use the continuity of the curvature to show that the limit must also osculate the boundary.
2. If a is only a limit of multiple points, then it is easy to show that the nearest boundary points to these multiple points can cluster only at the unique nearest

boundary point a' to a . We will show that the boundaries of the maximal balls at these points approach the osculating circle to the boundary at a' .

It is shown in [36], using slightly different definitions, that multiple points are in fact always dense in the medial axis. Also, in [14], it is shown under the much stronger assumption of piecewise real analytic boundary that any simple point on the medial axis has contact at a point at which the boundary curvature is a local maximum.

Lemma III.17. *Suppose $\{a_j\}_{j=1}^{\infty}$ is a sequence in $\text{MA}(F)$ converging to a simple point $a \in F$ with nearest boundary point $a' \in \partial F$. If the boundary of F is C^2 at a' , then a is in $\text{MA}(F)$.*

Proof. First, note that any limit point of the set

$$\{p \in \partial F \mid p \in B_F(a_j) \text{ for some } j\}$$

must be a nearest boundary point of a , so this set must have a' as its unique limit point.

First, suppose that there is a subsequence of simple points $\{a_{j_k}\}_{k=1}^{\infty}$. For each $k = 1, 2, \dots$, let $a'_{j_k} \in \partial F$ be the unique nearest boundary point of a_{j_k} . Then we have by Lemma III.14,

$$\frac{1}{d(a_{j_k}, a'_{j_k})} = \kappa(a'_{j_k}),$$

where $\kappa(x)$, $x \in \partial F$, is the curvature of the boundary at x . By continuity of κ and $d(\cdot, \cdot)$, we have $1/d(a, a') = \kappa(a')$, which implies that $a \in \text{MA}(F)$ by Lemma III.14.

Now suppose that for j large enough, a_j is a multiple point: for large enough j , let a'_j and a''_j be distinct nearest boundary points of a_j . We have $a'_j \rightarrow a'$ and $a''_j \rightarrow a'$. By Lemma III.14, a_j is at the intersection of the normal lines to the boundary at a'_j and a''_j . By Lemma II.3, this intersection point approaches the center of the osculating

circle to the boundary at a' as a'_j and a''_j approach a' , or approaches infinity if the curvature is zero at a' . But $a_j \rightarrow a$ as $j \rightarrow \infty$, so the boundary curvature must be nonzero at a' , and a must be the center of the osculating circle to the boundary at a' , hence $a \in \text{MA}(F)$. \square

Theorem III.18. *If F has C^2 boundary, then $\text{MA}(F)$ is closed.*

Proof. If a limit point $a \in F$ of $\text{MA}(F)$ is a simple point, it is in $\text{MA}(F)$ by the previous lemma. Otherwise, $a \in \text{Vor}(F) \subseteq \text{MA}(F)$. \square

Corollary III.19. *If F has C^2 boundary, then $\text{MA}(F)$ is a strong deformation retract of \tilde{F} .*

E. Regions with piecewise C^2 boundary

A class of domains frequently encountered in practice, namely the regions with polygonal boundary, is not addressed by the results of the previous section. In this section we prove that the medial axis is closed for a more general class of domains: the domains with piecewise C^2 boundary.

For the remainder of this section, F is a proper subset of the plane, $F = \overline{F^\circ}$, and F is assumed to have piecewise C^2 boundary. Our goal is to prove that the medial axis is closed for such a set, i.e., if a is a limit point of the medial axis, then a is also on the medial axis. Again, if a is a multiple point then we already know that this is the case, so we need to focus on simple points a that are the limit of points of the medial axis.

First, let a' be the nearest boundary point to a simple point a that is the limit of points on the medial axis; Lemma III.17 tells us that if the boundary is C^2 at a' , then a is in $\text{MA}(F)$. It still remains to show that if a' is a knot point, then a is in $\text{MA}(F)$.

If a is either a limit of simple points or a limit of multiple points such that the nearest boundary points of each point in the sequence lie on the same boundary segment, then we can ignore the second boundary segment at a' and apply the arguments from the previous section without modification. An additional argument is required if this is not the case.

We establish some notation for this entire section: let $\{a_j\}_{j=1}^{\infty}$ be a sequence in $\text{MA}(F)$ converging to a simple point $a \in F$ with nearest boundary point $a' \in \partial F$. We assume that a' is a knot point of the boundary: we denote the C^2 curves parametrized by length meeting at a' by:

$$\begin{aligned}\gamma_1 &: [-\alpha, 0] \rightarrow \mathbb{R}^2 \\ \gamma_2 &: [0, +\alpha] \rightarrow \mathbb{R}^2,\end{aligned}$$

where $\gamma_1(0) = \gamma_2(0) = a'$. Note that there may be two distinct boundary curvatures at a' : the curvature of γ_1 and the curvature of γ_2 . We use the notation $\kappa_i(a')$ where $i = 1, 2$ to distinguish these.

Lemma III.20. *Fix $i = 1$ or 2 . If a is on the normal line to γ_i at a' , then $1/d(a, a') \geq \kappa_i(a')$.*

Proof. This follows by the same argument using the Taylor series given in Lemma III.14. □

Lemma III.21. *The simple point a is in $\text{MA}(F)$.*

Proof. Again, note that any limit point of the set

$$\{p \in \partial F \mid p \in B_F(a_j) \text{ for some } j\}$$

must be a nearest boundary point of a , so this set must have a' as its unique limit point.

We have the sequence $\{a_j\}_{j=1}^{\infty}$ converging to a . If there is a subsequence of simple points, there must be a further (infinite) subsequence of simple points that are all on γ_i where i is 1 or 2. But then a is the center of the osculating circle to γ_i at a' , so by the above lemma, a is in $\text{MA}(F)$.

Similarly, if there is a subsequence of multiple points whose nearest boundary points are on a single boundary segment γ_i , a is again the center of the osculating circle to γ_i at a' , so we again have a in $\text{MA}(F)$.

Finally, suppose for j large enough that a_j is a multiple point with nearest boundary points a'_j on γ_1 and a''_j on γ_2 . Let $N_1 : [-\alpha, 0] \rightarrow \mathbb{R}^2$ and $N_2 : [0, \alpha] \rightarrow \mathbb{R}^2$ be the inward unit normal vector fields along γ_1 and γ_2 . For all j large enough we have $a'_j = \gamma_1(t'_j)$ for some $t'_j \in [-\alpha, 0]$ and $a''_j = \gamma_2(t''_j)$ for some $t''_j \in [0, \alpha]$. The intersection of the normal line $a'_j + uN_1(t'_j)$, $u \in \mathbb{R}$ at a'_j and the normal line $a''_j + vN_2(t''_j)$, $v \in \mathbb{R}$ at a''_j is the point a_j . For each j we may apply Cramer's rule to find the value u_j of u at the intersection of these lines:

$$u_j = \frac{\det \begin{pmatrix} a''_j - a'_j & N_2(t''_j) \end{pmatrix}}{\det \begin{pmatrix} -N_1(t'_j) & N_2(t''_j) \end{pmatrix}}$$

If γ_1 and γ_2 do not have a common tangent direction at a' , i.e., if $N_1(0) \neq \pm N_2(0)$, then the denominator of this expression approaches a nonzero constant. However, the numerator approaches zero because a'_j and a''_j both approach a' ; this implies that the parameter u_j approaches zero as $j \rightarrow \infty$, which in turn implies that a_j approaches a' , i.e., we have $a = a'$.

Suppose now that γ_1 and γ_2 do have a common tangent direction at a' , i.e., we have $N_1(0) = N_2(0)$. Then γ_1 and γ_2 have a common Frenet frame at a' . Near a' we can write γ_1 as the graph of a C^2 function $f_1 : [-\delta, 0] \rightarrow \mathbb{R}$ in the Frenet frame,

and we can write γ_2 as the graph of a C^2 function $f_2 : [0, \delta] \rightarrow \mathbb{R}$. The positive y -axis of the Frenet frame is the common normal direction of the two curves so by Lemma II.3 as $j \rightarrow \infty$, the intersection of the normal line to γ_1 at a'_j with the y -axis converges to the center $(0, 1/\kappa_1(a'))$ of the osculating circle to γ_1 at a' (or to $(0, \infty)$ if the curvature of γ_1 is zero at a'). Similarly, the intersection of the normal line to γ_2 at a''_j with the y -axis converges to the center $(0, 1/\kappa_2(a'))$ of the osculating circle to γ_2 at a' (or to $(0, \infty)$ if the curvature of γ_2 is zero at a'). Because a_j is the intersection of these normal lines, the y -coordinates a_j must converge to at least $y_0 = \min(1/\kappa_1(a'), 1/\kappa_2(a'))$. By Lemma III.20, this y -coordinates can be no greater than y_0 , so we must have $a = (0, y_0)$, hence $B_F(a)$ osculates one of the boundary segments at a' , so $a \in \text{MA}(F)$. \square

Theorem III.22. *If F has piecewise C^2 boundary, then $\text{MA}(F)$ is closed.*

Proof. If a limit point $a \in F$ of $\text{MA}(F)$ is a simple point, it is in $\text{MA}(F)$ by the previous lemma. Otherwise, $a \in \text{Vor}(F) \subseteq \text{MA}(F)$. \square

Corollary III.23. *If F has piecewise C^2 boundary, then $\text{MA}(F)$ is a strong deformation retract of \tilde{F} under the canonical retraction map.*

F. MAPRM in the plane

Returning to the realm of motion planning for a point moving in the plane amid polygonal obstacles, we now know that the medial axis of the free space will be strong deformation retract, under the extended retraction map, of \tilde{F} , nearly the whole configuration space.

It is reasonably clear from the definition of the retraction maps that global knowledge of the medial axis is not required to compute pointwise images under these maps. Points in the free space simply retract away from their nearest boundary point until

they reach the medial axis, and points outside the free space move toward their nearest boundary point and on through into the free space where the canonical retraction map is applied. It is possible to tell when a point has moved past the medial axis because its nearest boundary point will change as soon as this happens: this allows us to actually compute the retraction in practice using just a nearest boundary point primitive and a bisection method.

The MAPRM algorithm for construction of a roadmap consists of uniform sampling in the full configuration space, followed by application of the extended retraction map. The nodes are then connected to form a roadmap using a local planner as in the standard PRM methods.

The MAPRM algorithm in the plane is given in Algorithm III.1.

Algorithm III.1 MAPRM in 2D

Preprocessing:

Input. N , the number of nodes to generate.

Output. N nodes in F connected into a roadmap graph.

- 1: **repeat**
 - 2: Generate a uniformly random point p in \mathbb{R}^2 .
 - 3: Find the nearest point q on ∂F to p .
 - 4: **if** p is free **then**
 - 5: Take the retraction direction \vec{v} to be \overrightarrow{qp} , and let the start point s be p .
 - 6: **else**
 - 7: Take the retraction direction \vec{v} to be \overrightarrow{pq} , and let the start point s be q .
 - 8: **end if**
 - 9: Using bisection, move s in the direction \vec{v} until q is not the unique nearest point of ∂F to s . This moves s onto the medial axis of F
 - 10: **until** N nodes have been generated
 - 11: For each pair of nodes: if the pair can be connected with a straight line, insert an edge into the graph connecting them.
-

1. Rudimentary analysis

It is fairly clear that applying this sample-and-retract approach will increase the sampling rate in small corridors: a nonzero volume of nearby configurations in $B = \overline{\mathbb{R}^2 \setminus F}$ will be pushed into the corridor, increasing the sampling by a constant that is essentially independent of the volume of the corridor. In this section we make this idea somewhat more precise.

The medial axis provides a convenient definition for what is meant by a corridor. Let $r_F : F^\circ \rightarrow \text{MA}(F)$ be the canonical retraction map. A *corridor* in F is a connected open subset S of F° that contains the retracted images of all of its points (i.e., $r_F(S) \subseteq S$) while also including any points that will be retracted into S under r_F (i.e., $r_F^{-1}(S) \subseteq S$). These conditions essentially ensure that S is bounded on “all” sides by obstacles.

Clearly any point of such a corridor S remains in S under the extended retraction map. Furthermore, any point in B whose nearest boundary point (on ∂B) is also in the closure \overline{S} will be mapped into \overline{S} . Let $b_B : B \setminus \text{MA}(B) \rightarrow \partial B$ be the map that takes each point to its nearest boundary point. (This is well-defined since any multiple point will be in $\text{MA}(B)$.) The volume of points that map into the \overline{S} under the full retraction map is:

$$\mu(S) + \mu(b_B^{-1}(\partial S)),$$

where μ is a volume measure² on C .

Using uniform random sampling from C (with respect to μ) with collisions dis-

²A volume measure is provided, for example, by the Riemannian metric on $\text{SE}(3)$. We assume F , B , and S , etc., are μ -measurable.

carded, the probability that a single sampled configuration is contained in S is:

$$\frac{\mu(S)}{\mu(C)}. \quad (3.1)$$

Assuming the medial axis $\text{MA}(B)$ has measure zero, the probability that a single random sample from C is retracted by MAPRM into \bar{S} is:

$$\frac{\mu(S) + \mu(b_B^{-1}(\partial S))}{\mu(C)} \quad (3.2)$$

Assuming the term $\mu(b_B^{-1}(\partial S))$ is nonzero, as it is at least in our simple examples, (3.2) will exceed (3.1). See Figure 11. Note that the term $\mu(b_B^{-1}(\partial S))$ only involves the boundary of the corridor S and does not involve the volume of S .

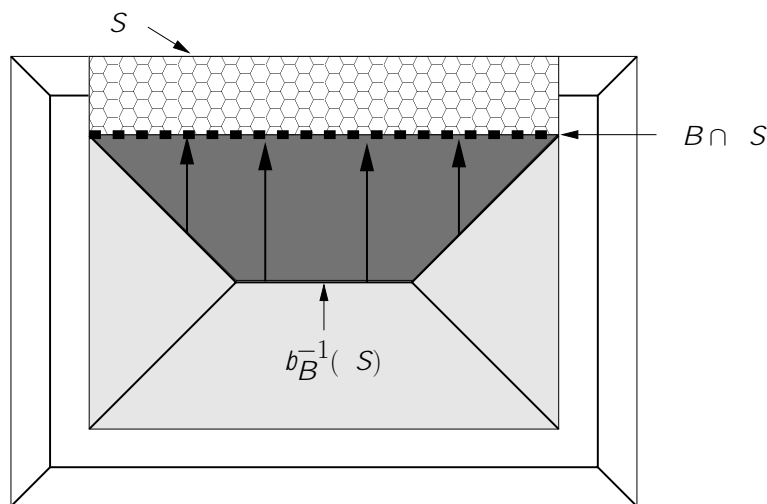


Fig. 11. The area added by the extended retraction map. The hatched area is S , the heavy dashed line is $\partial B \cap \partial S$, and the dark shaded area is $b_B^{-1}(\partial S)$.

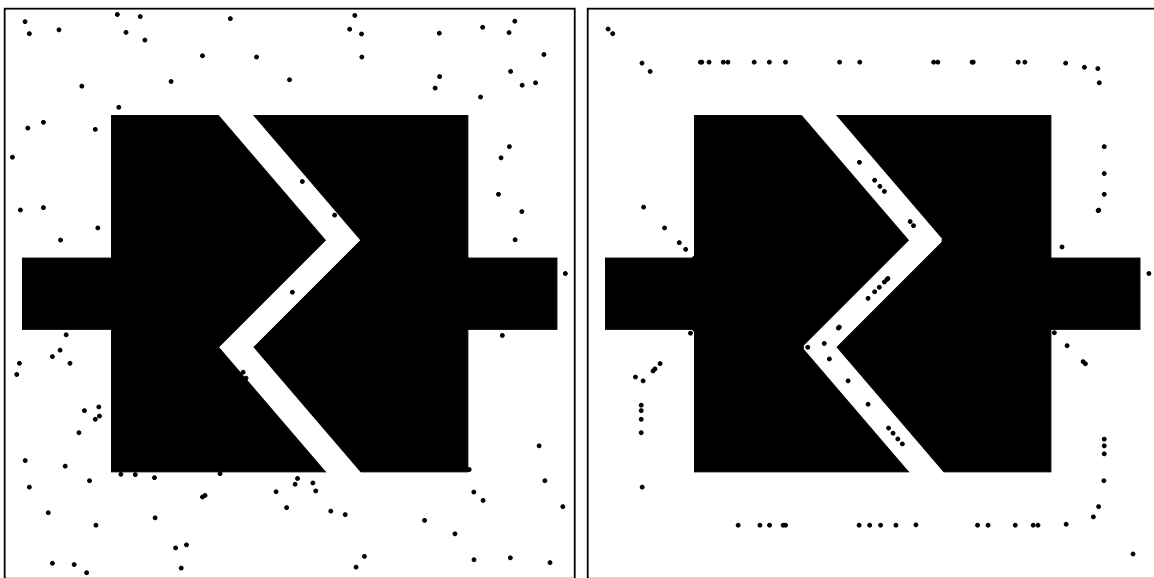
Intuitively, we may regard the function $A \mapsto \mu(b_B^{-1}(A))$ defined on subsets A of the boundary of B as a measure of how “thick” the obstacle is behind A . Equa-

tion (3.2) shows that our algorithm gives a greater increase in the sampling rate for corridors that happen to be bounded by thicker obstacles.

2. Examples

We give two examples: one in which MAPRM shows significant advantage over uniform sampling, and one in which it does not. We omit the connections between nodes and show only the results of the sampling.

Figure 12 shows an example of a free space containing a narrow corridor. Part (a)



(a)

(b)

Fig. 12. Uniform sampling (a) vs. MAPRM sampling (b).

shows the result of sampling 100 nodes from the free space: this required generating 168 random configurations. Part (b) shows the result of sampling 100 nodes using the MAPRM Algorithm III.1: 100 points were generated in the square and the extended

retraction map was applied to them. MAPRM produces many nodes in the corridor because the obstacles forming the corridor are “thick” which gives a reasonably large value for the $\mu(b_B^{-1}(\partial S))$ in Equation (3.2). Because of this, pushing the two obstacles closer together would not greatly affect the number of nodes MAPRM generates in the corridor.

Figure 13 shows an example in which MAPRM is less helpful. As before, nodes

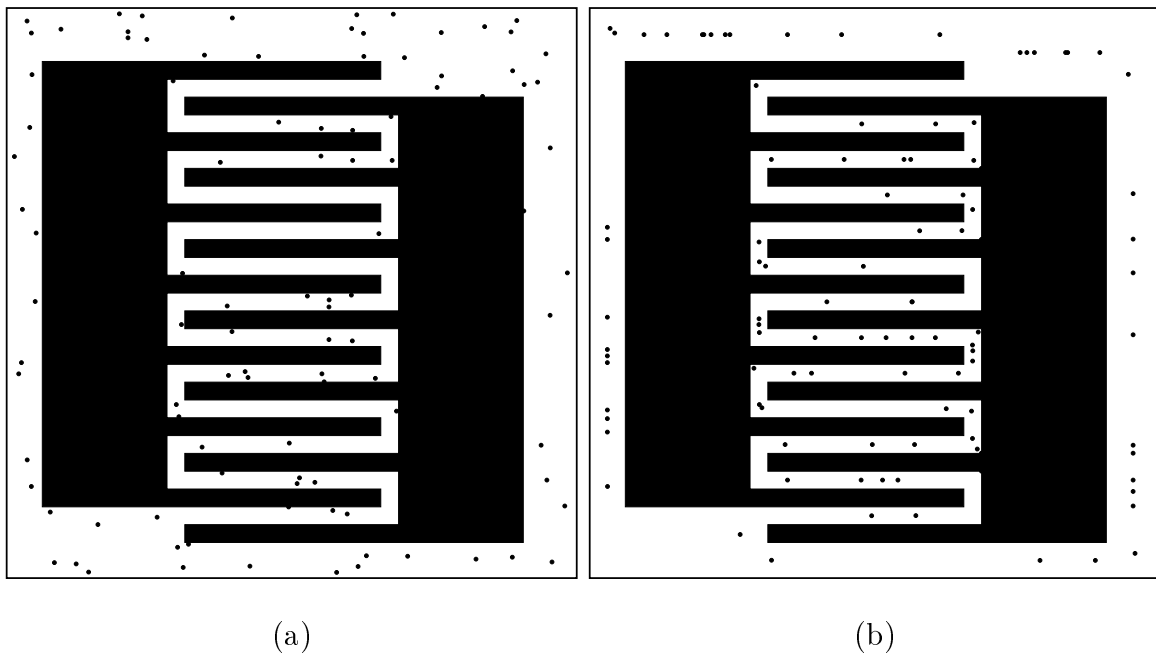


Fig. 13. Another example: uniform sampling (a) vs. MAPRM sampling (b).

in (a) are 100 free configurations (203 samples required), and nodes in (b) are 100 samples from the square retracted to $MA(F)$. The medial axis of the obstacles forming the corridor has a “maze-like” quality. The medial axis is very near the surface causing the $\mu(b_B^{-1}(\partial S))$ in Equation (3.2). to be smaller in this case.

The running time of both algorithms on these planar examples is insignificant. However, as we will see below in the three dimensional rigid body case, as the volume of the free space gets smaller, MAPRM actually takes significantly *less* time to generate a free node than uniform sampling, decreasing the overall running time.

CHAPTER IV

THE MEDIAL AXIS ON $SE(3)$

As explained in Chapter II, the problem of motion planning for a rigid body in three dimensions can be viewed as planning motions of a point or configuration in the free space F within the configuration space $SE(3)$. A natural way to try to generalize the sample-and-retract idea of the previous chapter is to now retract sampled configurations onto the medial axis of the free space along “lines” in the configuration space, now a 6-D manifold rather than a 2-D Euclidean space. Accomplishing this involves several steps.

First, we need to generalize the definition of the medial axis to make it applicable to $SE(3)$. This involves addressing the question of choice of metric on $SE(3)$. Our development uses a complete Riemannian metric structure on $SE(3)$, which carries with it a natural distance metric; however, as discussed in Chapter II, there are many possible choices of Riemannian metric which potentially yield different distance metrics. We will address this issue of choice of metric by selecting a class of metrics which are intuitively appealing from the standpoint of motion planning while enabling the resulting retraction maps to be evaluated computationally.

With the medial axis definition in hand, we will show that the retraction maps can be defined essentially as in the 2-D case. Just as in the 2-D case, it is not true in general that the retraction maps will be continuous; however, we will prove in Chapter V that in the case of a polyhedral workpiece moving among polyhedral obstacles the retraction maps are continuous. Again, using straight line (geodesic) homotopies, this allows us to show that the medial axis of the free space is a strong deformation retract of the free space itself. Again, the importance is that there is no danger in taking sampled nodes just on the medial axis: the medial axis captures the topology

of the free space completely.

Finally, it will be clear from our development of the retraction maps how to actually carry out the retraction in practice. Due to our assumptions on the Riemannian metric the retraction involves only translational motion with no rotation, at least for all free configurations and a large subset of the collision configurations. In Chapter V we will present the MAPRM algorithm which carries out the retraction in practice and give experimental results demonstrating its performance.

A. The definition of the medial axis

In generalizing the definition of the medial axis, there is the question of how weak a mathematical structure can be used. For example, is it possible to generalize to arbitrary metric spaces? It turns out that although the medial axis definition often makes sense on metric spaces, the essential property that we seek, namely that the medial axis be an SDR for a large class of subsets of the metric space, is only true under very special circumstances. First of all, to even define the retraction maps, we need some notion of shortest path or line, and we need unique maximal containing balls to exist for a large portion of the balls contained inside the set: this requires at least some rudimentary geometric structure on the space in question.

As discussed in Chapter II, Riemannian geometry provides a standard machinery for imposing geometric structure on smooth manifolds such as $SE(3)$. Once a Riemannian metric has been chosen, the geometric concepts of angle, distance, shortest path, etc., are all well defined. However, even the imposition of a Riemannian metric is insufficient for our purposes: we will also require that the manifold be complete, i.e., that closed and bounded subsets of the manifold are compact.

To see that completeness is needed, consider the case of the punctured plane

$\mathbb{R}^2 \setminus (0, 0)$ with its usual geometry. This is a Riemannian manifold but is not complete. The medial axis of the unit disc in this manifold is empty, so there is no hope that the medial axis will be an SDR. In fact, any connected region whose medial axis “should” pass through the origin will have a hole at that point causing the topology to differ from that of the original region. This motivates us to restrict our discussion to the case of complete Riemannian manifolds.

Let M be any complete Riemannian manifold. For any proper subset S of M , we define the *boundary distance* function $\rho_S : S \rightarrow \mathbb{R}$ by

$$\rho_S(x) = \text{dist}(x, \partial S).$$

Let F be a proper closed subset of M . Note that for $x \in F$, the largest closed ball centered at x that is still contained in F is $\overline{B}(x, \rho_F(x))$. We will still use the notation $B_F(x) = \overline{B}(x, \rho_F(x))$, and we refer to this as the *maximal ball in F centered at x* . The *medial axis* $\text{MA}(F)$ of F in M is defined as before: a point x is on the medial axis if there is no point y in F whose associated maximal ball properly contains the maximal ball centered at x ; i.e.,

$$\text{MA}(F) = \{x \in F \mid \nexists y \in F \text{ with } B_F(x) \subsetneq B_F(y)\}.$$

B. The choice of Riemannian metric

As explained in Chapter II, there is no single obvious choice of Riemannian metric to use for $\text{SE}(3)$. We are thus faced with the problem of making (and defending) a choice of metric on $\text{SE}(3)$. Rather than select a single metric, we address this issue by selecting a class of metrics which are intuitively appealing from the standpoint of motion planning and enable the resulting retraction maps to be evaluated computationally. Under such a metric, it will turn out that the medial axis and retraction maps have

a simple interpretation in terms of geometry and motions in the workspace. We will describe the class of metrics under consideration with a pair of assumptions together with an example to show that the class is non-empty.

Interpreting $\text{SE}(3)$ as the configuration space of a free-flying rigid body, the factor \mathbb{R}^3 in the semi-direct product $\text{SE}(3) = \text{SO}(3) \ltimes \mathbb{R}^3$ corresponds to pure translation, i.e., at fixed orientation of the workpiece. In other words any element (I, r) where $I \in \text{SO}(3)$ is the identity matrix and $r \in \mathbb{R}^3$, is just a displacement by r . Note that \mathbb{R}^3 has a metric that is bi-invariant under the action of $\text{SE}(3)$, i.e., under rigid transformation. This is just the usual Euclidean metric, which is unique up to scale. The geodesics under this metric are the straight lines in \mathbb{R}^3 ; thinking of \mathbb{R}^3 as a subgroup of $\text{SE}(3)$, this corresponds to translations of the workpiece along a straight line. Our first assumption on the metric is that these pure translations remain minimizing geodesics with the choice of metric on $\text{SE}(3)$. For convenience, we also assume that the length of a translation by a vector r is just $|r|$, the Euclidean length of r .

Assumption 1. *For $r_1, r_2 \in \mathbb{R}^3$, and $Q \in \text{SO}(3)$, the path $\gamma(t) = (Q, (1-t)r_1 + t r_2)$, $t \in [0, 1]$, from (Q, r_1) to (Q, r_2) is a minimizing geodesic of length $|r_2 - r_1|$.*

We note here that under any Riemannian metric satisfying Assumption 1, $\text{SE}(3)$ will be complete. To see this we can use part of the Hopf-Rinow Theorem which asserts that completeness of a Riemannian manifold M is equivalent to the closed and bounded subsets of M being compact. Under any metric satisfying just Assumption 1, any closed and bounded subset S of $\text{SE}(3)$ is compact: its projection $\pi_2(S) \subset \mathbb{R}^3$ onto the \mathbb{R}^3 factor is compact by the Heine-Borel Theorem, so $\text{SO}(3) \times \pi_2(S)$ is compact and contains the closed subset S , hence S is compact.

When trying to define a “distance” between two configurations with different orientations and translations, there is the question of how to weight rotations against

translations. Assumption 1 says that a pure translation of length d in the workspace is a path of length d in the configuration space. Intuitively, we would expect a purely rotational motion in which most points on the workpiece move a distance comparable to d to be a path of length comparable to d in the configuration space. This idea depends on the workpiece in question: given that the cost of a translation is independent of the workpiece, we would expect rotations to cost more for a larger workpiece. We would like to make the assumption that, for the metric under consideration, not only are rotations “comparable” to translations, but given the choice between making a rotation or a translation we prefer the translation. The particular assumption we will make says essentially that given a particular free configuration of the workpiece, the shortest path to a contact configuration is to translate the workpiece along a straight line directly toward the nearest obstacle until the workpiece makes contact.

To make this more precise, suppose that we are given the workpiece U in a particular configuration $g_1 \in \text{SE}(3)$, and let $p \in g_1 \cdot U$ be the coordinates of a particular point on U when U is in configuration g_1 . Given a desired destination $p' \in \mathbb{R}^3$ for p , there are many ways to move the workpiece to get p to its destination p' : we could simply translate or we could translate and rotate simultaneously in varying amounts. What we want to assume about the metric is that rotations are weighted heavily enough that the shortest motion of the workpiece to take p to p' is always the pure translation. From the standpoint of motion planning, this assumption will tell us that, given free configuration $g \in F \subseteq \text{SE}(3)$, where F is the free space of a given workpiece U and obstacle V , the shortest path to the boundary of the free space is a translation of the workpiece straight at the nearest obstacle point (i.e., translate the nearest point $p \in g \cdot U$ to its nearest point $q \in V$). Wording this assumption more formally, we have:

Assumption 2. Let $(Q, v) \in \text{SE}(3)$. Given a workpiece U , a compact subset of \mathbb{R}^3 , let $p \in U$, and $p' \in \mathbb{R}^3$. Let $S = \{h \in \text{SE}(3) \mid h \cdot p = p'\}$ be the set of all transformations mapping p to p' . We assume that any closest point in S to (Q, v) is of the form (Q, u) .

We need to verify that there is at least one metric satisfying these assumptions: an appropriately weighted product metric on $\text{SE}(3)$ will have the desired properties. We can see this as follows:

We are given closed sets $U \subset \mathbb{R}^3$ and $V \subseteq \mathbb{R}^3$, where U is assumed to be compact. Let R_U be any number large enough that U is contained¹ in the open ball $B((0,0), R_U)$. The Lie group $\text{SO}(3)$ has a bi-invariant metric, unique up to scale, which may be normalized so that a rotation of θ radians about a fixed axis will have length θ . (See [34].)

Let $\pi_1 : \text{SE}(3) \rightarrow \text{SO}(3)$ and $\pi_2 : \text{SE}(3) \rightarrow \mathbb{R}^3$ be the canonical projection maps, and let π_{1*} and π_{2*} be the induced differential maps (see [16]). Let $g \in \text{SE}(3)$; for tangent vectors $v, w \in T_g \text{SE}(3)$, define:

$$\langle v, w \rangle_g = R_U^2 \langle \pi_{1*}v, \pi_{1*}w \rangle_{\pi_1(g)}^{\text{SO}(3)} + \langle \pi_{2*}v, \pi_{2*}w \rangle_{\pi_2(g)}^{\mathbb{R}^3} \quad (4.1)$$

This is essentially the product metric on $\text{SE}(3)$, the sum of the pullbacks of the metrics on $\text{SO}(3)$ and \mathbb{R}^3 , but with the rotational part weighted by R_U , the radius of the bounding sphere. This metric is discussed in [34].

Incidentally, this metric is left invariant on $\text{SE}(3)$. This is easy to see because the left multiplication map on $\text{SE}(3)$, $L_{(Q,r)}((Q', r'))$, evaluates to $(L_Q(Q'), L_{(Q,r)}(r'))$, so the bi-invariance of the metric on $\text{SO}(3)$ and the invariance of the metric on \mathbb{R}^3

¹In practice, we usually position U so that its smallest enclosing sphere has center at the origin and we take R_U to be the slightly larger than the radius of the enclosing sphere.

under the action of $\text{SE}(3)$ give left invariance.

With this metric, the length of a C^1 path $\gamma = (\gamma_1, \gamma_2) : [a, b] \rightarrow \text{SE}(3)$ is given by:

$$\begin{aligned} \text{length}(\gamma) &= \int_a^b |\gamma'(t)| \, dt \\ &= \int_a^b \sqrt{R_U^2 |\gamma_1'(t)|^2 + |\gamma_2'(t)|^2} \, dt \end{aligned}$$

It is easy to show that under such a product metric, a path on the product is a minimizing geodesic if and only if it is a minimizing geodesic in each component. Again, the distance between two points given by a Riemannian metric is the length of a minimizing geodesic connecting those points.

Proposition IV.1. *The weighted product metric in equation 4.1 satisfies Assumptions 1 and 2.*

Proof. It is easy to see that Assumption 1 is satisfied.

To verify Assumption 2, fix $(Q, v) \in \text{SE}(3)$, $p \in U$, $p' \in \mathbb{R}^3$, and let

$$S = \{h \in \text{SE}(3) \mid h \cdot p = p'\}.$$

We need to show that if (R, w) is a closest configuration in S to (Q, v) , then $R = Q$.

Let $\gamma = (\gamma_1, \gamma_2) : [a, b] \rightarrow \text{SE}(3)$ be a minimizing geodesic from $(Q, v) = \gamma(a)$ to $(R, w) = \gamma(b)$. Seeking a contradiction, assume now that $R \neq Q$. Let

$$p_a = \gamma(a) \cdot p$$

and

$$p_b = \gamma(b) \cdot p = (R, w) \cdot p = p'.$$

First, note that the distance $d(p_a, p_b)$ in \mathbb{R}^3 between p_a and p_b cannot exceed the total length l of the path traced out by $\gamma(t) \cdot p$ as t runs from a to b . But we will show below that the length l is strictly less than the length of γ in $\text{SE}(3)$ using our Riemannian metric. Since the length in $\text{SE}(3)$ of the pure translation geodesic $(Q, v + s(p_b - p_a))$, where $s \in [0, 1]$, is just $d(p_a, p_b)$, this will give a path to $(Q, v + (p_b - p_a)) \in S$ that is shorter than γ , a contradiction.

We have

$$\begin{aligned} l &= \int_a^b \left| \frac{d}{dt} (\gamma(t) \cdot a) \right| dt \\ &= \int_a^b |\gamma'_1(t) \cdot a + \gamma'_2(t)| dt \end{aligned}$$

Define r to be the length of γ measured by the Riemannian metric, i.e.,

$$\begin{aligned} r &= \int_a^b |\gamma'(t)| dt \\ &= \int_a^b \sqrt{R_U^2 |\gamma'_1(t)|^2 + |\gamma'_2(t)|^2} dt \end{aligned}$$

(where R_U is the quantity used to define the Riemannian metric). To complete the proof, we need to show that $l < r$. Note that $\gamma'_2(t)$ is never zero for $t \in [a, b]$. For any $t \in [a, b]$, we have:

$$\begin{aligned} |\gamma'_2(t) \cdot p + \gamma'_1(t)|^2 &\leq |\gamma'_2(t) \cdot p|^2 + |\gamma'_1(t)|^2 \\ &\leq |\gamma'_2(t)|^2 |p|^2 + |\gamma'_1(t)|^2 \\ &< |\gamma'_2(t)|^2 R_U^2 + |\gamma'_1(t)|^2 \end{aligned}$$

which shows that the integrand in the expression for l is always less than the integrand

in the expression for r , proving that $r < l$. One inequality used above, namely,

$$|\gamma_2'(t) \cdot p|^2 < |\gamma_2'(t)|^2 |p|^2$$

requires some justification.

Because γ_2 is a geodesic on $\text{SO}(3)$, it represents a constant-speed rotation in \mathbb{R}^3 through some angle θ about an axis through the origin. This means that $|\gamma_2'(t) \cdot p|$, the speed of the point p under this rotation, is just the speed of the rotation, $\theta/(b-a)$ times the distance of p from the origin. On the other hand, we have assumed that the metric $\text{SO}(3)$ normalized so that the length of the entire rotation as measured in $\text{SO}(3)$ is just θ ; this means that $|\gamma_2'(t)| = \theta/(b-a)$ which gives the required inequality. \square

For the remainder of this chapter, we assume $\text{SE}(3)$ with a Riemannian metric satisfying Assumptions 1 and 2. When using a distance metric $d(\cdot, \cdot)$, we always mean the distance metric induced by the Riemannian metric.

C. The medial axis in \mathbb{R}^3

We will need a few facts about the medial axis in \mathbb{R}^3 . The definition and many of the results from the 2-D case carry over to \mathbb{R}^3 or to any Euclidean space with no modification, so we will just state the required results.

Let F be a proper closed subset of \mathbb{R}^3 . As in the \mathbb{R}^2 case, we have the *boundary distance* function $\rho_F : F \rightarrow \mathbb{R}$ given by

$$\rho_F(x) = \text{dist}(x, \partial F),$$

the *maximal ball in F centered at $x \in F$*

$$B_F(x) = \overline{B}(x, \rho_F(x)),$$

and the *medial axis* $\text{MA}(F)$ of F in \mathbb{R}^3

$$\text{MA}(F) = \{x \in F \mid \nexists y \in F \text{ with } B_F(x) \subsetneq B_F(y)\}.$$

We also have again the definition of multiple and simple points and $\text{Vor}(F)$. Again, we define the domains of definition of the retraction maps:

$$F' = \left(F \setminus \overline{\text{MA}(\mathbb{R}^2 \setminus F)}\right) \cup \text{MA}(F),$$

and

$$\tilde{F} = \text{MA}(F) \cup (\mathbb{R}^2 \setminus \overline{\text{MA}(\mathbb{R}^2 \setminus F)}).$$

Proposition IV.2. *Let F be a proper closed subset of \mathbb{R}^3 .*

1. *Multiple points of F are contained in $\text{MA}(F)$, i.e., $\text{Vor}(F) \subseteq \text{MA}(F)$.*
2. *Let $x, y \in F$. Then $B_F(x) \subseteq B_F(y)$ if and only if $\rho_F(y) = \rho_F(x) + d(x, y)$.*
3. *If F does not contain any half space, then for each point $x \in F'$, there is a unique point $a_x \in \text{MA}(F)$ such that $B_F(x) \subseteq B_F(a_x)$.*

Proof. The proofs are the same as in the 2-D case.

1. Analog of Proposition III.4.
2. Analog of Lemma III.5
3. Analogs of Propositions III.6, III.7, and III.8.

□

Again, we have the canonical and extended retraction maps:

$$r_F : F' \mapsto \text{MA}(F),$$

and

$$\tilde{r}_F : \tilde{F} \mapsto \text{MA}(F).$$

D. The retraction maps

Recall that given a compact subset U of \mathbb{R}^3 and a closed obstacle set V in \mathbb{R}^3 , the free space F is defined to be the closure of the set

$$\{g \in \text{SE}(3) \mid g \cdot U \cap V = \emptyset\}.$$

A direct consequence of Assumptions 1 and 2 is that the boundary points of a maximal ball in F will always meet the boundary of F at configurations that have the same orientation component as the center of the ball: this is Lemma IV.5 below.

We will show that this means that the medial axis of F can be computed by looking at the medial axes of fixed orientation “slices” of F . These results are not true for $\text{SE}(3)$ with a general Riemannian metric or if F is not the free space associated with some workpiece and obstacle. We emphasize again that we are given specific U and V , and that *we are assuming a Riemannian metric satisfying Assumptions 1 and 2 above*.

Recall that the free space $F \subseteq \text{SE}(3)$ is the closure of the set:

$$\{(Q, v) \in \text{SE}(3) \mid (Q \cdot U + v) \cap V = \emptyset\}.$$

For $Q \in \text{SO}(3)$, we define the injection map $i_Q : \mathbb{R}^3 \rightarrow \text{SE}(3)$ by $i_Q(v) = (Q, v)$, and we again have the projection map $\pi_2 : \text{SE}(3) \rightarrow \mathbb{R}^3$ by $\pi_2((Q, v)) = v$. Define

$$F_Q = \pi_{\mathbb{R}^3}(F \cap i_Q(\mathbb{R}^3)).$$

This means that $F_Q \subseteq \mathbb{R}^3$ consists of all $v \in \mathbb{R}^3$ such that $(Q, v) \in F$, i.e., F_Q is the fixed-orientation “slice” of F with orientation component Q . Then we have:

$$F = \bigcup_{Q \in \text{SO}(3)} i_Q(F_Q)$$

Because F_Q is a closed subset of \mathbb{R}^3 we have as usual the boundary distance function ρ_{F_Q} , and the maximal ball centered at $v \in F_Q$:

$$B_{F_Q}(v) = \overline{B}(v, \rho_{F_Q}(v)).$$

Note that Assumption 1 allows us to move freely back and forth between measuring distances on F and on F_Q when this makes sense. For example, if (Q, v) and (Q, w) are elements of F , then $d((Q, v), (Q, w)) = d(v, w)$, where $d((Q, v), (Q, w))$ is the distance in $\text{SE}(3)$ induced by the Riemannian metric and $d(v, w)$ is the Euclidean distance in \mathbb{R}^3 .

We would first like to show that with our choice of metric, the medial axis of the free space is actually just the union of the medial axes in \mathbb{R}^3 of the fixed orientation free spaces F_Q for all $Q \in \text{SO}(3)$. Again, this is *not* true for general metrics on $\text{SE}(3)$ and even with our assumptions on the metric, it is not true for arbitrary sets in $\text{SE}(3)$.

Theorem IV.3. *With F and F_Q as above,*

$$\text{MA}(F) = \bigcup_{Q \in \text{SO}(3)} i_Q(\text{MA}(F_Q)),$$

where $\text{MA}(F)$ is the medial axis of F in $\text{SE}(3)$ and $\text{MA}(F_Q)$ is the medial axis of F_Q in \mathbb{R}^3 .

We prove this as a series of lemmas. Recall that for $x \in X$, where X is a metric space, and $r > 0$, $B(x, r)$ denotes the open ball centered at x of radius r .

Lemma IV.4. *Let $u, v \in \mathbb{R}^3$ and $r, s \in \mathbb{R}$, and suppose that the balls $B(u, r)$ and $B(v, s)$ in \mathbb{R}^3 satisfy:*

$$B(u, r) \subseteq B(v, s).$$

Then for any $Q \in \text{SO}(3)$, the balls $B((Q, u), r)$ and $B((Q, v), s)$ in $\text{SE}(3)$ satisfy

$$B((Q, u), r) \subseteq B((Q, v), s).$$

Proof. In order to have $B(u, r) \subseteq B(v, s)$, we must have $d(u, v) \leq s - r$. For any $g \in B((Q, u), r)$, we have $d(g, (Q, u)) \leq r$, so:

$$\begin{aligned} d(g, (Q, v)) &\leq d(g, (Q, u)) + d((Q, u), (Q, v)) \\ &\leq r + (s - r) = s, \end{aligned}$$

by Assumption 1; i.e., g is in $B((Q, v), s)$. \square

This next lemma says that the maximal ball centered at a configuration in F can meet the boundary of the free space only at configurations that have the same orientation component. This is a direct result of both Assumption 2 and the fact that F is a free space. The lemma is not true if we substitute for F , say, the complement of F .

Lemma IV.5. *For any $(Q, v) \in F$, we have:*

$$B_F((Q, v)) \cap \partial F \subseteq i_Q(\mathbb{R}^3).$$

Proof. Let $(P, u) \in B_F((Q, v)) \cap \partial F$. Take $p' \in (P, u) \cdot U \cap V$, and let $p = (P, u)^{-1} \cdot U$.

As in Assumption 2, we look at the set

$$S = \{h \in \text{SE}(3) \mid h \cdot p = p'\}.$$

Any $h \in S$ must have $d(h, (Q, v)) \geq \rho_F((Q, v))$, and we have $(P, u) \cdot p = p'$. Also, $d((P, u), (Q, v)) = \rho_F((Q, v))$, so (P, u) is a closest point (configuration) to (Q, v) .

But by Assumption 2, this means that $P = Q$, hence $(P, u) \in i_Q(\mathbb{R}^3)$. \square

As an immediate consequence, we have:

Lemma IV.6. For $(Q, v) \in F$, $\rho_F((Q, v)) = \rho_{F_Q}(v)$.

Lemma IV.7. Let $Q \in \text{SO}(3)$ and u and v be elements of F_Q such that $B_F((Q, u)) \subseteq B_F((Q, v))$. Then $B_{F_Q}(u) \subseteq B_{F_Q}(v)$.

Proof. We have $B_F((Q, u)) \subseteq B_F((Q, v))$, so that $B(u, \rho_F((Q, u))) \subseteq B(v, \rho_F(Q, v))$. But $\rho_F((Q, u)) = \rho_{F_Q}(u)$ and similarly for v , so we have $B_{F_Q}(u) \subseteq B_{F_Q}(v)$. \square

Lemma IV.8. Let (Q, v) and (R, w) be in F . If $B_F((Q, v)) \subseteq B_F((R, w))$, then $Q = R$.

Proof. By Lemma IV.4, any point of $B_F((Q, v)) \cap \partial F$ must be in $i_Q(\mathbb{R}^3)$ and any point of $B_F((R, w)) \cap \partial F$ must be in $i_R(\mathbb{R}^3)$. But we have $B_F((Q, v)) \subseteq B_F((R, w))$ and $i_Q(\mathbb{R}^3)$ and $i_R(\mathbb{R}^3)$ are disjoint unless $R = Q$, so we must have $R = Q$. \square

Proof of Theorem IV.3. First, let $(Q, v) \in \text{MA}(F)$. We need to show that $v \in \text{MA}(F_Q)$. Suppose there is a $w \in F_Q$ such that $B_{F_Q}(v) \subseteq B_{F_Q}(w)$. Then by Lemma IV.4, $B((Q, v), \rho_{F_Q}(v)) \subseteq B((Q, w), \rho_{F_Q}(w))$. But by Lemma IV.6, $\rho_{F_Q}(x) = \rho_F((Q, x))$ for all $x \in F_Q$, so this means we have:

$$B_F((Q, v)) \subseteq B_F((Q, w)).$$

Because $(Q, v) \in \text{MA}(F)$, this means that $w = v$, which implies that $v \in \text{MA}(F_Q)$.

Now let $Q \in \text{SO}(3)$ and suppose that $v \in \text{MA}(F_Q)$. We need to show that $(Q, v) \in \text{MA}(F)$. Suppose that $B_F((Q, v)) \subseteq B_F((R, w))$ for some $(R, w) \in F$. In this case, by Lemma IV.8, we must have $R = Q$, and by Lemma IV.7, $B_{F_Q}(v) \subseteq B_{F_Q}(w)$. But because $v \in \text{MA}(F_Q)$, we have $w = v$, hence $(R, w) = (Q, v)$ and $(Q, v) \in \text{MA}(F)$. \square

We define:

$$F' = \bigcup_{Q \in \text{SO}(3)} i_Q(F'_Q).$$

Theorem IV.9. *Suppose that for no $Q \in \text{SO}(3)$ does F_Q contain a half space. Let $(Q, v) \in F'$. Then there is a unique point $(Q', v') \in \text{MA}(F)$ such that $B_F((Q, v)) \subseteq B_F((Q', v'))$. Furthermore, $Q = Q'$ and $v' = r_{F_Q}(v)$.*

Proof. Let $v' = r_{F_Q}(v)$. We have $B_F((Q, v)) \subseteq B_F((Q, v'))$ and $(Q, v') \in i_Q(\text{MA}(F_Q)) \subseteq \text{MA}(F)$ which gives existence.

To see that the point (Q, v') is unique, suppose we have $(R, w) \in \text{MA}(F)$ such that $B_F((Q, v)) \subseteq B_F((R, w))$. Lemma IV.8 tell us that $R = Q$, so we have $B_{F_Q}(v) \subseteq B_{F_Q}(w)$. By the uniqueness of containing balls in \mathbb{R}^3 , it must be that $w = v'$. \square

The preceding theorem allows us to define the canonical retraction map just as we did in the planar case: we map any point (Q, v) in F' to the unique point (Q', v') in $\text{MA}(F)$ such that $B_F((Q, v)) \subseteq B_F((Q', v'))$. Theorem IV.3 above tells us that the point (Q', v') is just $(Q, r_{F_Q}(v))$. This is what allows us to compute images of this retraction map on $\text{SE}(3)$: the canonical retraction is really just happening along fixed-orientation slices of the configuration space. This is not true of the extended retraction map: there are potentially regions of the configuration space for which the extended retraction map involves rotation as well as translation. We will characterize these areas below, at least for polygonal workspaces. There is also again the issue of continuity of the retraction map: again the retraction is not continuous in general, but we will show below that for polygonal workspaces the map is continuous.

E. Continuity of the canonical retraction map

As in the planar case, the medial axis is an SDR of F' under the canonical retraction map if and only if it is closed.

Theorem IV.10. *Suppose that for no $Q \in \text{SO}(3)$ does F_Q contain a half space. If $\text{MA}(F)$ is closed, then the canonical retraction map $r_F : F' \rightarrow \text{MA}(F)$ is continuous.*

Proof. Take $(Q, v) \in F'$ and let (Q_j, v_j) be a sequence in F' converging to (Q, v) . We have $r_F((Q_j, v_j)) = (Q_j, r_{F_{Q_j}}(v_j))$ and $r_F((Q, v)) = (Q, r_{F_Q}(v))$.

The sequence $r_{F_{Q_j}}(v_j)$ must be bounded: reduce to any convergent sub-sequence, and let v' be the limit point of that sequence. (Rather than introduce additional sub-indices, we renumber the sequence.) By the containment of balls just in each slice, we have for each j :

$$\rho_{F_{Q_j}}(r_{F_{Q_j}}(v_j)) = \rho_{F_{Q_j}}(v_j) + d(v_j, r_{F_{Q_j}}(v_j))$$

Because $\rho_{F_Q}(v)$ is a continuous function of Q and v , we have in the limit:

$$\rho_{F_Q}(v') = \rho_{F_Q}(v) + d(v, v'),$$

so that $B_{F_Q}(v) \subseteq B_{F_Q}(v')$ and hence by Lemma IV.7 $B_F(v) \subseteq B_F(v')$. But v' is a limit point of the medial axis, and the medial axis is assumed to be closed, so $v' \in \text{MA}(F)$. This implies that $r_{F_Q}(v) = v'$, which implies that $r_{F_{Q_j}}(v_j) \rightarrow r_{F_Q}(v)$. \square

Now using a homotopy which carries configurations along fixed-orientation geodesics in $\text{SE}(3)$, i.e., pure translations, we have shown that $\text{MA}(F)$ is a strong deformation retract of F' :

Theorem IV.11. *Let $F \subset \text{SE}(3)$ be the free space of a compact workpiece $U \subseteq \mathbb{R}^3$ with a fixed closed obstacle $V \subseteq \mathbb{R}^3$. Suppose that $\text{MA}(F)$ is closed and that for no $Q \in \text{SO}(3)$ does F_Q contain a half space. Then the canonical retraction map $r_F : F' \rightarrow \text{MA}(F)$ is continuous and $\text{MA}(F)$ is a strong deformation retract of F' under the map:*

$$h(t, (Q, v)) = (Q, (1-t)v + tr_{F_Q}(v))$$

where $t \in [0, 1]$.

We defer discussion of the extended retraction map to the following chapter.

F. The medial axis of a free space of polyhedra

In this section we show that the medial axis of the free space of a rigid polyhedral workpiece moving among polyhedral obstacles is closed. With the results of the previous sections this will show that the medial axis is always a strong deformation retract of the free space under the canonical retraction map.

A convex polyhedron is usually defined as the intersection of a finite number of open half spaces. Many of the problems found in practice do not involve only convex bodies; for this reason, we allow both the workpiece and the obstacles to be finite unions of such convex polyhedra. Also, for our application, we would like to have the objects be closed sets. This motivates us to define a polyhedron P as a set of the form:

$$P = \overline{\bigcup_{i=1}^N P_i},$$

where each P_i is of the form:

$$P_i = \bigcap_{j=1}^{N_i} H_{ij},$$

where each H_{ij} is an open half space, i.e.,

$$H_{ij} = \{x \in \mathbb{R}^3 \mid (x - p_{ij}) \cdot \hat{n}_{ij} < 0\},$$

where $p_{ij} \in \mathbb{R}^3$ and \hat{n}_{ij} is a unit vector in \mathbb{R}^3 .

For the remainder of this section, let $F \subseteq \text{SE}(3)$ be the free space for a compact polyhedral workpiece U with polyhedral workspace obstacle V . We assume that for no $Q \in \text{SO}(3)$ does F_Q contain a half space. This will be true if, for example, the complement of V is bounded.

Proposition IV.12. *With U , V , and F as above, let $(Q, v) \in \text{MA}(F) \cap F^\circ$. Then there are at least two points in $\partial F \cap B_F((Q, v))$; i.e., (Q, v) is a multiple point of F .*

Proof. Suppose that there is only a single point $(Q', v') \in \partial F \cap B_F((Q, v))$. By Lemma IV.5, we have $Q' = Q$. Let $a \in U$ and $b \in V$ be the unique points on U and V such that $a' = (Q, v) \cdot a$ and b are the closest points between $(Q, v) \cdot U$ and V , i.e., $d(a', b) = \text{dist}((Q, v) \cdot U, V)$. Let $u = v - v'$ and $\hat{u} = u/|u|$; then we have $a' = b + u$. We claim that for $\epsilon > 0$ small enough, a and b are the only points on U and V for which $d((Q, v + \epsilon\hat{u}) \cdot a, b) = \text{dist}((Q, v + \epsilon\hat{u}) \cdot U, V)$. If we can show this, then we will have $B_F((Q, v)) \subset B_F((Q, v + \epsilon\hat{u}))$ which then shows that (Q, v) is not in $\text{MA}(F)$.

Take H_V to be the halfspace $\{x \in \mathbb{R}^3 \mid (x-b) \cdot u \leq 0\}$. Because V is a polyhedron, there is a small neighborhood N_V around b so that $V \cap N_V$ is entirely contained in H_V . Similarly, we have $H_U = \{x \in \mathbb{R}^3 \mid (x-a') \cdot u \geq 0\}$, and a neighborhood N_U so that $(Q, v) \cdot U \cap N_U$ is contained in H_U . Let ϵ be the smaller of the two quantities

$$\frac{1}{2}(\text{dist}(V, (Q, v) \cdot U) - \text{dist}(V \setminus N_V, (Q, v) \cdot U))$$

and

$$\frac{1}{2}(\text{dist}((Q, v) \cdot U, V) - \text{dist}((Q, v) \cdot U \setminus N_U, V)).$$

Note that each of these expressions is positive, so we have $\epsilon > 0$. Now, moving $(Q, v) \cdot U$ by ϵ in the \hat{u} direction, we have

$$d((Q, v + \epsilon\hat{u}) \cdot a, b) = d((Q, v) \cdot a, b) + \epsilon.$$

And by the definition of ϵ , these must be the nearest points on $(Q, v + \epsilon\hat{u}) \cdot U$ and V . □

To interpret this result in the workspace note that Lemma IV.5 implies that $\partial F \cap B_F((Q, v)) \subset i_Q(\mathbb{R}^3)$, i.e., the nearest boundary points of F to (Q, v) are of the form (Q, v') , $v' \in \mathbb{R}^3$. This means that there are distinct translations of $(Q, v) \cdot U$ that put it into contact with V , i.e., there are points a and b on $(Q, v) \cdot U$ and a' and

b' on V such that $\rho_F((Q, v)) = \text{dist}((Q, v) \cdot U, V) = d(a, a') = d(b, b')$ and such that $b - a \neq b' - a'$.

Theorem IV.13. *With F as above, the medial axis $\text{MA}(F)$ is closed.*

Proof. Take a sequence $(Q_j, v_j) \in \text{MA}(F)$, $j = 1, 2, \dots$, converging to a point $(Q, v) \in F$. We need to show that $(Q, v) \in \text{MA}(F)$. By the above proposition and Lemma IV.5, we know that for each j there are distinct v'_j and v''_j such that (Q_j, v'_j) and (Q_j, v''_j) are in $B_F((Q_j, v_j)) \cap \partial F$; i.e., so that (Q_j, v'_j) and (Q_j, v''_j) are distinct nearest boundary points of (Q_j, v_j) . If the set

$$S = \{(Q_j, v'_j) \mid j = 1, 2, \dots\} \cup \{(Q_j, v''_j) \mid j = 1, 2, \dots\}$$

has more than one limit point, say (Q, v') and (Q, v'') are distinct limit points, then by continuity of the distance function, these will be two nearest boundary points of (Q, v) , whence $(Q, v) \in \text{MA}(F)$. Otherwise, the set S above has a unique limit point, call it (Q, v') .

What we will show in this case is that $(Q, v) = (Q, v')$ and that $(Q, v') \in \text{MA}(F)$. Let $n'_j = (v_j - v'_j)/|v_j - v'_j|$ and $n''_j = (v_j - v''_j)/|v_j - v''_j|$. The vector v_j is at the intersection point of the rays

$$v'_j + s n'_j$$

$$v''_j + t n''_j$$

for $s, t > 0$. Now the sequences v'_j and v''_j approach v' as $j \rightarrow \infty$, and v_j approaches v so the unit vectors n'_j and n''_j cannot approach a common direction, otherwise their intersection point would become arbitrarily far from v' as j grew larger. Therefore, the n'_j approach some direction n' and the n''_j approach some direction n'' and $n' \neq n''$. But then the intersection points v_j must approach v'_j as j gets larger, so we have $v_j \rightarrow v'$ as $j \rightarrow \infty$, i.e., $v = v'$.

Having established that $v = v'$ is on the boundary, we need to show that in fact v' is on the medial axis. But this is easy to see: any disc of nonzero radius in \mathbb{R}^3 containing v' would necessarily contain some points of either the sequence v'_j or v''_j .

□

CHAPTER V

THE MAPRM ALGORITHM

The above theory produces a retraction algorithm that is simple to express, at least for free configurations: given a randomly chosen free configuration (Q, v) of the polyhedral workpiece U , we find the nearest point on the obstacle V to $(Q, v) \cdot U$. Say $p \in (Q, v) \cdot U$ and $p' \in V$ are the closest points. We simply move U away from the nearest point p' along the line connecting p and p' . As U recedes from p' , at some point there there be two distinct equal-length translations to contact configurations: one taking p to p' and another one. At that point, the configuration of U is on the medial axis. In practice this is done using a bisection method.

Carrying out the extended retraction map is more complicated. Recall that the idea behind the extended retraction map is to move everything outside of F to its nearest point on ∂F (if a unique nearest point exists). The next section discusses the details of carrying this out on $\text{SE}(3)$.

A. Finding the nearest contact configuration

For a free configuration $(Q, v) \in F'$, the image under the extended retraction map is simply $(Q, r_{F_Q}(v))$, i.e., only the translational component of the original configuration is modified. For collision configurations, evaluating the extended retraction map involves first retracting to the nearest configuration in ∂F (a contact configuration) under the chosen Riemannian metric on $\text{SE}(3)$, and then applying the canonical retraction map. The nearest configuration to a given collision configuration does *not* in general have the same translational component as the original configuration. (Imagine a situation in which there is a small rotation that will free the workpiece, but there no small translation that will free it.)

Computing the nearest contact configuration for a collision configuration (Q, v) involves a computation which amounts to calculating explicitly the structure of the C -obstacles in the vicinity of (Q, v) . Given the expense of such a calculation, we should examine whether it makes sense to go to the trouble to retract such configurations.

A simpler version of the problem of finding the nearest contact configuration, involves finding the shortest *translation* required to free the workpiece, usually referred to as the *minimum translation distance*, denoted $MTD(Q, v)$. Some research has been conducted in this area, although only for convex bodies, see [12] for example. Given a workpiece and obstacles presented as polyhedra, an exhaustive approach to this problem could proceed as follows. For each pair of features (faces, edges, and vertices) on the workpiece and obstacle, we compute the minimum translation required to put those features in contact; this yields a configuration which can be either in contact or collision. After exhausting all pairs features, and throwing out any collisions, we then choose the smallest translation. The resulting translation is guaranteed to be $MTD(Q, v)$ only if the nearest contact configuration to (Q, v) places the workpiece in contact with the obstacles in a single point. Any contact configuration placing the workpiece in contact with the obstacles at more than one point is on the medial axis of the free space, so this exhaustion method will fail only for collision configurations whose nearest contact configuration is a contact configuration on the medial axis of the free space.

Now we need to consider what action to take for a collision configuration whose nearest contact configuration is not obtained by simply translating, but requires some rotation; i.e., the nearest contact configuration has the workpiece at a different orientation. The following proposition shows that for polyhedra, any such contact configuration is on the medial axis of the free space.

Proposition V.1. *Let $(Q, v) \in \text{SE}(3) \setminus F$ and let $(R, w) \in \partial F$ be a nearest contact configuration to (Q, v) . If $R \neq Q$, then $(R, w) \in \text{MA}(F)$.*

Proof. Let $\gamma : [0, L] \rightarrow \text{SE}(3)$ be a minimizing geodesic between $(Q, v) = \gamma(0)$ and $(R, w) = \gamma(L)$. The idea is that if (R, w) is not on the medial axis, then we can stop γ just short of configuration (R, w) , say at $t = L - \epsilon$, and then just translate U (at fixed orientation) by a small amount more so that the workpiece meets V only at one of the original contact points for $\gamma(L) \cdot U$. The assumption that (R, w) is not on the medial axis will give sufficient “slack” that under a slight rotation of U , the position can be corrected to give a contact configuration.

Suppose that (R, w) is not on the medial axis of F . We claim that in this case there must be at least one point $p' \in (R, w) \cdot U \cap V$ and an $\epsilon > 0$ such that the configuration

$$(I, p' - \gamma(L - \epsilon) \cdot p) \cdot \gamma(L - \epsilon)$$

is in ∂F , where $p = (R, w)^{-1} \cdot p'$. Note that this configuration maps p to $p' = (R, w) \cdot p$.

If we can show this claim to be true, then as in Assumption 2 on the metric, we look at $S = \{g \in \text{SE}(3) \mid g \cdot p = p'\}$. We have:

$$(I, p' - \gamma(L - \epsilon) \cdot p) \cdot \gamma(L - \epsilon) \cdot p = p'$$

and

$$(\gamma(L) \cdot \gamma(L - \epsilon)^{-1}) \cdot \gamma(L - \epsilon) \cdot p = p',$$

i.e., both of these transformations are in S . By Assumption 2, the nearest configuration in S to $\gamma(L - \epsilon)$ must differ from $\gamma(L - \epsilon)$ by a pure translation, which means that

$$d((I, p' - \gamma(L - \epsilon) \cdot p) \cdot \gamma(L - \epsilon), \gamma(L - \epsilon)) < d(\gamma(L), \gamma(L - \epsilon)).$$

But this has produced a contact configuration $(I, p' - \gamma(L - \epsilon) \cdot p) \cdot \gamma(L - \epsilon)$ that is

closer to $\gamma(0) = (Q, v)$ than $\gamma(L) = (R, w)$ is, a contradiction.

We still need to show that our initial claim is justified. Because (R, w) is not on the medial axis, there exists a unit vector $\hat{\beta}$ in \mathbb{R}^3 and a $\delta > 0$ so that $\{(R, w)\} = B_F((R, w)) \subset B_F((R, w + \delta\hat{\beta}))$. This means that $\rho_F((R, w + \delta\hat{\beta})) = \delta$, in other words, that

$$\text{dist}((R, w + \delta\hat{\beta}) \cdot U, V) = \delta.$$

We choose $\delta > 0$ small enough that $(R, w + \delta\hat{\beta})$ is not on the medial axis. Let $\gamma = (\gamma_1, \gamma_2)$. The idea is to take U first in the configuration (R, w) , move it by δ in the $\hat{\beta}$ direction, rotate back slightly (so that the rotation component is $\gamma_1(L - \epsilon)$), and then move U back into contact with V . If in configuration (R, w) there is only a single contact point $p' = (R, w) \cdot p$ between U and V , then after translating U by $\delta\hat{\beta}$ and making a sufficiently small rotation $\gamma_1(L - \epsilon) \cdot \gamma_1(L)^{-1}$, the points p and p' will still be a closest pair between U and V . Then U can be translated back so that p coincides with p' and U is in contact (but not collision) with V . If there are multiple contact points, we choose p so that the small rotation $\gamma_1(L - \epsilon) \cdot \gamma_1(L)^{-1}$, carried out with the point p' fixed, causes the other contact points to move in a direction whose projection onto $\hat{\beta}$ is non-negative.

□

We conclude that by allowing only translations and using the the exhaustion method described above we can retract any collision configuration whose final retracted image is not a contact configuration on the medial axis of the free space; i.e., we can retract anything in the inverse image of $\text{MA}(F) \cap F^\circ$ under the extended retraction map.

We should also remark that this cut-down version of the extended retraction map, i.e., ignoring configurations that can't be retracted using only translation, still

maps configurations outside F continuously onto ∂F , so that the continuity of the overall retraction is not disturbed.

The obvious question here is: does it harm the algorithm's performance to simply ignore such configurations (or retract them incorrectly)? We have no firm answer to this question, but some insight can be gained by looking at a simple 2-D example. In Figure 14, interpret the interior of the square as the free space F . For a general

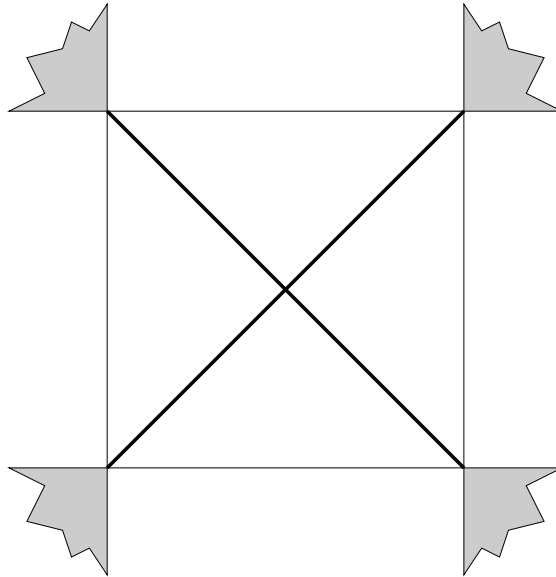


Fig. 14. Motivation in 2-D.

point on the medial axis of the square, there is a pair of 1-D rays of points in the plane that retract to it.

However, for each of the corner points, which are in $\text{MA}(F) \cap \partial F$, there is an entire 2-D cone of points (the shaded regions) which retract to them. When sampling and retracting, this means that the corner points are “hot spots” which have nonzero

probability of being hit when a single point is sampled and retracted. When a large number of points is sampled and retracted, the corner points will get hit repeatedly.

For points outside the square, the computation of the nearest point on the boundary of the square is not very computationally intensive, but assume for the moment that for points in the shaded regions the computation is much more expensive than for other points. If we simply neglected sampled points in the shaded regions, then the corner points would have now a pair of rays (the bounding rays of the cones) that retract to them, giving them approximately the same chance of getting hit as the other points. Our situation in the SE(3) case is essentially analogous to this, with one slight twist: we don't actually figure out whether a point is in what corresponds to the shaded regions: we just proceed as if it is not. We may find no nearest contact configuration for such a configuration, or we may retract it erroneously.

Algorithm V.1 gives our exhaustive method for finding the nearest contact configuration.

Algorithm V.1 Finding the nearest contact configuration

Input. A configuration (R, p) .

Output. A shortest translation (R, q) from configuration (R, p) that puts U in contact (but not collision) with an obstacle, or **failure**.

- 1: **if** (R, p) is a free configuration **then**
 - 2: Return $(R, p + (x - y))$, where $x \in (R, p) \cdot U$ and $y \in V$ are a pair of closest points between U and V .
 - 3: **else**
 - 4: For each feature (vertex, edge, face) of U , and each feature of the V_j , find the configuration (R, q) with smallest $|p - q|$ that puts these features in contact in a single point.
 - 5: Such a configuration may put U strictly in collision (not just contact) with an obstacle; discard any such configurations.
 - 6: If no configurations remain, output **failure**.
 - 7: Otherwise, output a remaining configuration (R, q) with smallest $|p - q|$.
 - 8: **end if**
-

The MAPRM sample-and-retract algorithm is given in Algorithm V.2.

Algorithm V.2 MAPRM for rigid bodies in three dimensions

Preprocessing:

Input. N , the number of nodes to generate.

Output. N nodes in F connected into a roadmap.

- 1: **repeat**
 - 2: Sample a configuration (R, p) from C .
 - 3: Use Algorithm V.1 to get the nearest contact configuration (R, q) to (R, p) .
 - 4: **if** (R, p) is free **then**
 - 5: Take the retraction direction \vec{v} to be \overrightarrow{qp} , and let the start point s be p .
 - 6: **else**
 - 7: Take the retraction direction \vec{v} to be \overrightarrow{pq} , and let the start point s be q .
 - 8: **end if**
 - 9: Let $a \in U$ and $b \in V$ be a pair of closest point on U and V when U is in configuration (R, s) .
 - 10: Starting in configuration (R, s) , translate U in the direction \vec{v} until there is another pair of closest points $a' \in U$ and $b' \in V$ with $b - a \neq b' - a'$. This configuration is on the medial axis of the free space.
 - 11: **until** N vertices have been output
 - 12: For each pair of vertices: if the pair can be connected with the local planner, insert an edge into the graph connecting them.
-

B. Implementation and experimental results

We implemented MAPRM for rigid bodies using V-Clip [30] to provide collision detection and closest pair calculations. We used a single local planner: translation with simultaneous rotation about the principal axis of rotation in the body frame. In using probabilistic roadmap methods in practice, the the time to connect nodes usually overwhelms the time spent in the sampling phase. As a result, connections are not usually attempted between all pairs of nodes, but only between promising pairs of nodes according to some heuristic. In our examples, so few nodes are generated that

it was feasible to attempt connections between all pairs necessary to determine the components of the roadmap.

1. Narrow corridor example

Our first example is shown in Figure 15. The workpiece is a cube of side length 2;

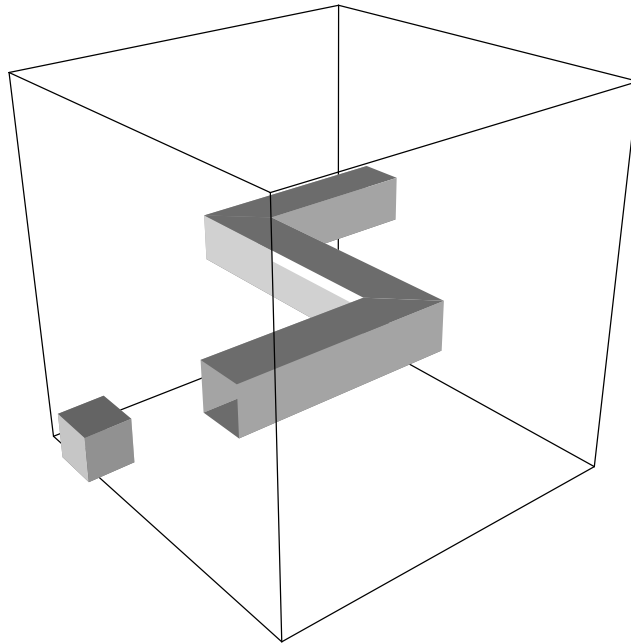


Fig. 15. Rigid body example. The workpiece is the small cube and the obstacle is a solid block with the indicated corridor cut through it.

the obstacle is a solid cube of side length 20 with the indicated corridor cut through

it. The corridor has 2.5×2.5 cross section.

We compared experiments using uniform sampling (with collisions discarded) against sampling with MAPRM. The same local planner, collision detection, connection scheme, etc., were used for both methods. Configurations were sampled with arbitrary rotation, and translations placing the center of the workpiece anywhere inside the $20 \times 20 \times 20$ cube. The retraction and collision checking calculations were carried out with a spatial tolerance of 0.01. Execution of each method was terminated when some component of the roadmap reached from one mouth of the corridor to the other.

The mean results for 15 runs are given in Table I. Experiments were run on a

Table I. Experimental results for cubical workpiece.

	MAPRM	Uniform
Sampling time (seconds)	690	7875
Connection time (seconds)	82	79
Total preprocessing time (seconds)	772	7954
# roadmap nodes required	404	351
# random config. sampled	39,568	114,058,889
Mean time to generate a node (seconds)	1.706	22.652

MIPS R10000 processor running at 200 Mhz. Observe that on average the MAPRM algorithm solved the problem in less than one-tenth of the time required by uniform random sampling. MAPRM generated nodes at about 13 times the rate of uniform sampling. Note the huge number of random samples required using uniform sampling. However, observe that MAPRM generally required more nodes in the roadmap to be able to solve the problem. We attribute this to the non-uniform distribution of nodes

generated by MAPRM along the medial axis: in general there will be somewhat fewer nodes sampled near corners than in the straight sections. However, the much greater sampling rate of MAPRM far outstrips this demand for additional nodes. This effect warrants further investigation.

2. Wider corridor example

We also compared performance on a simpler problem: the same obstacle block but with a workpiece cube of side length 1.5 units (rather than 2 units). Mean results for 15 trials are given in Table II. In this case both methods solved the problem very

Table II. Experimental results for a smaller workpiece.

	MAPRM	Uniform
Sampling time (second)	14.6	10.5
Connection time (seconds)	12.4	4.4
Total preprocessing time (seconds)	27.0	14.9
# roadmap nodes required	109	57
# random config. sampled	881	157,333
Mean time to generate a node (second)	0.133	0.184

quickly, but uniform sampling was quicker on average than MAPRM. Still, MAPRM generated free nodes at a higher rate than uniform sampling. Again, note the significantly larger number of roadmap nodes required by MAPRM.

We should also note that the variance from the mean was very high in both of these examples. For example, in the narrow corridor case, the mean sampling time for MAPRM was 690 seconds with standard deviation 163; the mean sampling time with uniform sampling was 7875 seconds with standard deviation 2514. This is probably

due to the fact that a few “key” nodes near the corners are sufficient to solve the problem. In more realistic examples, connections are usually only attempted between nearby pairs of nodes, so we would expect these deviations to decrease.

CHAPTER VI

CONCLUSIONS AND FURTHER RESEARCH

We have described a sampling method for probabilistic path planning for a rigid body which retracts sampled configurations onto the medial axis of the free space. The retraction method assumed that a Riemannian metric had been chosen on the configuration space which satisfied certain geometric properties. With such a metric, retraction of a free configuration onto the medial axis was shown to require only translation of the workpiece at fixed orientation. Furthermore, it was shown how the retraction could be carried out without first having to compute the medial axis explicitly. This permitted an algorithm to be developed which retracted nearly all sampled configurations, not just those in the free space, onto the medial axis in a computationally feasible manner.

While most probabilistic roadmap methods require only collision detection, our retraction algorithm requires additional geometric computations. The calculation is simple for nodes in the free space, requiring only a closest pair primitive in the workspace; this is produced as a by-product by many collision detection algorithms. For sampled nodes that land outside the free space, however, the retraction is more complicated, requiring a calculation of the closest boundary (i.e., contact) configuration. Our algorithm only guarantees retraction of nodes whose final destinations on the medial axis are non-contact configurations.

The goal of the method was to improve performance on problems requiring the traversal of narrow corridors in configuration space: theoretical evidence for the planar case and some preliminary experimental results for the rigid body case were given. It was shown in the planar case that the increase in the number of nodes found in a narrow corridor depends not on the volume of the corridor but on the characteristics

of the obstacles bounding it.

There is clearly a need for a more thorough experimental evaluation of this method. Although the method performs well on our examples, it employs more complicated geometric calculations than the standard uniform sampling and is consequently slightly more difficult to implement. For larger problems, we expect the time required for the calculation of the nearest contact configuration to become more significant: additional sophistication will probably be required in that calculation.

There is also a clear need for theoretical results regarding the overall probability of success of the method. This would permit us to evaluate the performance impact of various choices we have made, such as the choice of Riemannian metric.

We would like to apply this technique to higher dimensional problems such as articulated robots, an area in which probabilistic methods have been very successful. More theoretical work on the medial axis will probably be required in order to apply it to these situations.

REFERENCES

- [1] N. AMATO, O. B. BAYAZIT, L. K. DALE, C. JONES, AND D. VALLEJO, *OBPRM: An obstacle-based PRM for 3D workspaces*, in Proc. Int. Workshop on Algorithmic Foundations of Robotics, P. K. Agarwal, L. E. Kavraki, and M. T. Mason, eds., A. K. Peters, Wellesley, MA, 1998, pp. 155–168.
- [2] N. M. AMATO, O. B. BAYAZIT, L. K. DALE, C. V. JONES, AND D. VALLEJO, *Choosing good distance metrics and local planners for probabilistic roadmap methods*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1998, pp. 630–637.
- [3] N. M. AMATO, O. B. BAYAZIT, K. KIM, W. SON, AND G. SONG, *Providing haptic ‘hints’ to automatic motion planners*, Tech. Report 98-026, Dept. of Computer Science, Texas A&M University, 1998.
- [4] N. M. AMATO AND L. K. DALE, *Probabilistic roadmap methods are embarrassingly parallel*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1999, pp. 688–694.
- [5] N. M. AMATO AND Y. WU, *A randomized roadmap method for path and manipulation planning*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1996, pp. 113–120.
- [6] V. I. ARNOLD, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, NY, 2nd ed., 1989.
- [7] F. AURENHAMMER, *Voronoi diagrams: A survey of a fundamental geometric data structure*, ACM Comput. Surv., 23 (1991), pp. 345–405.

- [8] O. B. BAYAZIT, G. SONG, AND N. M. AMATO, *Enhancing randomized motion planners: Exploring with haptic hints*, Tech. Report 99-021, Dept. of Computer Science, Texas A&M University, 1999.
- [9] H. BLUM, *A transformation for extracting new descriptors of shape*, in *Models for the Perception of Speech and Visual Form*, W. Wathen-Dunn, ed., MIT Press, Cambridge, MA, 1967, pp. 362–380.
- [10] V. BOOR, M. H. OVERMARS, AND A. F. VAN DER STAPPEN, *The Gaussian sampling strategy for probabilistic roadmap planners*, in *Proc. IEEE Int. Conf. Robot. Autom.*, IEEE Robotics and Automation Society, Piscataway, NJ, 1999, pp. 1018–1023.
- [11] G. E. BREDON, *Topology and Geometry*, Springer-Verlag, New York, NY, 1993.
- [12] S. CAMERON, *Enhancing GJK: Computing minimum and penetration distances between convex polyhedra*, in *Proc. IEEE Int. Conf. Robot. Autom.*, IEEE Robotics and Automation Society, Piscataway, NJ, 1997, pp. 3112–3117.
- [13] J. CANNY, *The Complexity of Robot Motion Planning*, ACM – MIT Press Doctoral Dissertation Award Series, MIT Press, Cambridge, MA, 1987.
- [14] H. I. CHOI, S. W. CHOI, AND H. P. MOON, *Mathematical theory of medial axis transform*, *Pacific J. Math.*, 181 (1997), pp. 57–88.
- [15] M. P. DO CARMO, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976.
- [16] M. P. DO CARMO, *Riemannian Geometry*, Birkhauser Boston, Boston, MA, 1992.

- [17] P. FINN, L. E. KAVRAKI, J.-C. LATOMBE, R. MOTWANI, C. SHELTON, S. VENKATASUBRAMANIAN, AND A. YAO, *RAPID: Randomized pharmacophore identification for drug design*, in Proc. ACM Symp. on Computational Geometry, ACM, New York, NY, 1997, pp. 324–333.
- [18] C. M. HOFFMAN, *How to construct the skeleton of CSG objects*, in Computer-aided Surface Geometry and Design: The Mathematics of Surfaces IV, A. Bowyer, ed., Clarendon Press, Oxford, UK, 1994, pp. 421–437.
- [19] D. HSU, L. E. KAVRAKI, J.-C. LATOMBE, R. MOTWANI, AND S. SORKIN, *On finding narrow passages with probabilistic roadmap planners*, in Proc. Int. Workshop on Algorithmic Foundations of Robotics, P. K. Agarwal, L. E. Kavraki, and M. T. Mason, eds., A. K. Peters, Wellesley, MA, 1998, pp. 141–153.
- [20] D. HSU, J.-C. LATOMBE, AND R. MOTWANI, *Path planning in expansive configuration spaces*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1997, pp. 2719–2726.
- [21] L. KAVRAKI, *Random Networks in Configuration Space for Fast Path Planning*, Ph.D. dissertation, Stanford Univ., Stanford, CA, 1995.
- [22] L. KAVRAKI, M. KOLOUNTZAKIS, AND J.-C. LATOMBE, *Analysis of probabilistic roadmaps for path planning*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1996, pp. 3020–3025.
- [23] L. KAVRAKI, F. LAMIRAUX, AND C. HOLLEMAN, *Towards planning for elastic objects*, in Proc. Int. Workshop on Algorithmic Foundations of Robotics, P. K. Agarwal, L. E. Kavraki, and M. T. Mason, eds., A. K. Peters, Wellesley, MA, 1998, pp. 313–325.

- [24] L. E. KAVRAKI, J.-C. LATOMBE, R. MOTWANI, AND P. RAGHAVAN, *Randomized query processing in robot path planning*, in Proc. ACM Symp. Theory of Computing (STOC), ACM, New York, NY, 1995, pp. 353–362.
- [25] L. E. KAVRAKI, P. SVESTKA, J.-C. LATOMBE, AND M. H. OVERMARS, *Probabilistic roadmaps for path planning in high dimensional configuration spaces*, IEEE Trans. Robot. Autom., 12 (1996), pp. 566–580.
- [26] J.-C. LATOMBE, *Robot Motion Planning*, Kluwer Academic Publishers, Boston, MA, 1991.
- [27] S. LAVALLE, J. YAKEY, AND L. KAVRAKI, *A probabilistic roadmap approach for systems with closed kinematic chains*, in Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1999, pp. 1671–1676.
- [28] M. C. LIN AND S. GOTTSCHALK, *Collision detection between geometric models: A survey*, in Proc. 8th IMA Conf. on The Mathematics of Surfaces, R. Cripps, ed., Information Geometers, Winchester, UK, 1998, pp. 1–20.
- [29] J. LONCARIC, *Geometrical Analysis of Compliant Mechanisms in Robotics*, Ph.D. dissertation, Harvard Univ., Cambridge, MA, 1985.
- [30] B. MIRTICH, *V-Clip: Fast and robust polyhedral collision detection*, Tech. Report 97-05, Mitsubishi Electric Research Lab, Cambridge, MA, 1997.
- [31] F. MORGAN, *Riemannian Geometry: A Beginner's Guide*, A. K. Peters, Wellesley, MA, 2nd ed., 1998.
- [32] J. R. MUNKRES, *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

- [33] R. M. MURRAY, Z. LI, AND S. S. SASTRY, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, FL, 1994.
- [34] F. PARK, *Distance metrics on the rigid-body motions with applications to mechanism design*, *Journal of Mechanical Design*, 117 (1995), pp. 48–54.
- [35] J. SERRA, *Image Analysis and Mathematical Morphology*, Academic Press, London, UK, 1982.
- [36] J. SERRA, *Image Analysis and Mathematical Morphology Volume 2: Theoretical Advances*, Academic Press, London, UK, 1988.
- [37] E. C. SHERBROOKE, N. M. PATRIKALAKIS, AND F.-E. WOLTER, *Differential and topological properties of medial axis transforms*, *Graphical Models and Image Processing*, 58 (1996), pp. 574–592.
- [38] P. SVESTKA, *Robot motion planning using probabilistic roadmaps*, Ph.D. thesis, Utrecht Univ., Utrecht, The Netherlands, 1997.
- [39] P. SVESTKA AND M. H. OVERMARS, *Coordinated path planning for multiple robots*, *Robotics and Autonomous Systems*, 23 (1998), pp. 125–152.
- [40] D. VALLEJO, C. V. JONES, AND N. M. AMATO, *An adaptive framework for ‘single shot’ motion planning*, Tech. Report 99-024, Dept. of Computer Science, Texas A&M University, 1999.
- [41] P. J. VERMEER, *Medial Axis Transform to Boundary Representation Conversion*, Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 1994.
- [42] S. A. WILMARTH, N. M. AMATO, AND P. F. STILLER, *MAPRM: A probabilistic roadmap planner with sampling on the medial axis of the free space*, in

- Proc. IEEE Int. Conf. Robot. Autom., IEEE Robotics and Automation Society, Piscataway, NJ, 1999, pp. 1024–1031.
- [43] S. A. WILMARTH, N. M. AMATO, AND P. F. STILLER, *Motion planning for a rigid body using random networks on the medial axis of the free space*, in Proc. ACM Symp. on Computational Geometry, ACM, New York, NY, 1999, pp. 173–180.
- [44] R. H. WILSON, *On Geometric Assembly Planning*, Ph.D. dissertation, Stanford Univ., Stanford, CA, 1992.
- [45] F. E. WOLTER, *Cut locus and medial axis in global shape interrogation and representation*, Tech. Report 92-2, Dept. Ocean Eng., Design Lab, MIT, Cambridge, MA, 1992.
- [46] M. ZEFRAN, V. KUMAR, AND C. CROKE, *Choice of Riemannian metrics for rigid body kinematics*, in Proc. of the 1996 ASME Eng. Tech. Conf. and Comp. in Eng. Conf., ASME, New York, NY, 1996, pp. 1–11.

VITA

Name: Steven Albert Wilmarth

Address: 14740 Carona Drive, Silver Spring, MD 20905

Email: wilmarth@metsci.com

Education: M.S. in Mathematics, Rutgers University, New Brunswick,
NJ, May 1992.

B.S. in Engineering Physics, Tufts University, Medford,
MA, May 1989.

Employment: Metron, Inc., 11911 Freedom Drive Suite 800, Reston, VA,
20190. (703) 787-8700. February 1999 to present.