Shape Approximation in Kinematic Systems

Ernest Davis*
New York University
251 Mercer St.
New York, New York
DAVISE@CS.NYU.EDU

June 15, 2005

Abstract

The vast majority of physical calculations that involve the shapes of objects are carried out by approximating the actual shape in terms of a idealized, or nominal, shape description. It then becomes a problem to determine whether calculations based on idealized shapes in fact carry over to actual shapes. This paper addresses the following instance of that problem, in the domain of solid object kinematics: Suppose that it is known that the ideal shapes approximate the real shape to a specified tolerance, or degree of accuracy. In what respects, or to what degree, can we expect that the physical behavior will resemble the calculated behavior?

In studying this problem, we consider two definitions of "shape approximation" and three definitions of "similarity of physical behavior," all original to this paper, and we prove a number of theorems relating these. Moreover, we show that these relations are, at least in principle, computable, if the idealized shapes are semi-algebraic. The results in this paper are the first to show that a flexible representation of incomplete knowledge of shape can support computable reasoning in a rich physical domain.

Keywords: Shape approximation, shape tolerance, mechanical tolerance, kinematics, configuration space, Hausdorff distance.

1 Introduction

In most computations that involve the shapes of real objects, the true shape of the objects is approximated by a *nominal* or *ideal* shape. That is, one has a program that uses representations of ideal spatial regions, such as polyhedra, or cubic splines, or Fourier components. One constructs a representation in this language of a nominal shape that approximates closely the actual shape of the object; and one uses this nominal shapes for geometric calculations; and then one hopes that results of the calculations over the nominal shape are valid over the actual shape. For instance, if an engineer models an airplane wing in terms of a triangulation of its surface, it is not because he imagine that that is the actual shape of the wing, and that if he looks close enough he can find the edges and vertices in the physical wing; it is because he has some reason to believe that

*This research was supported by NSF grants #IRI-9300446 and #IRI-9625859. Thanks to David MacAllester for the suggestion that I look for computability results by using reduction to Tarski's theorem; to Leo Joskowicz for many helpful comments and suggestions; also to Dana Angluin, Mladen Covic, Philip Davis, Drew McDermott, Bud Mishra, Igor Najfeld, Chee Yap, and Ken Yip.

calculations based on the triangulation will be close enough to the actual values for the real shape, for the questions he wants to answer and the degree of precision that he requires.

There are several reasons that approximations of this kind are necessary in practice. First, the true shape may not be precisely expressible in terms of the representation language; for example, the representation language uses polyhedra and the true shape is a smooth curve, or the representation language is algebraic and the true shape is a helix. Second, approximating a complex shape by a simpler one may substantially reduce computation costs (e.g. Fleischer et al., 1992). Third, in many cases, the true shape cannot be fully determined. For example, it may not be possible to measure the shape with sufficient accuracy, or the shape may be an ideal to be manufactured using a process with some error tolerance, and so on (e.g. Joskowicz and Taylor, 1996; Requicha, 1983).

The use of this kind of approximation — which, it is worth reiterating, is ubiquitous, almost universal, in applied geometric computation — immediately raises two questions. First, what does it mean for one shape to approximate another? Second, how can one go from calculations based on the nominal shape to conclusions about the actual shape? The first question is the subject of the theory of *tolerancing*, which has been extensively studied, both from an engineering standpoint (e.g. Neumann, 1994) and from a theoretical standpoint (e.g. Yap and Chang, 1996). Much less is known about the second question, particularly for physical as opposed to purely geometrical computations.

In practice, the question is generally addressed using engineering rather than theoretical methods. An engineer has a sense from experience of how accurate an approximation he needs for a given application, and then can test the reliability of a calculation either by physical experimentation or by computational techniques such as increasing the accuracy and checking to see whether the result change much. Nonetheless, theoretical results have substantial value, as in most domains (also, of course, substantial limitations.) Our particular interest in these results is to study explicit shape approximations as a form of qualitative spatial information, and to study what aspects of shapes are important in physical reasoning by determining what aspects must be preserved under approximation in order guarantee desired physical consequences.

This paper addresses the question of the reliability of calculations based on shape approximations for calculations in the kinematic theory of rigid solid objects (KRSO). KRSO deals with the possible motions of a collection of solid objects that are idealized as perfectly rigid in shape. The theory is narrow but extremely important in practical terms; for example, it is powerful enough to characterize a large fraction of the mechanisms that are used in practice (Joskowicz and Sacks, 1991). Thus this paper addresses the following general question: Suppose we have calculated the behavior of some kinematic system using ideal descriptions of the shapes of the objects involved. Does it then follow that a real mechanism, in which the shapes of the objects approximate this ideal, will have a similar behavior? Our answer consists of three theorems of the following form: "If the actual shapes of objects lie close enough, under a specified tolerance, to the nominal shape, then certain kinematic properties of the actual shapes lie close to the calculated properties for the nominal shapes." We consider three different kinematic properties and two different shape tolerances. Since the kinematic behavior of a system of objects is highly discontinuous as a function of the shape, these results are far from obvious. As far as we know, this is the first result obtained that shows that physical behaviors can be well behaved under small shape tolerances.[1]

To address this question, we must define what it means (a) for one shape to approximate another and (b) for the behavior of one mechanism to be similar to the behavior of another. As regards (a), we will consider two general criteria of shape approximation. The first, in terms of the dual-Hausdorff distance, considers that shape **A** approximates **B** if every point in **A** is close to a point in **B** and vice versa, and every point outside **A** is close to a point outside **B** and vice versa. The

[1]The results of (Joskowicz, Sacks, and Srinivasan, 1997) relate to parametric tolerance; it is assumed that the actual shape lies in a family of regions that vary in terms of a fixed, finite number of one-dimensional parameters, such as the family of rectangles. The discussion here uses the much broader notion of shape tolerance.

second criterion, called "approximation in tangent", adds the further requirement that the tangents at corresponding points on the boundaries of **A** and **B** are close.

In addressing (b) above, we will characterize the behavioral properties of a kinematic system in terms of its *configuration space*; that is, the set of physically feasible positions and orientations of the objects. Thus, for the purposes of this paper, we will consider the behavior of two mechanical systems to be "similar" if their respective configuration spaces are close. Again, we will consider several different possible criteria of "closeness" between two configuration spaces; which, if any, of these is appropriate in a given circumstance depends on the particular application and the question being addressed.

Thus, we can reword the general question posed above: If the shapes of one system approximate those of another system, is the configuration space of the first close to the configuration space of the second, under the various definitions of "approximation" of shape and "closeness" of configuration space? In this paper, we shall prove several theorems that guarantee that a sufficiently precise approximation of shape preserves significant properties of configuration space. In particular, we show that

- It is often possible to guarantee that every configuration in the configuration space of system A are close to a configuration of system B by requiring that the shapes of A closely approximate those of B in terms of the dual-Hausdorff distance. (Section 2)
- It is often possible to guarantee further that every path through the configuration space of system A is close to a path through the configuration space of system B by requiring that the shapes of A closely approximate those of B in terms of the dual-Hausdorff distance. (Section 3)
- It is often possible to guarantee further that each connected component of the configuration space of A is close to a connected component of the configuration space of B by requiring that the shapes of A approximate those of B in tangent. (Section 4)

We also show that, if the nominal shapes are semi-algebraic, then there are algorithms that will compute a positive tolerance satisfying the theorems described above. All these results use a reduction to Tarski's theorem (1951) that the first-order language of algebra over real variables is decidable. It should be noted that the straightforward expression of the tolerance theorems involve quantifying over all regions within tolerance of the nominal shapes, and that quantification over regions is not generally within the scope of Tarski's theorem. Indeed, first-order language over regions, even with very limited predicates, are generally undecidable (Grzegorczyk, 1951). Thus, the existence of a reduction to Tarski's theorem is not a trivial statement.

The immediate practical application of these results is limited for the following reason: The mechanical systems that are used in practice generally involved very tight fits between pieces in order to achieve a system with a small number of degrees of freedom, typically one of freedom. Gears mesh closely, pistons fit tightly in cylinders, and so on. Unfortunately, it is precisely in these circumstances that the results derived here tend not to apply. Nonetheless, the results here have real value. First, kinematic constraints are a critical part of physical reasoning for a wide range of problems, including many systems that are only loosely constrained and have the full space of degrees of freedom. Second, it is encouraging to have attained any results relating physical behavior to shape tolerance, and it may be hoped that these can be extended to results of more practical importance.

Three appendices are attached to the paper. Appendix A summarizes the ten different spaces and associated distance functions used in this paper. Appendix B gives a brief review of the elementary point-set topology used. Appendix C contains the proofs of the theorems in this paper.

1.1 Geometric preliminaries and tolerances

We begin by defining some basic geometric primitives and some definitions of what it means for one shape to approximate another. We use standard two- or three-dimensional Euclidean geometry. A *region* is a subset of Euclidean space.

We write "$\mathrm{d}(\mathbf{p},\mathbf{q})$" to denote the Euclidean distance between points $\mathbf{p}$ and $\mathbf{q}$.

Definition 1.1: The distance between two regions $\mathrm{d}(\mathbf{A},\mathbf{B})$ is defined, as usual, as the minimal distance between the two.

$$\mathrm{d}(\mathbf{A},\mathbf{B}) = \inf_{\mathbf{p}\in\mathbf{A},\mathbf{q}\in\mathbf{B}} \mathrm{d}(\mathbf{p},\mathbf{q})$$

Definition 1.2: Two regions *overlap* if their interiors have a non-empty intersection. The degree of overlap of regions $\mathbf{A}$ and $\mathbf{B}$, denoted "$\mathrm{o}(\mathbf{A},\mathbf{B})$" is the radius of the largest sphere contained in $\mathbf{A}\cap\mathbf{B}$.

Definition 1.3: We define three distance functions on regions. The *Hausdorff* distance from $\mathbf{A}$ to $\mathbf{B}$ is defined as the maximum of either the maximal distance from a point $\mathbf{p}\in\mathbf{A}$ to $\mathbf{B}$ or the maximal distance from a point $\mathbf{q}\in\mathbf{B}$ to $\mathbf{A}$.

$$\mathrm{d}_H(\mathbf{A},\mathbf{B}) = \max(\sup_{q\in A}\inf_{p\in B} \mathrm{d}(\mathbf{p},\mathbf{q}), \sup_{p\in B}\inf_{q\in A} \mathrm{d}(\mathbf{p},\mathbf{q}))$$

The *complement-Hausdorff* distance from $\mathbf{A}$ to $\mathbf{B}$, denoted "$\mathrm{d}_{Hc}(\mathbf{A},\mathbf{B})$", is the Hausdorff distance between the complements of $\mathbf{A}$ and $\mathbf{B}$. The *dual-Hausdorff* distance from $\mathbf{A}$ to $\mathbf{B}$, denoted "$\mathrm{d}_{Hd}(\mathbf{A},\mathbf{B})$", is the maximum of the Hausdorff distance and the complement-Hausdorff distance.

Example 1.1: Let $\mathbf{A}$ be the square with vertices $\langle 1,1\rangle, \langle -1,1\rangle\langle -1,-1\rangle, \langle 1,-1\rangle$ and let $\mathbf{B}$ be the circle centered at the origin of radius 1.2 (Figure 1). Then the distance from the point $\mathbf{a1} = \langle 1,1\rangle$ in $\mathbf{A}$ to the closest point $\mathbf{b1} = \langle 0.6\sqrt{2}, 0.6\sqrt{2}\rangle$ in $\mathbf{B}$ is $\sqrt{2}-1.2 \approx 0.214$. Moreover, this is the greatest distance from any point in $\mathbf{A}$ to the closest point in $\mathbf{B}$. The distance from the point $\mathbf{b2} = \langle 1.2,0\rangle$ in $\mathbf{B}$ to the nearest point $\mathbf{a2} = \langle 1,0\rangle$ in $\mathbf{A}$ is 0.2. Moreover, this is the greatest distance from any point in $\mathbf{B}$ to the nearest point in $\mathbf{A}$. Therefore, the Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$, $\mathrm{d}_H(\mathbf{A},\mathbf{B}) = \max(0.214, 0.2) = 0.214$.

Let $\mathbf{A}^c$ and $\mathbf{B}^c$ be the closures of the complements of $\mathbf{A}$ and $\mathbf{B}$ respectively. Then the complement-Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$ is equal to the Hausdorff distance between $\mathbf{A}^c$ and $\mathbf{B}^c$. (Taking the closure does not affect the Hausdorff distance, and makes it possible to talk about maxima rather than least upper bounds.) The distance from the point $\mathbf{b1} = \langle 0.6\sqrt{2}, 0.6\sqrt{2}\rangle$ in $\mathbf{B}^c$ to the closest point $\mathbf{a3} = \langle 1, 0.6\sqrt{2}\rangle$ in $\mathbf{A}^c$ is equal to $1 - 0.6\sqrt{2} = 0.151$. Moreover, this is the greatest distance from any point in $\mathbf{B}^c$ to $\mathbf{A}^c$. The distance from the point $\mathbf{a2} = \langle 1,0\rangle$ in $\mathbf{A}^c$ to the point $\mathbf{b2} = \langle 1.2,0\rangle$ in $\mathbf{B}^c$ is 0.2. Moreover, this is the greatest distance from any point in $\mathbf{A}^c$ to $\mathbf{B}^c$. Thus, the compliment-Hausdorff distance from $\mathbf{A}$ to $\mathbf{B}$, $\mathrm{d}_{Hc}(\mathbf{A},\mathbf{B}) = \mathrm{d}_H(\mathbf{A}^c,\mathbf{B}^c) = \max(0.151, 0.2) = 0.2$.

The dual-Hausdorff distance from $\mathbf{A}$ to $\mathbf{B}$, $\mathrm{d}_{Hd}(\mathbf{A},\mathbf{B}) = \max(\mathrm{d}_{Hc}(\mathbf{A},\mathbf{B}), \mathrm{d}_H(\mathbf{A},\mathbf{B})) = 0.214$.

Example 1.2: Consider the three regions shown in figure 2. $\mathbf{A}$ is the upper comb-like region, $\mathbf{B}$ is the lower comb-like region, and $\mathbf{R} = \mathbf{A}\cup\mathbf{B}$ is the complete rectangle. Let t be the width of a tine of $\mathbf{A}$ and let h be the height of the rectangle. Let $\mathbf{p}$ be a point on the bottom line in the center of a tine of $\mathbf{B}$. Then the distance from $\mathbf{p}$ to the nearest point in $\mathbf{A}$ is roughly $t/2$. Moreover, no other point of $\mathbf{R}$ is further from $\mathbf{A}$, and, of course, every point of $\mathbf{A}$ is also a point of $\mathbf{R}$ and is therefore distance 0 from a point of $\mathbf{R}$. Hence, the Hausdorff distance from $\mathbf{A}$ to $\mathbf{R}$ is roughly $t/2$.

On the other hand, if $q \in \mathbf{A}^c$ is a point halfway up a tine of $\mathbf{B}$, then the distance from q to $\mathbf{R}^c$ is $h/2$. Therefore, the complement-Hausdorff distance and the dual-Hausdorff distance from $\mathbf{A}$ to $\mathbf{R}$ is $h/2$.

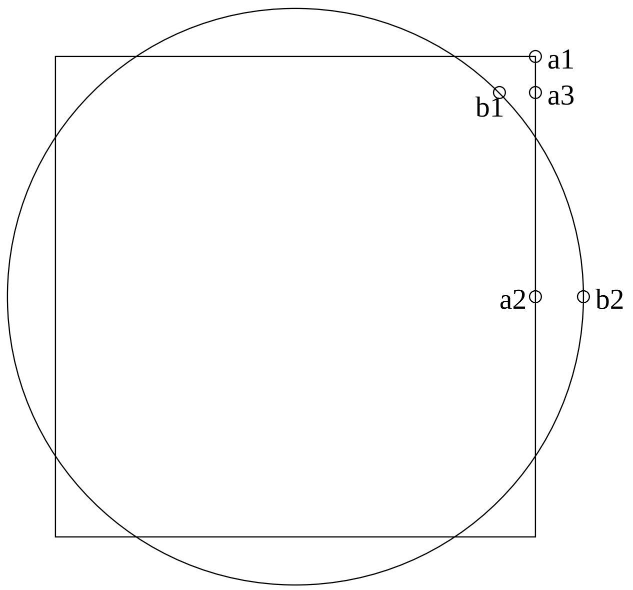


Figure 1: Hausdorff and complement-Hausdorff distances: Example 1.1

The Hausdorff distance is well known in the literature. The dual-Hausdorff distance is original here, as far as we know.

Lemma 1.1. The Hausdorff distance, the complement-Hausdorff distance, and the dual-Hausdorff distance, are all metrics over the space of compact (i.e. closed and bounded) regions.

Proof: Straightforward. ∎

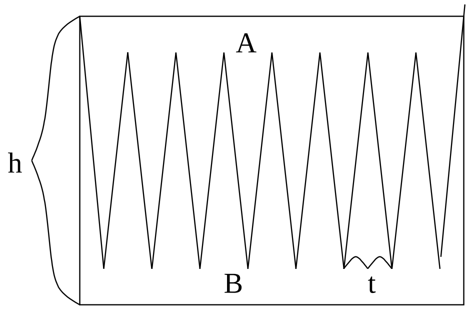
These three metrics define three different topologies on the space of compact regions. The dual-Hausdorff distance is strictly finer than the other two; the Hausdorff and complement-Hausdorff are incomparable. Therefore, if a function is continuous over the topology defined by the Hausdorff distance or over the topology defined by the complement-Hausdorff distance, it is also continuous over the topology defined by the dual-Hausdorff distance.

It can be shown that the distance function "$\mathrm{dist}(\mathbf{A}, \mathbf{B})$" is continuous relative to the Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$ and that the overlap function "$\mathrm{o}(\mathbf{A}, \mathbf{B})$" is continuous relative to the complement-Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$. The overlap function is not a continuous function of regions using the Hausdorff metric. For example, in figure 1, both combs $\mathbf{A}$ and $\mathbf{B}$ are close to $\mathbf{R}$ in the Hausdorff metric, but $\mathrm{o}(\mathbf{A}, \mathbf{B}) = 0$, while $\mathrm{o}(\mathbf{R}, \mathbf{R})$ is large. Note that $\mathbf{A}$ and $\mathbf{R}$ are not close in the complement-Hausdorff metric. The standard distance function between regions is not continuous over the complement-Hausdorff metric; the example is analogous.

Given a region $\mathbf{R}$ and a distance ϵ, lemma 1.2 below gives an elegant and useful characterizations of the regions that contain $\mathbf{R}$ and are within Hausdorff distance ϵ of $\mathbf{R}$; and a similar characterization of the regions that are contained in $\mathbf{R}$ and are within complement-Hausdorff distance ϵ of $\mathbf{R}$. First, we need a couple of definitions:

Definition 1.4: Let $\mathbf{x}$ be a point and $\epsilon > 0$ a distance. The open ball of radius ϵ around $\mathbf{x}$, denoted $\mathrm{B}(\mathbf{x}, \epsilon)$ is defined, as usual, as the set of points $\mathbf{y}$ such that $\mathrm{d}(\mathbf{x}, \mathbf{y}) < \epsilon$.

Definition 1.5: Let $\mathbf{R}$ be a closed region and $\epsilon > 0$ a distance. The *uniform expansion of* $\mathbf{R}$ *by* ϵ,



$\mathbf{R} = \mathbf{A} \cup \mathbf{B}$
The Hausdorff distance from $\mathbf{R}$ to $\mathbf{A}$ is roughly $t/2$.
The complement-Hausdorff distance from $\mathbf{R}$ to $\mathbf{A}$ is $h/2$.

Figure 2: Hausdorff and complement-Hausdorff distances: Example 1.2

denoted "expand($\mathbf{R}, \epsilon$)," is the closure of the union of B($\mathbf{x}$,ϵ) for all $\mathbf{x} \in \mathbf{R}$; equivalently, it is the set of all points $\mathbf{y}$ such that $\mathrm{d}(\mathbf{y}, \mathbf{R}) \leq \epsilon$. The *uniform contraction of* $\mathbf{R}$ *by* ϵ, denoted "contract($\mathbf{R}, \epsilon$)," is the set of all points $\mathbf{y}$ such that B($\mathbf{y}$,ϵ) $\subset \mathbf{R}$. Equivalently, it is the set of all points $\mathbf{y}$ such that $\mathrm{d}(\mathbf{y}, \mathbf{R}^c) \geq \epsilon$.

Lemma 1.2: If $\mathbf{S} \supset \mathbf{R}$, then $\mathrm{d}_H(\mathbf{R}, \mathbf{S}) \leq \epsilon$ if and only if $\mathbf{S} \subset \mathrm{expand}(\mathbf{R}, \epsilon)$. If $\mathbf{S} \subset \mathbf{R}$, then $\mathrm{d}_{Hc}(\mathbf{R}, \mathbf{S}) \leq \epsilon$ if and only if $\mathbf{S} \supset \mathrm{contract}(\mathbf{R}, \epsilon)$.

The proof is immediate from the definitions. Note that the second part of the lemma holds even if the uniform contraction of $\mathbf{R}$ by ϵ is the null set. In that case, the lemma asserts that all subsets of $\mathbf{R}$ lie within ϵ of $\mathbf{R}$ in the complement-Hausdorff distance.

1.2 Rigid mappings and configuration space

We now develop a language for describing the motions of rigid objects under kinematic constraints. We use the standard approach of thinking in terms of a *configuration space*. A *configuration* is a specification of the positions of a collection of objects, and the configuration space is the set of all configurations. A configuration is either *feasible*, if no two objects overlap, or *infeasible* if at least one pair of objects overlaps. The kinematic properties of a given collection of objects are determined by the space of feasible configurations of those objects.

Definition 1.6: A *regular region* is a subset of $\Re^n$ that is non-empty, bounded, and equal to the closure of its interior.

We will assume throughout this paper that every solid object occupies a regular region. In section 4, it will be necessary to add additional constraints on the shape. Curiously, the results in sections 2 and 3 do not depend on the shapes being connected. If it were physically possible to construct an object that was disconnected but nonetheless moved rigidly, these results would still apply.

We assume that we have a fixed collection of objects. With each object, we will associate a regular region that is its shape in some standard position. The region that it occupies in a given situation is then defined by the application of a particular rigid mapping, or *placement*, to its standard shape.

Example 1.3: Consider a square object with side length 2. We could take its standard shape to be the square with vertices $\langle 1, 1\rangle, \langle -1, 1\rangle \langle -1, -1\rangle, \langle 1, -1\rangle$. If it is rotated counter-clockwise by $\pi/6$ around the upper-right corner, then the placement is given by the relation

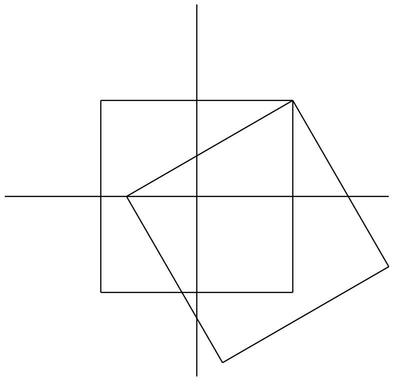


Figure 3: A square under two different mappings

$x' = x\sqrt{3}/2 - y/2 + (3/2 - \sqrt{3}/2).$
$y' = x/2 + y\sqrt{3}/2 + (1/2 - \sqrt{3}/2)$

(Note that $\cos(\pi/6) = \sqrt{3}/2$ and $\sin(\pi/6) = 1/2$.) The placement associated with this new position of the object is then the pair consisting of the original square with this rigid transformation. (Figure 3)

We will need to describe the amount of change in a configuration space as a result of changing. The first step to this is to define a notion of the distance between the effects of mappings M_1 and M_2 on a fixed reference region $\mathbf{R}$. (There is no useful definition of the distance between two rigid mappings M_1 and M_2 in an absolute sense.)

Definition 1.7: Let $\mathbf{R}$ be a compact region and let M_1 and M_2 be rigid mappings. Then the distance between M_1 and M_2 relative to $\mathbf{R}$, denoted "$\mathrm{p}^{\mathbf{R}}(M_1, M_2)$", is defined as the maximal displacement of any point in $\mathbf{R}$ in going from M_1 to M_2.

$$\mathrm{p}^{\mathbf{R}}(M_1, M_2) = \max_{\mathbf{r} \in \mathbf{R}} \mathrm{d}(M_1(\mathbf{r}), M_2(\mathbf{r}))$$

Example 1.4: Let $\mathbf{R}$ be the square used in example 1.3, and let M_1 and M_2 be the two positions shown. Since the change from M_1 to M_2 is a rotation about $\langle 1, 1 \rangle$, the point in $\mathbf{R}$ that moves the most from M_1 to M_2 is the point furthest from $\langle 1, 1 \rangle$, which is the lower left hand corner. This moves from $\langle -1, -1 \rangle$ to $\langle 2 - \sqrt{3}, -\sqrt{3} \rangle$, which is a distance of $2\sqrt{3} - 2 = 1.46$. Thus $\mathrm{p}^{\mathbf{R}}(M_1, M_2) = 1.46$, in this case.

Lemma 1.3: For any fixed regular region $\mathbf{R}$, the function $\mathrm{p}^{\mathbf{R}}(M_1, M_2)$ is a metric on the space of placements.

The proof is straightforward.

1.3 Collections of objects

We now go from describing the state of a single object to describing the state of a collection of objects. We assume that the objects are numbered $1 \ldots k$, so we can describe their shapes as a k-tuple of regions, and their positions as a k-tuple of rigid mappings. Throughout this paper, we will indicate the ith component of a tuple V as $V[i]$, reserving subscripts to distinguish different tuples.

Definition 1.8: A *display* is a k-tuple of regular regions. Intuitively, these are the shapes of k objects.

Definition 1.9: A display D' is a *contraction* of display D if, for $i = 1 \ldots k$, $D'[i] \subseteq D[i]$. D' is an *expansion* of D if D is a contraction of D'.

Definition 1.10: Display D' is the *uniform contraction (expansion)* of display D by distance ϵ if $D'[i]$ is the uniform contraction (expansion) of $D[i]$ by ϵ for each i.

Definition 1.11: A *configuration* is a k-tuple of rigid mappings C. Intuitively, these are the displacements of each object from its standard position as given in a display. The *configuration space on k objects* is the set of all such k-tuples.

Definition 1.12: A *scenario* is a pair of a display and a configuration. If $\langle D, C \rangle$ is a scenario, then, slightly abusing notation, we will write $CD[i]$ for $C[i](D[i])$, the region occupied by the ith object in the scenario.

Definition 1.13 We extend the metrics d_H, d_{Hc}, d_{Hd} to displays by taking the maximum of the function over indices. That is, $\mu(x, y) \equiv \max_i \mu(x[i], y[i])$, where μ is one of the above functions and x and y are displays.

For any display D, we define the metric p^D over configurations as $\mathrm{p}^D(C_1, C_2) = \max_i \mathrm{p}^{D[i]}(C_1[i], C_2[i])$. That is, it is the maximum distance moved by any point of any object in D in going from configuration C_1 to C_2. The *clearance* of a scenario is the minimal distance between places of two different objects in the scenario. The *maximal overlap* of a scenario is the maximal overlap of two different objects in the scenario.

$$\begin{array}{l} \mathrm{clearance}(D, C) = \min_{i \neq j} \mathrm{d}(CD[i], CD[j]) \\ \mathrm{overlap}(D, C) = \max_{i \neq j} \mathrm{o}(CD[i], CD[j]) \end{array}$$

Definition 1.14: A scenario $\langle D, C \rangle$ is *feasible* if no two objects overlap; that is, $\mathrm{overlap}(D, C) = 0$. $\langle D, C \rangle$ is *contact-free* if no two objects are in contact; that is, $\mathrm{clearance}(D, C) > 0$. $\langle D, C \rangle$ is *forbidden* if it is not feasible; that is $\mathrm{overlap}(D, C) > 0$. For any display D, the set of configurations C such that $\langle D, C \rangle$ is feasible is denoted "free(D)"; the set of configurations C such that $\langle D, C \rangle$ is forbidden is denoted "forbidden(D)"; and the set of configurations C such that $\langle D, C \rangle$ is contact-free is denoted "cfree(D)".

The more common tolerance measure over regions in the literature is the *zone-tolerance* measure, defined in (Requicha, 1983) (see also Yap and Chang, 1996). We define the ϵ-zone of **A** to be the region within distance ϵ of the boundary of **A**. Then **B** is within zone-tolerance ϵ of **A** if the boundary of **B** lies within the ϵ-zone of **A**. We define the zone-tolerance measure from **B** to **A** as the maximum ϵ such that **B** lies within the ϵ-zone of **A**, or, equivalently, as the maximum distance from any point in the boundary of **B** to the boundary of **A**. (Note that this is not symmetric in **B** and **A**.) The zone-tolerance measure is often equal to the dual-Hausdorff distance, and never greater, but sometimes much less.

Theorem 1.4: The dual-Hausdorff distance from **B** to **A** is greater than or equal to the zone-tolerance measure from **B** to **A**.

Figure 4 shows a number of cases where the dual-Hausdorff distance from **B** to **A** is much greater than the zone-tolerance measure. Part (I) shows the solid square **A** with a zone around its boundary indicated by dashed lines. Parts II, III, and IV show figures B whose boundary lies entirely in the zone, and hence qualify as being with in the specified zone-tolerance of **A**, even though the Hausdorff distance between the two regions is large. Note that in parts III and IV, it is also the case that **A** is within the same zone tolerance of **B**.

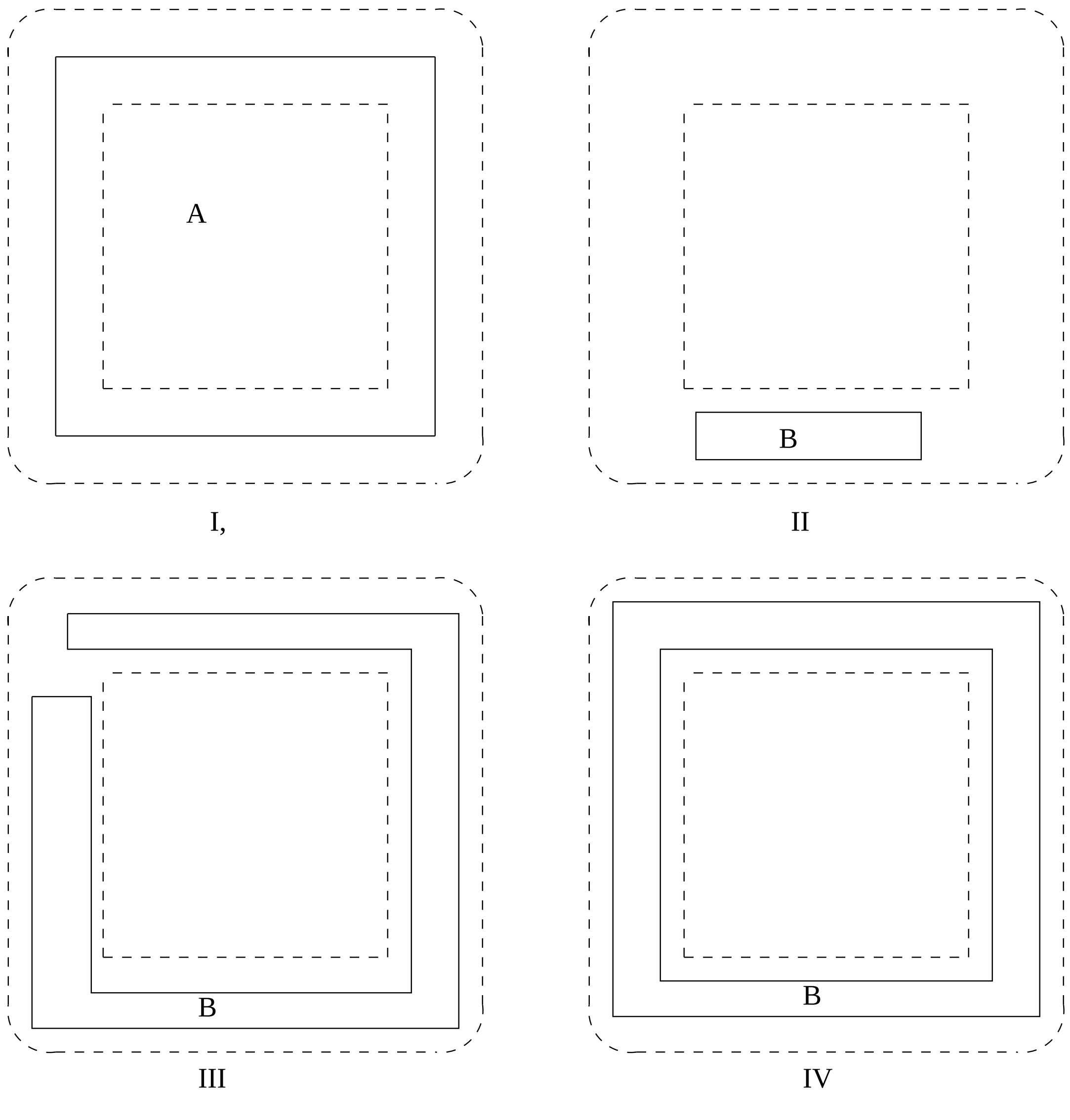


Figure 4: Zone tolerance

2 Approximation of configuration space in the Hausdorff distance

In this section, we present the first of our approximation results. We wish to say that the feasible configuration spaces for two different display are "close". To do this, we must define the distance between two configuration spaces. We achieve this by generalizing the Hausdorff construction to apply to subsets of an arbitrary metric space.

Definition 2.1: Let μ be a metric over space $\mathcal{O}$. Let $\mathcal{S}$ and $\mathcal{T}$ be subsets of $\mathcal{O}$. Then the function $\mu_H(\mathcal{S},\mathcal{T})$ is defined as

$$\mu_H(\mathcal{S},\mathcal{T}) = \max(\sup_{s\in\mathcal{S}} \inf_{t\in\mathcal{T}} \mu(s,t),\ \sup_{t\in\mathcal{T}} \inf_{s\in\mathcal{S}} \mu(s,t))$$

Thus, the domain of μ_H is $2^{\mathcal{O}} \times 2^{\mathcal{O}}$ and the range is the non-negative reals union infinity. It is easily verified that μ_H is a metric over the space of closed subsets of $\mathcal{O}$ (in the extended sense of "metric" that allows infinite values.)

In particular, applying the Hausdorff construction to the metric p^D on configurations gives a metric p^D_H over the space of closed regions in configuration space.

Example 2.1 (Figure 5): Let **A** be the square $[0,1]\times[0,1]$; let **B** be the square $[0,0.8]\times[0,0.8]$; let **C** be the square $[0,0.5]\times[0,0.5]$. Let **Z** be the figure $[0,6]\times[0,6]-[3.0,3.5]\times[3,6]$ a 6×6 square with a notch of width 0.5 cut out of it.

Now, let Q be any feasible configuration over $\langle \mathbf{B},\mathbf{Z}\rangle$. Then the scenario $\langle\langle \mathbf{A},\mathbf{Z}\rangle, Q\rangle$ may well be infeasible. (Changing scenario $\langle\langle \mathbf{B},\mathbf{Z}\rangle, Q\rangle$ to $\langle\langle \mathbf{A},\mathbf{Z}\rangle, Q\rangle$ amounts to expanding the square 0.8×0.8 to 1.0×1.0 while keeping shape of **Z** and the positions of the two objects fixed.) However, it is intuitively clear, and indeed easily proved, that, so to speak, only one side or one corner of **A** can overlap **Z** in Q, and therefore the overlap can be removed by moving **A** and **Z** apart, moving each a distance of at most $0.2\sqrt{2}/2 = 0.14$ (the distance from the top right corner of **A** to the top-right corner of **B**.) Hence, the Hausdorff distance from free($\langle \mathbf{A},\mathbf{Z}\rangle$) to free($\langle \mathbf{B},\mathbf{Z}\rangle$) is 0.14.

On the other hand, consider next the display $\langle \mathbf{C},\mathbf{Z}\rangle$. Let U be the configuration, "Translate the first object by $\langle 3.0, 3.0\rangle$; keep the second object fixed." Then the scenario $\langle\langle \mathbf{C},\mathbf{Z}\rangle, U\rangle$ places **C** at the bottom of the notch of **Z**, and is free. It is clear that U is not at all close to any free configuration for **A**, **Z**. That is, if you start with the scenario $\langle\langle \mathbf{C},\mathbf{Z}\rangle, U\rangle$ and then you expand **C** to **A**, and now you want to move **A** to get rid of the overlap, there is no direction to move **A** in which this is easy. In fact, the closest configuration to U that is feasible over $\langle \mathbf{A},\mathbf{Z}\rangle$ is the configuration V, "Translate the first object by $\langle 4.5, 3.0\rangle$; translate the second by $\langle -1.5, 3.0\rangle$." The distance between configurations U and V is 1.5, and, indeed, this is the Hausdorff distance p^D_H from free($\langle \mathbf{A},\mathbf{Z}\rangle$) to free($\langle \mathbf{C},\mathbf{Z}\rangle$). (Which display D we use as reference here doesn't matter, because in each case we happen to be dealing with translations.)

The function "free(D)" is, in fact, discontinuous under expansion at $\langle \mathbf{C},\mathbf{Z}\rangle$; there are arbitrarily small expansions of **C** or of **Z** in which U and all configurations close to U are infeasible. Not coincidentally, as we shall see in theorem 2.8, the display $\langle \mathbf{C},\mathbf{Z}\rangle$ has the feature that free(D) is not equal to the closure of cfree(D); configuration U is feasible, but there are no nearby configurations that are contact-free.

Theorem 2.1 states that the above example is characteristic of a general rule. If we start with a display, such as $\langle \mathbf{A},\mathbf{Z}\rangle$ and then shrink it by a sufficiently small amount, the Hausdorff distance between the free spaces changes only a small amount. We see this above in shrinking **A** to **B**. If we shrink it too much, however, there may be a large discontinuous change. This happens in the above example when **A** is contracted to **C**. The amount of shrinkage must be measured relative to

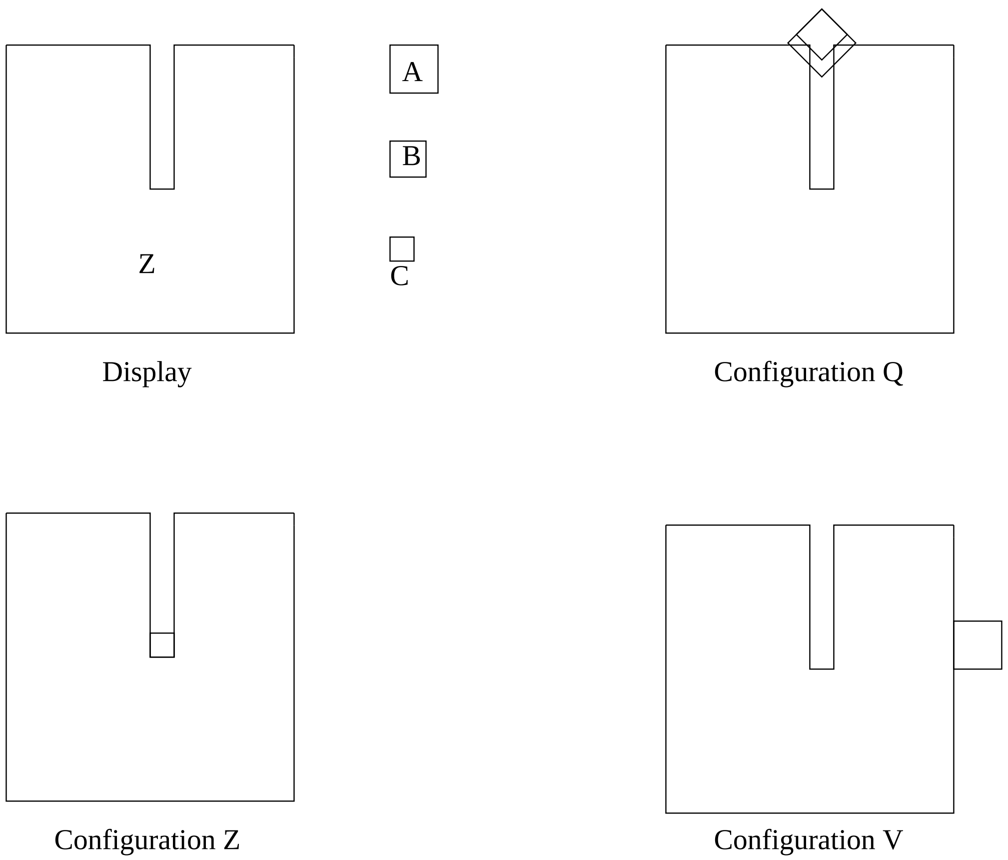


Figure 5: Discontinuity in free(D): Example 2.1

the complement-Hausdorff distance.

Theorem 2.1: Let D be a display. For any distance $\epsilon > 0$ there exists a distance $\delta > 0$ such that the following holds: if D' is a contraction of D and the complement-Hausdorff distance $d_{Hc}(D, D') < \delta$ then the Hausdorff distance between the free spaces $\mathrm{p}_H^D(\mathrm{free}(D), \mathrm{free}(D')) < \epsilon$.

The proofs of theorem 2.1 and all the other theorems in this section are given in appendix A.2.

The above result can be phrased more elegantly with the introduction of some new terminology. First, we observe that, though the metric p^D over configuration space and the metric p_H^D over regions in configuration space depend on the display D, the topologies (over configuration space and over the power set of configuration space, respectively) induced by these metrics are independent of D (Lemma 2.2 below). We can therefore speak of the topology induced by p and the topology induced by p_H, meaning the topologies induced by p^D and by p_H^D for any D. In particular, if f is a function whose range is regions of configuration space, then we will say "f is continuous with respect to p_H" to mean that f is continuous with respect to p_H^D for any display D.

Lemma 2.2: Let D and E be two displays over k objects. Then the two metrics p^D and p^E induce the same topology on configuration space. Also, the two metrics p_H^D and p_H^E over regions of configuration space induce the same topology. We therefore refer to this topology as "p_H", without specifying a reference display.

Next we define the concepts of "continuous under contraction" and "continuous under expansion". (These are analogues of the more common notions of one-sided continuity.)

Definition 2.2: A function $f(D)$ from the space of displays to a topological space $\mathcal{O}$ is *continuous under contraction (expansion)* if the following holds: Let D be a display, and let U be a neighborhood of $f(D)$ in $\mathcal{O}$. Then there exists a neighborhood V of D, such that, for every $D' \in V$, if D' is a contraction (expansion) of D, then $f(D') \in U$. Intuitively, "f is continuous under contraction" means that if you shrink all the shapes in D by a small amount, then you remain close to $f(D)$. "Function f is continuous under expansion" means that if you expand all the shapes in D by a small amount, then you remain close to $f(D)$.

We can now rephrase theorem 2.1 in a more elegant form:

Theorem 2.1 (rephrased): The function "free(D)", mapping a display to a region in configuration space, is continuous under contraction, assuming that the domain is topologized using d_{Hc} and the range is topologized using p_H.

In theorem 2.1, it is possible to compute the value of ϵ given the display D and the distance δ:

Theorem 2.3: Let D be a display and let $\delta > 0$ be a distance. Let $E = \mathrm{contract}(D, \delta)$. Let $\epsilon = \mathrm{p}_H^D(\mathrm{free}(D),\mathrm{free}(E))$. Then D, ϵ, and δ satisfy the conditions of Theorem 2.1; that is, if F is a contraction of D and $\mathrm{d}_{Hc}^D(F, D) < \delta$, then $\mathrm{p}_H^D(\mathrm{free}(D),\mathrm{free}(F)) \leq \epsilon$.

Theorem 2.3 allows you to compute ϵ from δ; that is, if you know how precisely you can manufacture the shape, you can compute how exactly you will match the configuration space. If you want to go in the opposite direction — that is, you are required to achieve a desired precision ϵ in the configuration space, and you want to know how precisely you need to manufacture the object — then you can use a binary search, since ϵ is a monotonically non-decreasing function of δ. Alternatively, if the display is semi-algebraic, theorem 2.4 will show that you can express δ as a semi-algebraic function of ϵ.

Definition 2.3: The *language of real algebra* is the first-order language with equality containing the two functions "plus" and "times" where variables range over the real numbers.

We will use extensively Tarski's (1951) famous theorem that the language of real algebra is decidable.

Definition 2.4: A region $\mathbf{R}$ is semi-algebraic in n-dimensional Euclidean space, relative to a given coordinate system $\mathcal{C}$, if there is an open formula $\Phi(x_1 \ldots x_n)$ in the language of real algebra such that the point with coordinates $x_1 \ldots x_n$ in $\mathcal{C}$ is in $\mathbf{R}$ if and only $\Phi(x_1 \ldots x_n)$. A display D is semi-algebraic with respect to $\mathcal{C}$ if $D[i]$ is semi-algebraic for each i.

Example 2.2. The relation $x \leq y$ is semi-algebraic, as it is defined by the formula, $\exists_Z \, x + z^2 = y$. The unit sphere is semi-algebraic, as it is defined by the formula $x^2 + y^2 + z^2 \leq 1$.

It is easily shown that, if D is semi-algebraic, then all the operations in theorem 2.3 are semi-algebraic and therefore computable, by theorem 2.4.

Theorem 2.4: Assume that display D is semi-algebraic. Then there is an algebraic formula $\Phi(\epsilon, \delta)$ which holds if and only if Theorem 2.1 is satisfied for D, ϵ and δ. Moreover, Φ is easily computable, given the form of D. Hence, by Tarski's theorem, for fixed semi-algebraic D, it is possible to compute δ from ϵ or ϵ from δ.

In practice, it is much easier to use theorem 2.3 to calculate ϵ from δ or vice versa than to use theorem 2.4. The importance of theorem 2.4 is that it is more easily extended to further computability results. For example, suppose that the shapes of D are not given exactly, but that they are characterized in terms of some parameters, which are themselves restricted by some algebraic constraint, such as "Object A is an $h \times w$ rectangle where $w \leq h \leq 1.5w$." Then one can conclude from theorem 2.4 that it is possible to answer questions such as, "Is a shape approximation accurate to within 0.1 sufficient to compute a configuration space approximation accurate to within 0.5 for all possible values of h and w? For some possible values of h and w?" No such conclusion follows from theorem 2.3.

Theorem 2.1 guarantees that, if you have an actual set of shapes D and you want to calculate the free space to accuracy δ, you can achieve this by approximating D with a sufficiently accurate contraction D' and calculating the free space of D'. What it does not guarantee, however, is that if you have found D' and know its accuracy as an approximation of D, that you can then be sure that the calculation of free space is accurate to within δ. In fact, this guarantee cannot be achieved, in general.

Theorem 2.5: There exists a display D and a distance $\epsilon > 0$ with the following property: Let D' be any semi-algebraic display, and let $\delta = \mathrm{d}_{Hc}(D, D')$. Then there exists a display E such that $\mathrm{d}_{Hd}(E, D') < \delta$ but $\mathrm{p}_H^D(\mathrm{free}(D), \mathrm{free}(E)) > \epsilon$.

That is, there are some particular displays D with corresponding fixed values of ϵ that have the following unfortunate property: If we construct any semi-algebraic approximation D' within tolerance δ of D, and we calculate free(D') and we try to estimate how close free(D) must be to free(D'), it *always* appears possible that free(D) could be more than ϵ away from free(D'). Knowing that D is within δ of D' is *never* enough to guarantee that free(D) is close to free(D'). One such display D contains two objects: a helical corkscrew, and an object with a hole that fits the corkscrew.

This may seem paradoxical in view of theorem 2.1, but it is in fact inevitable in a domain where important functions are discontinuous, and where discontinuities appear at values that cannot be exactly represented. Consider the following analogous case: Suppose that you approximate a real numbers by a rational estimate with a rational tolerance. For example, you might approximate $\sqrt{2}$ as "$7/5 \pm 1/50$" or as "$99/70 \pm 1/10000$". Now consider the discontinuous function

$$f(x) = \begin{cases} 2 - x^2 & \text{if } x^2 \leq 2 \\ -1 & \text{otherwise} \end{cases}$$

Thus $f(\sqrt{2}) = 0$. Now, $f(p)$ can be made arbitrarily close to $f(\sqrt{2})$ by choosing p to be less than $\sqrt{2}$ and sufficiently close to it; for example $f(7/5) = 0.04$. However, if the range $p \pm r$ contains $\sqrt{2}$ then it necessarily contains values that are greater than $\sqrt{2}$ for which f will have value -1.

The same kind of thing is going on in theorem 2.5 with calculating configuration space from semi-algebraic approximations with tolerances. It is possible to find semi-algebraic contractions D' of D for which free(D') is arbitrarily close to free(D). However, if you consider a tolerance around D' large enough to include D, it will necessarily include some supersets of D' with much smaller free spaces, as the function "free(D)" is discontinuous under expansion.

A theorem complementary to theorem 2.1 holds for the contact-free space. The computability results are exactly analogous to theorems 2.2 and 2.4.

Theorem 2.6: The function "cfree(D)", mapping a display to a region in configuration space, is continuous under expansion, using the metric d_H on displays and the topology induced by p_H on regions of configuration space.

Theorem 2.7: Let D be a display and let $\delta > 0$ be a distance. Let E =expand(D, δ). Let $\epsilon = \mathrm{p}_H^D$(cfree($D$),cfree($E$)). If F is an expansion of D and $\mathrm{d}_H^D(F, D) < \delta$, then p_H^D(cfree(D),cfree(F)) $\leq \epsilon$.

Lemma 2.8 describes the relationship between the contact-free space and the free space for a display.

Lemma 2.8 For any display D, cfree(D) is the interior of free(D), relative to the topology induced by metric p.

It is not the case, however, that free(D) is always the closure of cfree(D); it may be a superset of the closure of cfree(D). For example, consider the display $\langle \mathbf{C}, \mathbf{Z} \rangle$ from example 2.1. The free space of this display includes configurations where the square **C** is placed in the notch of **Z**. However, such configurations are not "close" to any contact-free configuration of $\langle \mathbf{C}, \mathbf{Z} \rangle$. Rather, they lie along a one-dimensional "spur" of the free space that is eliminated when you take the interior to generate the contact-free space.

Discontinuities appear in the functions "free(D)" and "cfree(D)" exactly for those displays where free(D) is not equal to the closure of cfree(D).

Theorem 2.9: The following three statements are equivalent:

A. The functions "free($\cdot$)" is continuous at display D, using the metric d_{Hd} on displays.

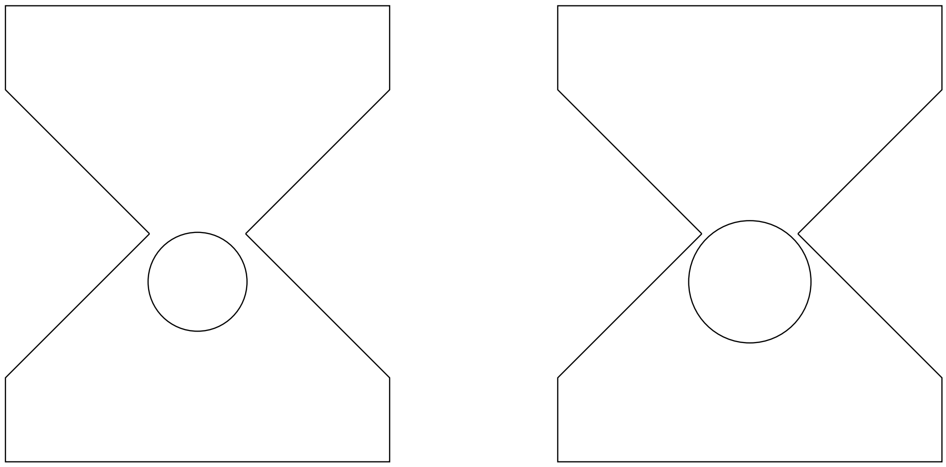
B. The function "cfree($\cdot$)" is continuous at display D, using the metric d_{Hd} on displays.

C. free(D) is equal to the closure of cfree(D).

Much stronger results hold for the forbidden space; the function "forbidden(D)" is always continuous, and δ is equal to ϵ. However, this result is much less interesting, since in practice one is almost always interested in approximating the free space, not the forbidden space.

Theorem 2.10: For any displays D, D', p_H^D(forbidden(D),forbidden(D')) $\leq \mathrm{d}_H(D, D')$. Consequently, the function "forbidden(D)" is continuous, if the topology of the domain is given by d_H and the topology of the range is given by p_H.

Thus every configuration in forbidden(D) is within $\mathrm{d}_H(D, D')$ of some configuration in forbidden(D'), and vice versa, by symmetry. The Hausdorff distance from forbidden(D) to forbidden(D') is therefore at most $\mathrm{d}_H(D, D')$. ∎.



A B

In A, the configurations where the ball is on top and those where it is on the bottom are in a single connected component of free space. In B, they are two separate connected components.

Figure 6: Free space vs. path traces

3 Paths

In the previous section, we considered that two kinematic systems were similar if every feasible configuration in one is close to a feasible configuration in the other. For purposes of physical reasoning, though, it seems natural to say that physical similarity is more a matter of similarity in the paths that can be followed rather than in the single configurations that can be attained. Consider, for example, the system shown in figure 6, and compare the case where the ball is slightly narrower than the gap with the case where the ball is slightly wider. There is only a slight difference in the positions that can be attained. The position where the ball is just halfway through the gap, which is attainable in the former case, is very close to positions where the ball is pressed against the gap, which are attainable in the latter case. However, the motions that are possible in the two systems are very different; in the former, the ball can pass between the top and the bottom; in the latter, it is always stuck in one or the other half.

To characterize this kind of change in feasible motions, we first need a definition of a path and of the distance between two paths. There is more than one possible definition, but the most reasonable, on the whole, seems to be take a *path* through a space to be a continuous function from a standard time interval [0,1] to the space, and then to take the different between paths ϕ and ψ to be just the maximal value over $t \in [0,1]$ of the distance from $\phi(t)$ to $\psi(t)$. In this case, we are interested in paths through configuration space, so the "distance" between points in the paths is given by the metric p^D. We can then use the Hausdorff construction to define the distance between two sets of paths.

Definition 3.1: Let U be a metric space with metric μ. A *path* through U is a continuous function from the real interval [0,1] into U. The set of paths through U is denoted "paths(U)" The distance between paths ϕ and ψ, denoted $\mu_t(\phi,\psi)$ is defined as

$$\mu_t(\phi,\psi) = \max_{t\in[0,1]} \mu(\phi(t),\psi(t))$$

Definition 3.2: Let U be a metric space with metric μ. Let Φ and Θ be two subsets of paths(U). The Hausdorff distance between Φ and Θ is denoted "$\mu_{tH}(\Phi, \Theta)$" is defined, as usual, as

$$\mu_{tH}(\Phi, \Theta) = \max(\sup_{\phi \in \Phi} \inf_{\psi \in \Theta} \mu_t(\phi, \psi), \sup_{\psi \in \Theta} \inf_{\phi \in \Phi} \mu_t(\phi, \psi)).$$

In particular, for sets of paths through the configuration space of display D, we will be interested in the case where metric μ is the metric p^D, so the corresponding distance between sets of paths will be denoted p^D_{tH}.

Example 3.1: In figure 6, let D be the display where the diameter of the ball is just 0.5, and let E be the display where the diameter of the ball is 0.5001. Then paths(free(D)) contains the path ϕ where the ball goes from the lower half of the frame to the upper left hand corner. The path ψ in free(E) that most closely tracks ϕ goes up as close to the gap as it can manage and then stays there. (If ϕ includes in addition some rotations of the ball, ψ will rotate the ball in parallel, to minimize the maximal distance between corresponding points.) The maximal distance in terms of p^D between $\phi(t)$ and $\psi(t)$ is about 1.2, so this is the value of $\mathrm{p}^D_t(\phi, \psi)$, and also the value of p^D_{tH}(paths(free(D)), paths(free(E))).

As example 3.1 illustrates, the function paths(U) is often discontinuous. However, we can show that the function of D, paths(free(D)), is continuous under contraction of D, while the function paths(cfree(D)) is continuous under expansion of D. We do need to add one additional technical constraint on the spaces free(D) and cfree(D), given in definition 3.3.

Definition 3.3: A region O in a topological space is *ordinarily connected* if O has finitely many connected components and every connected component of O is path-connected. A region O is *locally ordinarily connected* if (i) O is ordinarily connected; and (ii) for any point $p \in O$ and any neighborhood N of p, there exists a neighborhod $S \subset N$ of p such that $O \cap S$ is ordinarily connected. That is, $O \cap S$ is ordinarily connected for arbitarily small neighborhoods S of p.

Local ordinary connectivity of the free space is a property that can be reasonably expected of physically realizable objects. It can only fail if either the free space has infinitely many connected components (for example, the objects have infinitely many hooks and eyes) or one of the connected components of the free space is not path-connected, which involves other anomalies (see the discussion at the end of appendix B.) In particular, it can be shown that if the display is semi-algebraic, then the free space is locally ordinarily connected (Mishra, 1993).

Theorem 3.1: (Analogue of theorem 2.1.) Let D be a display such that free(D) is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that, if E is a contraction of D and the complement-Hausdorff distance $\mathrm{d}_{Hc}(D, E) < \delta$ then p^D_{tH}(paths(free(D)), paths(free(E))) $< \epsilon$.

Theorem 3.2: (Analogue of theorem 2.2.) Let D be a display and let $\delta > 0$. Let E = contract(D, δ), and let $\epsilon = \mathrm{p}^D_{tH}$(paths(free($D$)),paths(free($E$))). If F is a contraction of D and $\mathrm{d}_{Hc}(F, D) < \delta$, then p^D_H(free(D),free(F)) $\leq \epsilon$.

Theorem 3.3: (Analogue of theorem 2.3.) Let D be a semi-algebraic display and let $\delta > 0$. Let ϵ be the maximal value of p^D_{tH}(free(D),free(F)) where F is a contraction of D and $\mathrm{d}_{Hc}(F, D) \leq \delta$. Then ϵ can be computed to arbitrary precision.

Note that theorem 3.3 has a weaker form than theorem 2.3. In theorem 2.3, we were able to assert that ϵ and δ were related by a semi-algebraic relation; here, we can only assert that ϵ is computable to arbitrary precision from δ. I conjecture that the following stronger form holds:

> Given any semi-algebraic relations $\Phi(p, q_1 \ldots q_k)$ and $\Psi(s, t_1 \ldots t_m)$, and algebraic function μ, the quantity
>
> $$\mu_{tH}(\text{paths}(\{p \mid \Phi(p, q_1 \ldots q_k)\}), \text{paths}(\{s \mid \Psi(s, t_1 \ldots t_m)\}))$$

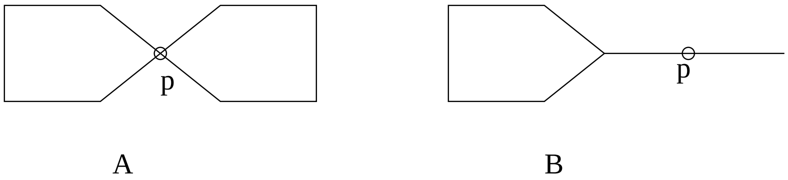


Figure 7: Regions not locally internally connected

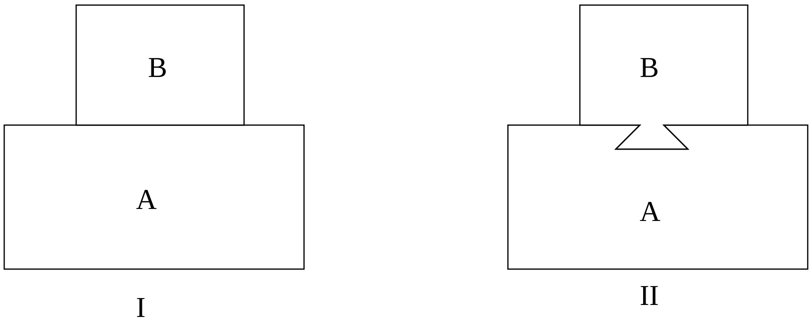


Figure 8: Path tracking vs. correspondence of connected components

is a semi-algebraic function of $q_1 \dots q_m, t_1 \dots t_m$.

Theorem 3.4: (Analogue of theorem 2.4.) Let D be a display such that free(D) is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that, if E is a expansion of D and the Hausdorff distance $\mathrm{d}_H(D, E) < \delta$ then $\mathrm{p}^D(\text{paths}(\text{cfree}(D)), \text{paths}(\text{cfree}(E))) < \epsilon$.

Definition 3.4: A set U is *locally internally connected* if for every point $\mathbf{p}$ in U, and for every neighborhood O of $\mathbf{p}$, $\text{Int}(U) \cap O$ is non-empty and path-connected.

Example 3.2: Figure 7 shows two regions that are not locally internally connected. In figure 7.A, for every sufficiently small neighborhood N of $\mathbf{p}$, $\text{Int}(U) \cap N$ is disconnected. In figure 7.B, for every sufficiently small neighborhood N of $\mathbf{p}$, $\text{Int}(U) \cap N$ is empty.

Theorem 3.5: For any display D, if the configuration space free(D) is locally internally connected, then the function "paths(free($\cdot$))" is continuous at display D, using the metric d_{Hd} on displays and the topology p_{tH} on sets of paths.

4 Lifting from contact and the approximation of tangents

As figure 8 illustrates, it is possible to change a system where all feasible configurations are connected to one that contains configurations that are "stuck" using changes that can be arbitrarily small in both the Hausdorff and complement-Hausdorff distance, and that can be either expansions or contractions. The kinds of criteria that we have used in the previous sections thus do not suffice to prevent this kind of change in the free space.

In terms of configuration space, what happens is that a new small "island" of free space is created in scenario II, close to, but disconnected from, the main region of free space. If scenario II is a contraction of scenario I, then the new island is created inside the forbidden space of I; otherwise, it may be created by having a small pocket of forbidden space close around the island. Note that this is consistent with the consequences of theorems 2.1 and 3.1; the new free space still lies close to the old space, and every path through the new free space can be closely tracked by a path through the old space and vice versa.

Fortunately, figure 4 also suggests where a solution might lie. The construction of the "hook and eye" that so radically changes the behavior of the object requires only a small variation in terms of the points occupied by the object, but requires a large change in the local surface normal. This suggests that if the approximation is required to be close both in distance and in the direction of the local surface normal, then it may be possible to infer that a small modification does not generate new connected components in configuration space.

This is indeed possible, at least two dimensions. In this section, we will state the following result: Given a display consisting of two-dimensional objects such that in every configuration it is possible to separate the objects by lifting or twisting one off the other, then, with certain exceptions, if the objects are modified by a sufficiently small amount with with sufficiently small change in boundary direction, then the connected components of the free space remain essentially unchanged. Connected components do not appear, disappear, merge, or split.

Constructing the formal statement of this theorem will take some work, so we begin by sketching the general direction we will go. Our target is theorem 4.2, which has a form analogous to theorems 2.9 and 3.5. These theorems all have the following general form: "If the objects in display D *have reasonable shapes* and free(D) *satisfies certain normality conditions* and display E approximates D sufficiently well *in a suitable measure of shape approximation* then free(E) is close to free(D) *under some desired criteria for closeness of regions of configuration space.*"[2] In theorems 2.9 and 3.5, the condition "reasonable shapes" is interpreted as "normal bounded regions" and the shape approximation is the dual-Hausdorff distance d_{Hd}. In theorem 2.9, the normality condition on free(D) is that it is equal to the closure of cfree(D), and the approximation measure on regions of configuration space is p^D_H. In theorem 3.5, the normality condition on free(D) is that it is locally internally connected, and the approximation measure on regions of configuration space is p^D_{tH}(paths(A), paths(B)).

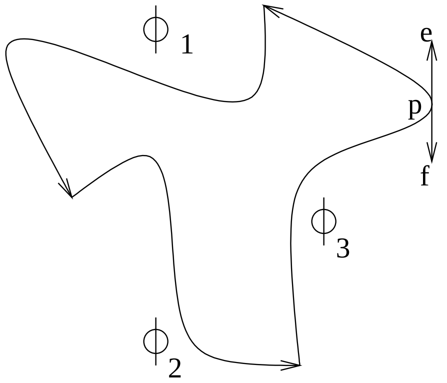
In theorem 4.2, the condition "reasonable shapes" is interpreted as *piecewise, smooth, cuspless (PSC) regions* (definition 4.6); roughly, that the boundary of the regions be smooth except at finitely many corners. The approximation criterion on shapes is *approximation in tangent* (definition 4.8); roughly, every boundary point in one shape is close to a boundary point in the other shape where the tangents are nearly the same. The normality condition on free(D) is that free(D) is *always strongly separable* (definition 4.16); I have not found a simple way to describe this condition that is close to correct. Finally, the approximation measure on regions of configuration space is *connected-component similarity in the space of paths* (definition 4.17). Essentially, this condition ensure that the tracking distance p^D_{tH} is small, and in addition that no islands of free space are opened up.

4.1 A generalization of the tangent

In this section, we develop a characterization of the local behavior of a region that generalizes the notion of the tangent for regions with corners of various kinds.

First, some notation. We denote the complement of regions $\mathbf{R}$ as $\mathbf{R}^c$, the boundary of region $\mathbf{R}$ as "Bd($\mathbf{R}$)", the interior of $\mathbf{R}$ as "Int($\mathbf{R}$)", and the exterior of $\mathbf{R}$ (i.e. the interior of $\mathbf{R}^c$) as

[2] Adding the condition that E be a contraction, or an expansion, does not seem to buy anything in this setting; hence there are no analogues here of theorems 2.1 or 3.1.



e and f are the forward and backward tangents at $\mathbf{p}$.

Figure 9: Piecewise smooth curve (definition 4.3)

"Ext($\mathbf{R}$)". For any non-zero vector $\vec{v}$, we will write "dir($\vec{v}$)" to mean the unit vector parallel to $\vec{v}$; $\mathrm{dir}(\vec{v}) = \vec{v}/\mid \vec{v} \mid$. We will overload this symbol by also writing "dir($\mathbf{a}, \mathbf{b}$)" for $\mathrm{dir}(\mathbf{b} - \mathbf{a})$, the direction from $\mathbf{a}$ to $\mathbf{b}$. We will use interval-like notation to denote arcs of directions: If $\hat{e}$ and $\hat{f}$ are unit vectors, then $(\hat{e}, \hat{f})$ denotes the open arc of vectors strictly counterclockwise between $\hat{e}$ and $\hat{f}$ and $[\hat{e}, \hat{f}]$ denotes the closed arc of vectors non-strictly counterclockwise between $\hat{e}$ and $\hat{f}$.

Definition 4.1: A *finite smooth curve* is a continuous function $\phi(t)$ from the unit interval [0,1] to the plane such that the derivative $\dot{\phi}(t)$ everywhere exists, is continuous, and is non-zero. Therefore, the tangent $\mathrm{dir}(\dot{\phi}(t))$ exists, is unique, and is continuous everywhere.

Definition 4.2: A *directed cycle* is a sequence of continuous curves $\langle \phi_1, \phi_2 \ldots \phi_k \rangle$ such that

- For $i = 1 \ldots k-1$, $\phi_i(1) = \phi_{i+1}(0)$.
- $\phi_k(1) = \phi_1(0)$.
- For any i, j, and t_1, t_2, if $\phi_i(t_1) = \phi_j(t_2)$ then either $[i = j$ and $t_1 = t_2]$ or $[j = i+1 \bmod k, t_1 = 1, t_2 = 0]$ or $[i = j+1 \bmod k, t_2 = 1, t_1 = 0]$.

That is, the end of each curve is the beginning of the next, and the end of the last is the beginning of the first, and the curves do not otherwise cross or meet one another.

Definition 4.3 A directed cycle $\phi = \langle \phi_1, \phi_2 \ldots \phi_k \rangle$ is *piecewise smooth* if each ϕ_i is a finite smooth curve. (Figure 9)

Definition 4.4: Let $\phi = \langle \phi_1, \phi_2 \ldots \phi_k \rangle$ be a piecewise smooth directed cycle, and let $\mathbf{p}$ be a point on ϕ. The *forward* and *backward tangents* to ϕ at $\mathbf{p}$, denoted "forw($\phi, \mathbf{p}$)" and "back($\phi, \mathbf{p}$)" are defined as follows:

- If $\mathbf{p} = \phi_i(1)$ for some i, then, letting $j = i+1 \bmod k$, $\mathrm{forw}(\phi, \mathbf{p}) = \mathrm{dir}(\dot{\phi}_j(0))$ and $\mathrm{back}(\phi, \mathbf{p}) = -\mathrm{dir}(\dot{\phi}_i(1))$.
- Otherwise if $\mathbf{p} = \phi_i(t)$ for $t \neq 0$, $t \neq 1$, then $\mathrm{forw}(\phi, \mathbf{p}) = \mathrm{dir}(\dot{\phi}_i(t))$ and $\mathrm{back}(\phi, \mathbf{p}) = -\mathrm{dir}(\dot{\phi}_i(t))$.

Definition 4.5: A piecewise smooth directed cycle has a *cusp* at point $\mathbf{x}$ if $\mathrm{forw}(\phi, \mathbf{x}) = \mathrm{back}(\phi, \mathbf{x})$ (Figure 10).

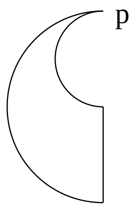


Figure 10: Curve with a cusp

Cusps turn out to be problematic, and we will exclude from consideration regions that give rise to them.

We will add a new requirement on the shape of an objects; namely, that its interior is connected. This excludes regions that hang together on a single point, like figure 11.A. Physically, this is clearly a plausible restriction. Regions are allowed to meet themselves at a point, as in figure 11.B, as long as their interior is connected from some other direction. What we particularly wish to exclude is the case where two objects "pass through" one another at a point **p**, as in figure 11.C, which is clearly physically impossible. A region whose interior is connected will be said to be *internally connected.*

Definition 4.6: Let **A** be an internally connected region. A *boundary cycle of* **A** is is the boundary of a connected component of $\mathbf{A}^c$, directed so that, moving forward on the boundary, the region **A** is always on the left.

Clearly, a boundary cycle of **A** is a subset of the boundary of **A**. The boundary of any internally connected region **A** consists of one boundary cycle counterclockwise around the outside of **A** and a number of boundary cycles clockwise around internal holes of **A**. Two boundary cycles are either disjoint or meet at a single point. (Figure 12)

Definition 4.7: An regular internally connected region **A** is *piecewise smooth and cuspless (PSC)* if every boundary cycle of **A** is piecewise smooth and has no cusps.

4.2 Approximation in tangent

We now define a new notion of approximation of one region by another, which is sensitive to the tangent directions.

Definition 4.8: Let ϕ and ψ be piecewise smooth directed cycles; let $\epsilon > 0$ be a distance and let $\alpha > 0$ be a dimensionless number. For any points **p** on ϕ and **q** on ψ, we say that **p** on ϕ corresponds to **q** on ψ with parameters (ϵ, α) if

- $\mathrm{d}(\mathbf{p}, \mathbf{q}) \leq \epsilon$;
- $\mathrm{d}(\mathrm{forw}(\phi, \mathbf{p}),\ \mathrm{forw}(\psi, \mathbf{q})) \leq \alpha$; and
- $\mathrm{d}(\mathrm{back}(\phi, \mathbf{p}),\ \mathrm{back}(\psi, \mathbf{q})) \leq \alpha$.

We say that ψ *approximates* ϕ *in tangent* (ϵ, α) if, for every point **p** on ϕ there is a point **q** on ψ that corresponds (ϵ, α) and if, for every point **q** on ψ there is a point **p** on ϕ that corresponds (ϵ, α).

Definition 4.9 (Figure 13:) Let **A** and **B** be PSC regions. Let $\epsilon, \alpha > 0$. **B** *approximates* **A** *in tangent* (ϵ, α) if

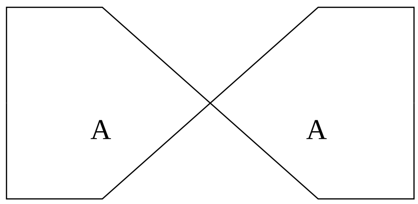


Figure A connectes to itself only at a point, and is not allowed.

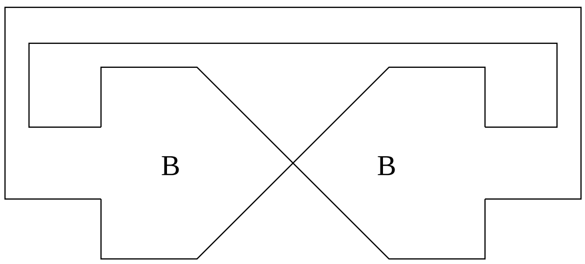


Figure B meets itself at a point, but is internally connected, so it is allowed.

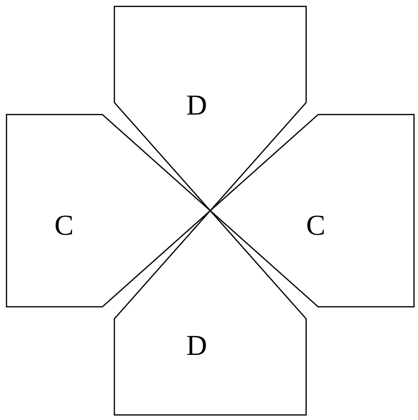
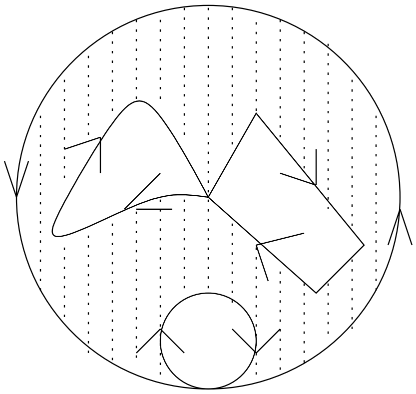


Figure C and D pass through one another at a point. They are not allowed.

Figure 11: Internally connected figures



The shaded region is a PSC region.
The large circle is the outer boundary.
The small circle, the quadrilateral right of center, and the irregular curve left of center are the boundary curves, directed as shown by the arrows.

Figure 12: PSC region with boundary curves

- $\mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B}) \leq \epsilon$;
- For every boundary cycle ϕ of $\mathbf{A}$ there exists a boundary cycle ψ of $\mathbf{B}$, such that ψ approximates ϕ in tangent (ϵ, α); and
- For every boundary cycle ψ of $\mathbf{B}$ there exists a boundary cycle ϕ of $\mathbf{A}$, such that ψ approximates ϕ in tangent (ϵ, α).

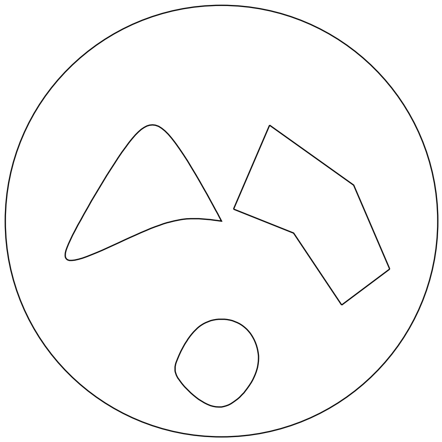
We will sometimes omit the words "in tangent" and simply write "$\mathbf{B}$ approximates $\mathbf{A}$ (ϵ, α)."

Example 4.1: Let $\mathbf{A}$ be the solid unit disc and let $\mathbf{B}$ be the inscribed solid regular n-gon. Then $\mathbf{B}$ approximates $\mathbf{A}$ with parameters $\epsilon = 1 - \cos(\pi/n), \alpha = 2\sin(\pi/2n)$. Proof: Let ϕ and ψ be the directed boundaries around $\mathbf{A}$ and $\mathbf{B}$. Associate a point $\mathbf{a}$ on ϕ with a point $\mathbf{b}$ on ψ if the line $\mathbf{b}$ lies on the radius $\mathbf{o}, \mathbf{a}$, where $\mathbf{o}$ is the center of the circle. The maximum distance from $\mathbf{b}$ to $\mathbf{a}$ occurs when $\mathbf{b}$ is a midpoint of one of the sides, and is equal to $1 - \sqrt{3}/2$. The maximum distance between $\mathrm{forw}(\phi, \mathbf{a})$ and $\mathrm{forw}(\psi, \mathbf{b})$, and between $\mathrm{back}(\phi, \mathbf{a})$ and $\mathrm{back}(\psi, \mathbf{b})$ occurs when $\mathbf{b}$ is a vertex of the polygon, where it is equal to the distance from the direction of the side of the polygon to the tangent of the circle. As the two vectors differ by an angle of π/n, the distance between them is $2\sin(\pi/2n)$. (Figure 14)

Example 4.2:. Let $\mathbf{A}$ be the unit square, and let $\mathbf{B}$ be the rectangle with vertices $\langle 0, 0 \rangle$, $\langle 1.1, 0 \rangle$, $\langle 1.1, 0.9 \rangle$, $\langle 0.0, 0.9 \rangle$. Then $\mathbf{B}$ approximates $\mathbf{A}$ with parameters $\epsilon = 0.1\sqrt{2}$, $\alpha = 0$. Proof: Associate each point $\langle x, y \rangle$ in $\mathbf{A}$ with the point $\langle 1.1x, 0.9y \rangle$ in $\mathbf{B}$. Then corresponding points have exactly the same tangents; the maximal distance between corresponding points is $0.1\sqrt{2}$.

In the same example, it is also true that $\mathbf{B}$ approximates $\mathbf{A}$ with parameters $\epsilon = 0.1, \alpha = \sqrt{2}$.. Proof: Associate any point $\mathbf{a} \in \mathrm{Bd}(\mathbf{A})$ with the closest point in $\mathrm{Bd}(\mathbf{B})$ and associate any point $\mathbf{b} \in \mathrm{Bd}(\mathbf{B})$ with the closest point in $\mathrm{Bd}(\mathbf{A})$. The maximum distance from each point to its associate point is then 0.1. The maximum difference between the tangents occurs at the mapping of right-angle vertices to edge points and vice versa, where one of the tangents of the straight edge lies 90 degrees away from the nearer side of the right-angle. The same difference is obtained at points where a point on a horizontal side is mapped into a point on a vertical side, and vice versa.

This example illustrates that, in general, there is a tradeoff between ϵ and α; if you look further



This region closely approximates figure 12 in tangent.

Figure 13: Approximation in tangent

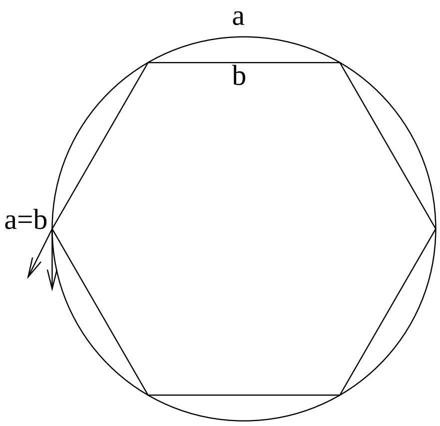


Figure 14: Example 4.1

off for a match, you may be able to find a closer directional match. The extremes of this tradeoff are given in the following examples:

Example 4.3: Let **A** and **B** be any two PSC regions. Let ϵ_1 be the maximum over all boundary curves ϕ of **A** of the minimum over all boundary curves ψ of **B** of $\mathrm{d}_H(\phi, \psi)$; and let ϵ_2 be the maximum over all boundary curves ψ of **B** of the minimum over all boundary curves ϕ of **A** of $\mathrm{d}_H(\phi, \psi)$. Let $\epsilon = \max(\epsilon_1, \epsilon_2, \mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B}))$. Then **A** approximates **B** in tangent with parameters $(\epsilon, 2)$ Proof: Associate each boundary curve of **A** with the nearest boundary curve of **B** and vice versa, and associate each point on a boundary curve with the nearest point on the corresponding boundary curve. Then the distance between corresponding points is no greater than the above value of ϵ. The difference between the tangents at the two associated points can never be greater than 2, the diameter of the unit circle.

Example 4.4: Let **A** and **B** be PSC regions with smooth boundaries, and let D be the maximum value of $\mathrm{dist}(\mathbf{a}, \mathbf{b})$ for $\mathbf{a} \in \mathbf{A}$, $\mathbf{b} \in \mathbf{B}$. Then **A** approximates **B** in tangent with parameters $\epsilon = D$, $\alpha = 0$. Proof: Since each boundary curve ϕ is smooth, the forward tangent to ϕ achieves every directions $\hat{u}$ in the unit circle, and $\mathrm{back}(\phi, \mathbf{p})$ is always equal to $\mathrm{forw}(\phi, \mathbf{p})$. Thus with any point **a** on $\mathrm{Bd}(\mathbf{A})$ we can associate a point **b** on $\mathrm{Bd}(\mathbf{B})$ such that the tangents to **B** at **b** are equal to the tangents to **A** at **a**.

Example 4.5 Let **A** be the solid circle of radius 1, and let **B** be the ring between the circle of radius 0.5 and the circle of radius 1. Of course, the outer boundaries of the two regions are the same, and so can be associated with one another with zero discrepancy. The question is how to associate the inner boundary ψ of **B** with the outer boundary of ϕ of **A**. One way, as in example 4.3, is to associate each point in ψ with the nearest point in ϕ. The distance between corresponding points is 0.5; since the tangents are anti-parallel, the distance betwen the tangents is 2. Thus, **B** approximates **A** in tangent $(0.5, 2)$. Another approach, as in example 4.4, is to associate each point in ψ with the antipodal point. Then the distance between corresponding points is 1.5, but the tangents are parallel, so **B** approximates **A** in tangent (1.5, 0.0). There is also a continuum of intermediate solution where we associate a point **p** in ψ with a point **q** in ϕ such that the angle $\mathbf{poq} = \alpha$, where **o** is the center of the circles and α is any angle between 0 and π. Using this, **B** approximates **A** in tangent $(\sqrt{1.25 - \cos(\alpha)}, \sqrt{2 + 2\cos(\alpha)})$.

4.3 Local separability of regions

We next define a concept of two regions that meet at a point **p** being *separable* in a direction $\hat{u}$ in the neighborhood of **p**.

Definition 4.10: Let $\mathbf{A}, \mathbf{B}$ be non-overlapping PSC regions, and let **p** be a point in $\mathbf{A} \cap \mathbf{B}$. A direction $\hat{u}$ is a *colliding* direction from **A** into **B** at **p**, if, in every neighborhood **O** of **p** there exist $\mathbf{a} \in \mathrm{Int}(\mathbf{A}) \cap \mathbf{O}$ and $\mathbf{b} \in \mathrm{Int}(\mathbf{B}) \cap \mathbf{O}$ such that $\hat{u} = \mathrm{dir}(\mathbf{b} - \mathbf{a})$. Intuitively, if you move **A** in the direction $\hat{u}$ and hold **B** still, then you will cause them to overlap in the neighborhood of **p**.

Definition 4.11: A direction $\hat{w}$ *separates* **B** from **A** at **p** if

- $\hat{w}$ is not a colliding direction from **B** into **A** at **p**; and
- $-\hat{w}$ is a colliding direction from **B** into **A** at **p**.

Direction $\hat{w}$ *strongly separates* **B** from **A** at **p** if $\hat{w}$ is in the interior of an arc of directions all of which separate **B** from **A**.

That is, moving **B** in direction $\hat{w}$ separates it from **A**, and moving it in the opposite direction $-\hat{w}$ causes it to collide with **A**. Moreover, the same holds true in some range of angles around $\hat{u}$.

We denote the set of vectors that strongly separate $\mathbf{B}$ from $\mathbf{A}$ at $\mathbf{p}$ as "sep($\mathbf{B}, \mathbf{A}, \mathbf{p}$)".

Definitions 4.12 and 4.13 and lemma 4.1 give a simple characterization of the directions that separate $\mathbf{B}$ from $\mathbf{A}$ at any contact point $\mathbf{p}$ in terms of the tangents to $\mathbf{A}$ and $\mathbf{B}$ at $\mathbf{p}$.

Definition 4.12: Let $\mathbf{A}$ and $\mathbf{B}$ be PSC regions that do not overlap. The *boundary of* $\mathbf{A}$ *facing* $\mathbf{B}$, denoted "FBd($\mathbf{A}, \mathbf{B}$)" is defined as follows. Let $\mathbf{O}$ be the connected component of $\mathbf{A}^c$ that contains $\mathbf{B}$. Then FBd($\mathbf{A}, \mathbf{B}$) is the boundary cycle of $\mathbf{A}$ that lies on the boundary of $\mathbf{O}$.

Definition 4.13 (Figure 15): Let $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ be four directions in non-strict counterclockwise order, such that $\hat{e} \neq \hat{f}$ and $\hat{g} \neq \hat{h}$. The function $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h})$ is defined as follows:

i. If the angle from $\hat{e}$ to $\hat{h}$ is less than or equal to π, then $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = (\hat{g}, -\hat{f})$.

ii. If the angle from $\hat{g}$ to $\hat{f}$ is less than or equal to π, then $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = (-\hat{e}, \hat{h})$.

iii. If the angles from $\hat{e}$ to $\hat{f}$, from $\hat{f}$ to $\hat{g}$, from $\hat{g}$ to $\hat{h}$ and from $\hat{h}$ to $\hat{e}$ are all less than or equal to π, then $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = (\min(-\hat{e}, \hat{g}), \max(-\hat{f}, \hat{h}))$, where min and max are in the sense of the smaller positive rotation between the two.

iv. If the angle from $\hat{h}$ to $\hat{g}$ is less than or equal to π, then $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = (-\hat{h}, -\hat{g})$.

v. If the angle from $\hat{f}$ to $\hat{e}$ is less than or equal to π, then $\text{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = (\hat{f}, \hat{e})$.

It is easily verified that in figures where two or more of these vectors are exactly angle π apart, so that more than one rule applies, the different rules give the same results.

Lemma 4.1: Let $\mathbf{A}, \mathbf{B}$ be regions that meet but do not overlap, and let $\mathbf{p}$ be a point in $\text{Bd}(\mathbf{A}) \cap \text{Bd}(\mathbf{B})$. Let $\phi = \text{FBd}(\mathbf{A}, \mathbf{B})$ and let $\psi = \text{FBd}(\mathbf{B}, \mathbf{A})$. Then
$\text{sep}(\mathbf{B}, \mathbf{A}, \mathbf{p}) = \text{sep1}(\text{forw}(\phi, \mathbf{p}), \text{back}(\phi, \mathbf{p}), \text{forw}(\psi, \mathbf{p}), \text{back}(\psi, \mathbf{p}))$.

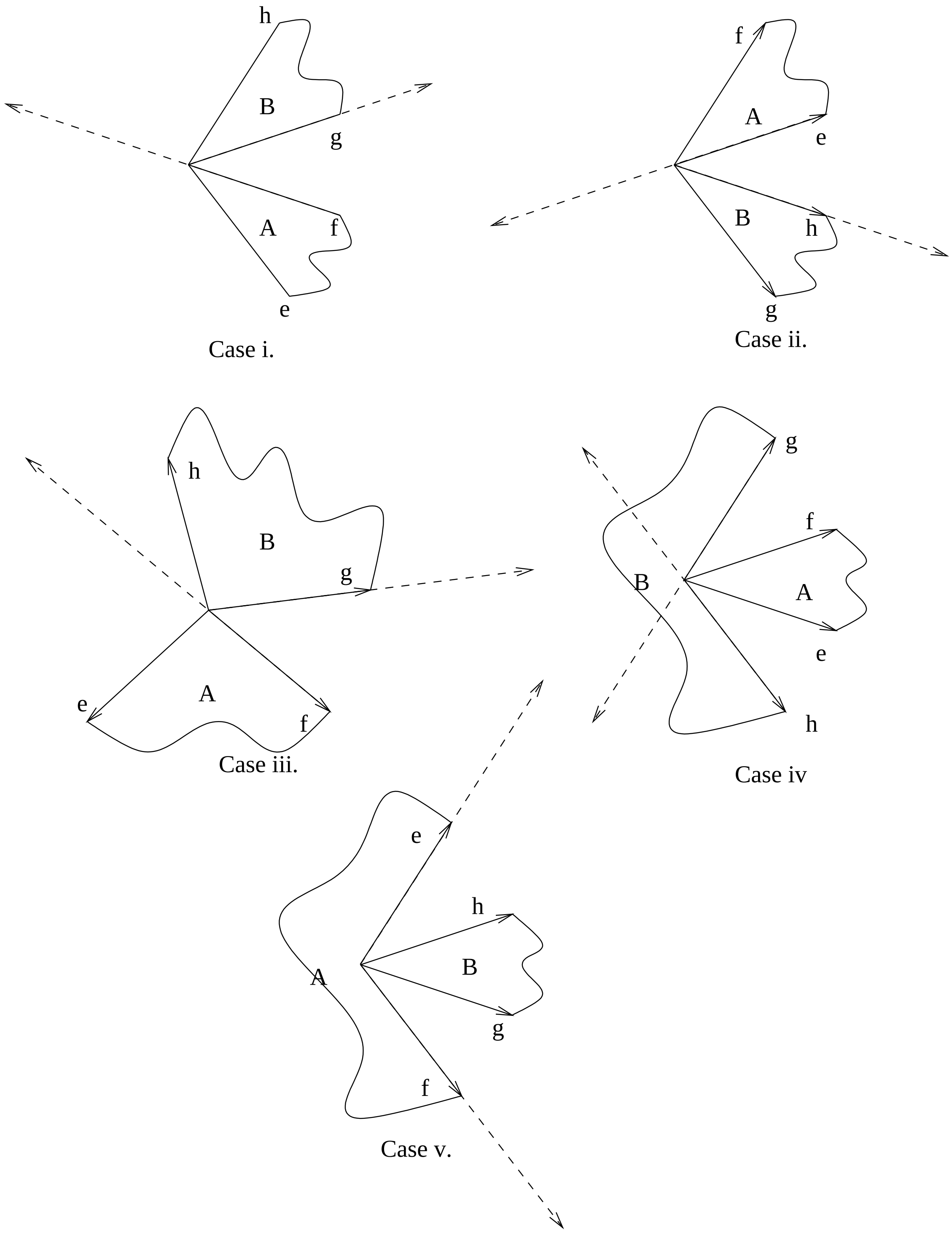
Definition 4.14: A *motion* is either (a) a unit vector $\hat{u}$, representing translation in the $\hat{u}$ direction; or (b) a pair $\langle \mathbf{o}, S \rangle$ of a point $\mathbf{o}$ and a sign $S = \pm 1$ representing counter-clockwise (if $S = 1$) or clockwise (if $S = -1$) rotation about the point $\mathbf{o}$.

Definition 4.15: The *flow* from point $\mathbf{p}$ to motion M, denoted "flow($\mathbf{p}, M$)" is the direction of motion of point $\mathbf{p}$ under motion M. It is defined as follows:

- If $M = \hat{v}$ then $\text{flow}(\mathbf{p}, M) = \hat{v}$.
- If $M = \langle \mathbf{o}, S \rangle$ and $\mathbf{p} = \mathbf{o}$, then $\text{flow}(\mathbf{p}, M) = \vec{0}$.
- If $M = \langle \mathbf{o}, 1 \rangle$ and $\mathbf{p} \neq \mathbf{o}$, then flow($\mathbf{p}, M$) is the unit vector perpendicular to dir($\mathbf{o}, \mathbf{p}$) and counter-clockwise from it.
- If $M = \langle \mathbf{o}, -1 \rangle$ and $\mathbf{p} \neq \mathbf{o}$, then flow($\mathbf{p}, M$) is the unit vector perpendicular to dir($\mathbf{o}, \mathbf{p}$) and clockwise from it.

Definition 4.16: Let $\mathbf{A}$ and $\mathbf{B}$ be two non-overlapping PSC regions. Motion M strongly separates $\mathbf{B}$ from $\mathbf{A}$ if, for every point $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$, $\text{flow}(\mathbf{p}, M) \in \text{sep}(\mathbf{B}, \mathbf{A}, \mathbf{p})$.

Definition 4.17: A display D over two objects A and B is *always strongly separable* if D[A] and D[B] are both PSC regions, and, for every any feasible configuration C over D, there exists a motion M that strongly separates CD[B] from CD[A].



Dashed lines indicate the range of directions separating B from A.

Figure 15: Separating directions

4.3.1 Connected-component similarity

Finally, we define the desired similarity measure between configuration spaces.

Definition 4.18: Let A and B be two regions in a space with metric μ. Let $\mathcal{A}$ be the set of connected components of A and let $\mathcal{B}$ be the set of connected components of B. For any connected component C of A, $C \in \mathcal{A}$ and any connected component D of A, $D \in \mathcal{B}$, measure the distance between C and D as $\mu_H(C, D)$. Then define the *connected-component similarity* between A and B, denoted "$\mu_S(A, B)$" as the Hausdorff distance between $\mathcal{A}$ and $\mathcal{B}$, relative to the metric μ_H. That is,

$$\mu_S(A, B) = (\mu_H)_H(\mathcal{A}, \mathcal{B}) = \max(\max_{C \in \mathcal{A}} \min_{D \in \mathcal{B}} \mu_H(C, D), \max_{D \in \mathcal{B}} \min_{C \in \mathcal{A}} \mu_H(C, D))$$

For theorem 4.2, we will take the space to be the space of paths through configuration spaces C and E, and the metric μ to be the tracking distance p^D_t. Then the measure we wish to constrain is the connected-component similarity between paths(C) and paths(E), p^D_{tS}(paths(D), paths(E)).

For example, in a situation like that of figure 8, the free space in A has only one connected component, while the free space in B has two, the set of all "stuck" positions and the set of all "unstuck" positions. The connected components of the sets of paths through these free spaces correspond to these. The set paths(free(A)) has one connected components, containing all feasible paths. The set paths(free(B)) has two connected components: the paths in which the two objects are stuck, and the paths in which they are unstuck. The connected-component similarily between paths(A) and paths(B) is thus equal to the Hausdorff tracking distance p^D_{tH} between the set of stuck paths in B and the feasible paths in A, which is infinite, as there are paths through free(A) that go artitrarily far from the stuck position. Hence, p^D_{tS}(paths(A), paths(B)) $= \infty$. (In this example, it is also the case that $\mathrm{p}^D_{tS}(A, B) = \infty$. In general μ_S(paths(X),paths(Y)) $\geq \mu_S(X, Y)$.)

4.3.2 Theorems

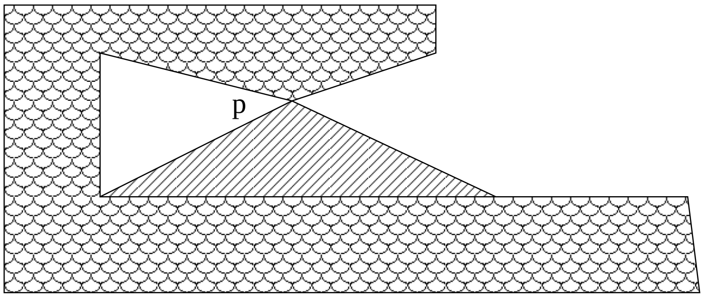
We can now state our main theorem:

Theorem 4.2: Let D be a display over two objects A and B that is always strongly separable, such that free(D) is locally ordinarily connected. Then for any $\eta > 0$ there exist $\alpha > 0$ and $\psi > 0$ such that, if D' is a display that approximates D in tangent (α, ψ), then $\mathrm{p}^D_{tS}(D, D') < \eta$.

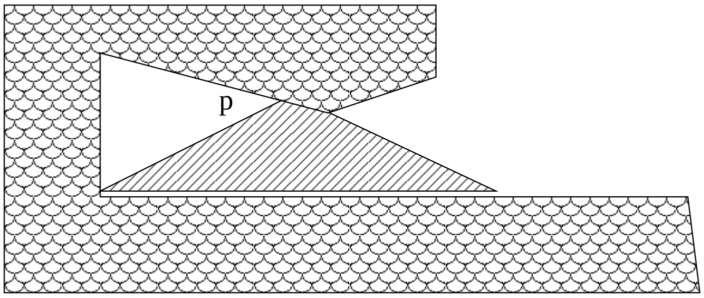
Some justification should be given for the restrictions placed on "strongly separable directions" in definition 4.10, which are not obviously necessary, and which end up restricting the scope of theorem 4.2. First, the reason that, in definition 4.2, we require, not only that a separating direction $\hat{w}$ is not a colliding direction, but also that $-\hat{w}$ *is* a colliding direction, is to exclude cases like those in figure 16. In part A of this figure, the triangular object can be moved away from the frame, but in part B, which can be made arbitrarily close in terms of approximation in tangent, it cannot. Though there is a range of motions, rightward with a small degree of upward, that separate the triangle from the frame, their reverses are not colliding directions at the point $\mathbf{p}$, (though, of course, they are collding directions elsewhere), and so the condition of the theorem fails. The condition that $-\hat{w}$ is a colliding direction ensures that one object cannot be expanded in a direction that will block the other.

The reason that definition 4.11 requires an open interval of separating directions is illustrated in figures 17 and 18. Both of these illustrate cases, in (A), where one object is separable from another, but only in one single direction. In each case, a contraction that is arbitrarily small, in terms of "approximation in tangent," suffices to create a "stuck" configuration in B.

For a semi-algebraic display, values satisfying theorem 4.2 can be computed using an algebraic formula.

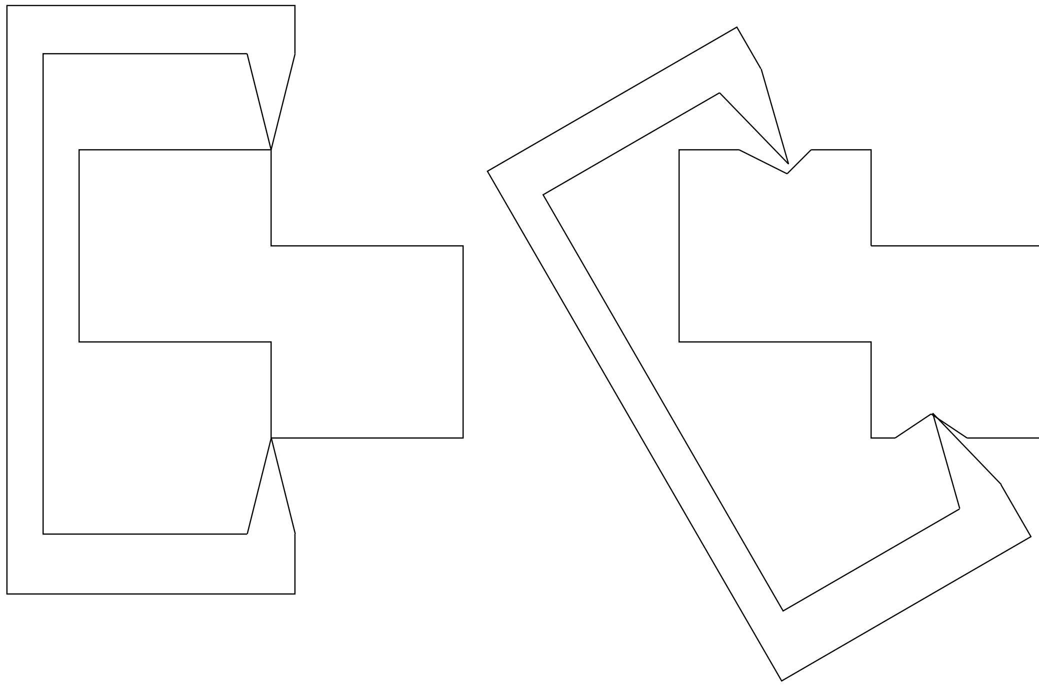


A



B

Figure 16: Failure of theorem 4.2 if $-\hat{w}$ is not a colliding direction

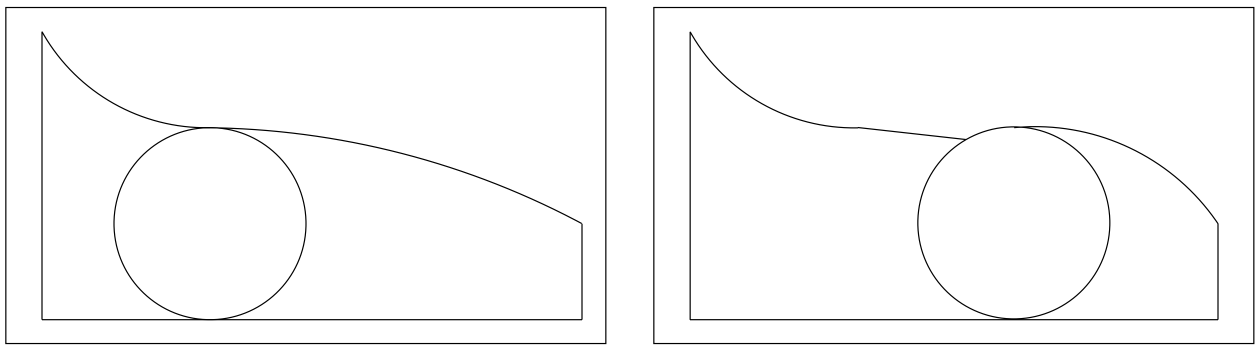


A

B

The pincer is "stuck" in (B) if the sides of the notch are steeper
than the circle through one tooth centered at the other tooth.
With a sufficiently small rotation, this can be accomplished
with an arbitrarily small and shallow notch.

Figure 17: Failure of theorem 4.2 if there is only one separating direction: I



A

B

In (A) the ball is free to move to the left.
However, a "trapped" position can be created by carving out a "notch" as in (B),
of arbitrarily small depth and change in angle.

Figure 18: Failure of theorem 4.2 if there is only one separating direction: II

Theorem 4.3: Let D be a semi-algebraic display over two objects A and B that is always strongly separable. Then there exists an algebraic formula $\Phi(\epsilon, \alpha, \psi)$ with the following properties:

- For any $\epsilon > 0$ there exists $\alpha > 0$ and $\psi > 0$ such that $\Phi(\epsilon, \alpha, \psi)$.
- If $\Phi(\epsilon, \alpha, \psi)$ and D' is a display that approximates D in tangent (α, ψ), then $\mathrm{p}^D_{tS}(\text{paths}(\text{free}(D)), \text{paths}(\text{free}(D')) < \epsilon$.
- The form of Φ can be computed from the forms of the regions in D.

It is possible to extend theorem 4.2 to systems of several objects along the following lines: Let ϕ be a continuously differentiable path through configuration space. For point $\mathbf{p}$, define the *relative motion* of $\mathbf{p}$ in object J relative to object I under path ϕ as the velocity of $\mathbf{p}$ under the motion of J relative to its velocity under the motion of I: $(d/dt\phi(T)[J](\mathbf{p})) - (d/dt\phi(T)[I](\mathbf{p}))$ A motion ϕ is strongly separating in scenario $\langle D, C\rangle$ where $C = \phi(0)$ if, for every pair of objects $I \neq J$, and for every point $\mathbf{p} \in CD[I] \cap CD[J]$, the motion of $\mathbf{p}$ in J relative to I is non-zero and its direction strongly separates $CD[J]$ from $CD[I]$ at $\mathbf{p}$. Display D is always strongly separable if for every $C \in$free(D) there exists a motion ϕ that is strongly separating in scenario $\langle D, C\rangle$.

Theorem 4.4: Let D be a display over n objects that is always strongly separable, and such that free(D) is locally ordinarily connected. Then there exist $\alpha > 0$ and $\psi > 0$ such that, if D' is a display that approximates D in tangent (α, ψ), then $\mathrm{p}^D_{tS}(\text{paths}(\text{free}(D)), \text{paths}(\text{free}(D')) < \epsilon$.

It seems plausible to conjecture that the analogous theorem holds in $\Re^k$ for $k > 2$ as well, and, indeed the same proof applies directly in the case of smooth shapes. It is also plausible to conjecture that, if D' approximates D sufficiently well in tangent, then free(D') is homeomorphic to free(D) and, further, that free(D') can be made to approximate free(D) arbitrarily well in tangent. This last result is likely to be necessary, or at least useful, in proving that small changes to a mechanism do not change its qualitative behavior, since corners in configuration space correspond to jammed positions of a mechanism, and it is therefore desirable to give conditions under which no such corners can emerge. However, I have not proven either of these results.

5 Conclusions

This paper has accomplished the following:

- We have defined two original measures of shape approximation: the dual-Hausdorff distance and approximation in tangent.
- We have defined three criteria for evaluating the similarity of the configuration spaces associated with kinematic systems: the Hausdorff distance p^D_H, the Hausdorff-tracking distance p^D_{tH}, and the path-set similarity measure p^D_{tS}. These express characteristics of the domain that intuitively seem significant.
- We have stated a number of theorems giving sufficient conditions under which approximately correct shape descriptions give rise to approximately correct configuration spaces, under the above senses of approximation. The accuracy of the approximation is computable. As far as we know, this is the most flexible representation for incomplete knowledge of shapes which has been shown to support computable reasoning in a substantial physical domain. (By contrast, the tolerances used in (Joskowicz, Sacks, and Srinivasan, 1997) are much less expressive, being only parametric tolerancing; the first-order languages of *ad hoc* spatial relations used in (Hayes, 1985) and (Davis, 1988) are almost certainly non-decidable; and the physical applications proposed for the RCC calculus (Randell, Cui, and Cohn, 1992) are unconvincing.)

The theorems proven here are significant, more as encouragement for this line of research than for their direct application. A central problem in "diagrammatic reasoning"[3] in the sense of doing calculations based on exact shape descriptions, is the problem of knowing when inferences based on idealized shapes are in fact valid for the actual shapes. The results in this paper establishes cases where this can be done soundly in an important physical domain. If such results can be obtained for other domains as well, then the power of diagrammatic reasoning, in this sense, will be significantly increased.

In a larger setting, the results in this paper are a small contribution toward solving the problem of geometric idealization. Russell (1948, p. 238) describes this problem as follows:

> When, in surveying, we use the process of triangulation, it is admitted that our triangles do not have accurate straight lines for their sides nor exact points at their corners, but this is glozed over by saying that the sides are *approximately* straight and the corners *approximately* points. It is not at all clear what this means, so long as it is maintained that there are no exact straight lines or points to which our rough-and-ready lines and points approximate. We may mean that sensible lines and points have approximately the properties set forth by Euclid, but unless we can say, within limits, how close the approximation is, such a view will make calculation vague and unsatisfactory.

One aspect of this problem, then, is to determine, for each geometric property of the idealized shapes, whether we should expect it to hold, exactly or approximately, for the real world objects being described. One means to making this determination is find that a property carries over from an idealized shape to an approximate shape. This paper has studied this for kinematic properties. That is to say, one aspect of establishing that two physical object may be reasonably described as "square" is to establish that their kinematic interactions resemble those that we expect in the configuration space of two ideally square objects. This paper gives reasons to expect that behavior of approximately square objects will, in fact, often approximate the behavior of ideally square objects.

Alan Perlis (1982) wrote, "One can't proceed from the informal to the formal by formal means," which is true, of course. However, one can to some extent justify the informal process of going from an real-world problem to a highly idealized formal description by giving a formal account of the relation between this extremely idealized formal description and a more realistic formal description. For idealization in shape, in the domain of kinematic behavior, that is what this paper accomplishes.

References

Ernest Davis, "A Logical Framework for Commonsense Predictions of Solid Object Behavior" : *AI in Engineering,* vol. 3, 1988. pages 125-140.

Rudolf Fleischer, Kurt Melhorn, Gunter Rote, Emo Welzl, and Chee Yap, "Simultaneous Inner and Outer Approximation of Shape," *Algorithmica,* vol. 8, 1992, pp. 365-389.

J. Glasgow, N.H. Narayanan, and B. Chandrasekaran, *Diagrammatic Reasoning,* MIT Press, 1995.

A. Grzegorczyk, "Undecidability of some Topological Theories." *Fundamenta Mathematicae* vol. 38, 1951, pages 137-152

Patrick Hayes. "Naive Physics 1: Ontology for Liquids." In Hobbs, J. and Moore, R. (eds.) *Formal Theories of the Commonsense World.* Norwood, N.J.: Ablex Pubs. 1985

[3] This buzzword (Glasgow, Narayanan, and Chandrasekaran, 1995) is used in a half-dozen incompatible ways, but this seems the most common.

Leo Joskowicz, "Simplification and Abstraction of Kinematic Behaviors," *Proc. IJCAI-89,*, pp. 1337-1342.

Leo Joskowicz and Elisha Sacks, "Computational Kinematics," *Artificial Intelligence,* vol. 51, 1991, pp. 381-416.

Leo Joskowicz, Elisha Sacks, and Vijay Srinivasan, "Kinematic Tolerance Analysis," *Computer-Aided Design,* Vol. 29(2), 1997.

Leo Joskowicz and Russell Taylor, "Interference-Free Insertion of a Solid Body into a Cavity: Algorithms and a Medical Application," *International Journal of Robotics Research,* Vol. 15, No. 3 1996.

Bhubaneswar Mishra, *Algorithmic Algebra,* Springer-Verlag, New York, 1993.

James R. Munkres, *Topology: A First Course,* Prentice-Hall, 1975.

Alvin G. Neumann, "The new ASME Y14.5M standard on dimensioning and tolerancing." *Manufacturing Review*, Vol. 7, No. 1, 1994, pages 16-23.

Alan Perlis, "Epigrams in Programming" *SIGPLAN*, September 1982.

D. A. Randell, Z. Cui, and A. G. Cohn (1992). "A Spatial Logic Based on Regions and Connection," *Third Intl. Conf. on Principles of Knowledge Representation and Reasoning.* pp. 165-176.

Aristides A.G. Requicha, "Towards a Theory of Geometric Tolerancing," *The International Journal of Robotics Research,* vol 2, no. 4, 1983, pp. 45-60.

Bertrand Russell, *Human Knowledge: Its Scope and Limits.* Simon and Schuster, New York. 1948.

Chee Yap and Ee-Chien Chang, "Issues in the Metrology of Geometric Tolerancing," in J.P. Laumond and M. Overmars (eds.) *Algorithms for Robot Motion Planning and Manipulation,* A.K. Peters, Wellesley, Mass. 1996.

Appendix A: Summary of geometric types and distances

We have dealt with a rather large number of different types of geometric entities, including both specific and abstract categories, and corresponding measures of distance. For purposes of reference, these are tabulated here.

1. Points in two- or three-dimensional Euclidean space. These are indicate by bold-face, lowercase letters, such as $\mathbf{p}$. We use the Euclidean distance between two points $\mathbf{p}$ and $\mathbf{q}$, denoted "$\mathrm{d}(\mathbf{p}, \mathbf{q})$".

2. Directions. A direction is denoted using a vector with a hat, such as $\hat{t}$. By identifying a direction with the corresponding unit vector, we use the Euclidean distance $\mathrm{d}(\hat{u}, \hat{v})$ to measure the distance between two directions $\hat{u}$ and $\hat{v}$.

3. Sets of points in a metric space. Let O be a space with metric μ, and let U and V be subsets of O.

   a. The *Hausdorff construction* (definition 2.1) defines a distance measure between U and V, denoted $\mu_H(U, V)$.

   b. The *connected-component similarity measure* $\mu_S(U, V)$ is another distance measure between U and V. (Definition 4.18)

4. Regions in Euclidean space. These are indicated by bold-face, upper-case letters, such as $\mathbf{R}$. We define four different measure of distance between spatial regions:

   a. The Hausdorff construction applied to the Euclidean distance function gives the distance measure $\mathrm{d}_H(\mathbf{R}, \mathbf{S})$. (Definition 1.3)

   b. The complement-Hausdorff distance, $\mathrm{d}_{Hc}(\mathbf{R}, \mathbf{S})$. (Definition 1.3)

   c. The dual-Hausdorff distance, $\mathrm{d}_{Hd}(\mathbf{R}, \mathbf{S})$. (Definition 1.3)

   d. Approximation in tangent (α, ϵ). (Definition 4.9)

5. Displays. A display is a tuple of regions, corresponding to the shapes of a collection of objects. The same distance functions used for regions are extended to apply to displays. (Definition 1.13)

6. Configurations. A configuration is a specification of the positions of a collection of objects. We will define a distance measure, $\mathrm{p}^D(C_1, C_2)$ between two configurations C_1, C_2. The display D is a parameter of the distance measure; each different display defines a different distance measure. (Definition 1.13).

7. Regions in configuration space. The Hausdorff construction, applied to the measure p^D, gives the distance measure $\mathrm{p}^D_H(R, S)$ between two regions R, S in configuration space.

8. Paths through a metric space. Greek letters are used for paths. If O is a space with metric μ, then we measure the distance between paths ϕ and ψ using the *tracking distance*, denoted $\mu_t(\phi, \psi)$. (Definition 3.1)

9. Sets of paths through metric space. Let O be a space with metric μ, and let U and V be sets of paths through O.

   a. Applying the Hausdorff construction (3.a) to the tracking distance μ_t gives measure $\mu_{tH}(U, V)$. (Definition 3.2)

   b. Applying the connected-components similarity construction (3.b) to the tracking distance μ_t gives measure $\mu_{tS}(U, V)$.

10. Paths and sets of paths through configuration space. We specialize the metric μ in (8) and (9) to be the metric p^D defined in (6). Then we get

    a. The measure $\mathrm{p}^D_t(\phi, \psi)$ between two paths ϕ and ψ in configuration space, by applying the construction (8) to (6).

    b. The measure $\mathrm{p}^D_{tH}(A, B)$ between two sets A and B of paths through configuration space, by applying (9a) to (6).

    c. The measure $\mathrm{p}^D_{tS}(A, B)$ between two sets A and B of paths through configuration space, by applying (9b) to (6).

    As in (6), D, a display, is a parameter of the distance measure; each different display gives a different measure.

Appendix B: Point-set topology

In this appendix, we will review some basic definitions and results from point-set topology over Euclidean space. These can be found in any introductory topology textbook, such as [Munkres, 75].

Definition B.1: A *metric* over a set O is a real-valued function μ over $O \times O$ satisfying the following properties:

- $\mu(X, Y) = 0$ if and only if $X = Y$. (Positive definite)
- $\mu(X, Y) = \mu(Y, X)$. (Symmetric)
- $\mu(X, Z) \leq \mu(X, Y) + \mu(Y, Z)$. (Triangle inequality)

The pair of the set O and the metric μ is called a *metric space*.

Definition B.2: A *topology* consists of a space O and a collection $\mathcal{U}$ of subsets of O, called "open sets" satisfying the following:

- $U \in \mathcal{U}$ and $\emptyset \in \mathcal{U}$.
- If $U \in \mathcal{U}$ and $V \in \mathcal{U}$ then $U \cap V \in \mathcal{U}$. ($\mathcal{U}$ is closed under finite intersection.)
- If $\mathcal{S} \subset \mathcal{U}$ then $\cup_{U \in \mathcal{S}} \in \mathcal{U}$. ($\mathcal{U}$ is closed under arbitrary union.)

Definition B.3: A subset $C \subset O$ is *closed* in a given topology over O if the complement $O - C$ is open. A *neighborhood* of point p is an open set containing p. The *interior* of set S, denoted "Int(S)" is the union of all open subsets of S; equivalently, the maximal open subset of S. The *closure* of S is the intersection of all closed sets containing S; equivalently, the minimal closed set containing S. The *boundary* of S, denoted "Bd(S)" is the set difference of the closure minus the interior.

Definition B.4: Let O be a space with metric μ. The standard topology for that metric space is defined as follows: A set U is open if and only if, for every point $x \in U$ there exists a value $\epsilon > 0$ such that $B(x, \epsilon) \subset O$. (Recall that $B(x, \epsilon)$ is the open ball of radius ϵ around x; that is, the set of all y such that $\mu(y, x) < \epsilon$.

It is easily shown that definition B.4 does, indeed, define a topology satisfying the conditions of definition B.2.

Example B.1: Let $\Re$ be the real line with the metric $\mu(X, Y) = \mid X - Y \mid$. The open interval $U = (a, b)$ is open in the standard topology because, for any $x \in U$, the open ball around x of radius $\min(\mid x - a \mid, \mid x - b \mid)$ is a subset of U. (Note that since the interval is open, x cannot be equal to a or b, so the radius is non-zero.) The closed interval $U = [a, b]$ is not open in the topology because, if we choose $x = a$, any open ball around x must include points less than a, and is therefore not a subset of U.

Definition B.5: Let O and P be topological spaces. A function f from O to P is *continuous* if, for every open set $U \subset P$, the inverse image $f^{-1}(U)$ is open in O.

Example B.2: The function

$$f(x) = \begin{cases} x & \text{if } x \leq 0 \\ x + 1 & \text{if } x > 0 \end{cases}$$

is discontinuous. Note that if U is the open set $(-1, 1)$ then the inverse image $f^{-1}(U)$ is the union of the two intervals $(-1, 0] \cup (1, 2)$, which is not an open set.

In a metric space, continuity can also be given by the standard "epsilon-delta" definition.

Theorem B.1: Let O be a space with metric μ and let P be a space with metric τ. A function f from O to P is continuous if and only if the following condition holds: for every point $o \in O$ and for every $\epsilon > 0$ there exists a $\delta > 0$ such that, if $q \in O$ and $\mu(q, o) < \delta$ then $\tau(f(q), f(o)) < \epsilon$. ∎.

The notion of "compactness" is central, though the definition is not at all obvious. First, a preliminary definition.

Definition B.6: An *open covering* of set S is a collection $\mathcal{O}$ of open sets whose union contains S. A *subcovering* of $\mathcal{O}$ is a subcollection whose union still covers S.

Definition B.7: A set S is *compact* if every open covering of S has a finite subcovering.

Example B.3: Let S be the open interval (0,1). S is not compact, because the collection $\mathcal{O}$ of open intervals { (1/2,1), (1/3,1), (1/4,1) ...} has no finite subcovering. (Any finite subcollection will reach only down to $1/m$, for some finite m.) The closed interval [0,1] is compact. Note that $\mathcal{O}$ does not cover [0,1] because it does not cover the points 0 and 1.

Compact sets are important because of theorems such as the following:

Theorem B.2: If O is compact and f is continuous, then $f(O)$ is compact.

Theorem B.3: In Euclidean space $\Re^k$, a set is compact if and only if it is closed and bounded.

Theorem B.4: If O is compact, and f is a continuous real-valued function over O, then f attains a maximum and a minimum value on O.

Definition B.8: Let S be an infinite subset of a topological space O. A point $x \in O$ is a *cluster point* of set S if, for every neighborhood N of x, there is an element $y \in N \cap S$ such that $y \neq x$. Note that x may or may not be an element of S.

Example: If S is the set of reciprocal integers { 1, 1/2, 1/3 ...} then 0 is a cluster point of S. The set of integers has no cluster point.

Theorem B.5: If O is compact and S is an infinite subset of O, then S has a cluster point in O.

The following lemmas about compact sets, which we will use, are less standard, so we include the proofs.

Lemma B.6: Let U be a compact metric space with metric μ. Let f be a non-negative continuous function over U. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that, for all X, if $f(X) < \delta$ then there exists a Y such that $\mu(X,Y) < \epsilon$ and $f(Y) = 0$. That is, any X with a very small value of f lies close to some Y where $f = 0$.

Proof: Let $B = \{Z \in U \mid \exists_Y f(Y) = 0 \wedge \mu(Z,Y) < \epsilon\}$. That is, B is all points within ϵ of some zero of f. Since B is open, $U - B$ is compact, so f attains a minimum value $\delta > 0$ on $U - B$. This δ then satisfies the condition of the lemma. ∎

Lemma B.7: Let U be a compact metric space with metric μ and let f be a continuous, real-valued function over U. Then, for any $\epsilon > 0$, there exists a $\delta > 0$, such that every point $X \in U$ where $f(X) > 0$ is within ϵ of a point Y where $f(Y) > \delta$.

Proof: Fix a value $\epsilon > 0$. For any $\delta > 0$, define the set $Q_\delta = \{Y \mid \exists_X f(X) > \delta \wedge \mu(X,Y) < \epsilon\}$. Thus, for any $b < a$, $Q_b \supset Q_a$. Let W be the closure of the set $\{X \mid f(X) > 0\}$; then clearly $\cup_{\delta>0} Q_\delta \supset W$. Since W is compact, there must be a finite subcover of W, $Q_{\delta 1}, Q_{\delta 2} ... Q_{\delta k}$. Choosing δ to be the smallest of these δ_i, we infer that $Q_\delta \supset W$, which is the desired result. ∎

Lemma B.8: Let U be a compact space. For $u \in U$ and $x_1 \ldots x_k \in \Re$, let $\Phi(u, x_1 \ldots x_k)$ be a property of u such that

a. If $0 < y_i < x_i$ for $i = 1 \ldots k$, and $\Phi(u, x_1 \ldots x_k)$ then $\Phi(u, y_1 \ldots y_k)$.

b. For any $u \in U$ there exists a neighborhood V of u and values $z_1 > 0 \ldots z_k > 0$ such that, for all $v \in V$, $\Phi(v, z_1 \ldots z_k)$.

Then there exist $x_1 > 0 \ldots x_k > 0$ such that for all $u \in U$, $\Phi(u, x_1 \ldots x_k)$.

Proof: Let $U_n = \{u \mid \Phi(u, 1/n \ldots 1/n)\}$. Then the collection of U_n is a covering of U by open sets.

Since U is compact, this has a finite subcover. But, for $n > m$, $U_n \supset U_m$. Thus the last U_n in the finite subcover contains all the rest and is so equal to U. Thus $\Phi(u, 1/n \ldots 1/n)$ for all $u \in U$. ∎

Finally, we need the concept of "connectedness". There are actually two such concepts of connectedness in the topological literature.

Definition B.9: Two non-empty sets O and P are *separable* if there exist disjoint open sets U and V such that $O \subset U$ and $P \subset V$. A set U is *disconnected* if it can be partitioned into two separable sets O and P such that $U = O \cup P$. A set U is *connected* if it is not disconnected. Equivalently, U is connected if it satisfies the following condition: For any O, P, U, V if $U = O \cup P$, O and P are non-empty, U and V are open, $O \subset U$ and $P \subset V$ then $U \cap V$ is non-empty.

Example B.4: The set $U = [-1, 0) \cap (0, 1]$ is disconnected; set $O = [-1, 0), P = (0, 1], U = (-2, 0), V = (0, 2)$. The single interval $[-1, 1]$ is connected.

Definition B.10: A *path* through set S is a continuous function from the interval [0,1] into S.

Definition B.11: A set S is *path-connected* if for any two points $s, t \in S$, there is a path from s to t through S.

A set that is path connected is necessarily connected. The converse does not hold; there are anomalous sets that are connected but not path connected. For example, consider the planar region.

$$S = \{\langle x, \sin(1/x) \mid x \in (0, 1)\} \cup \{\langle 0, y \rangle \mid y \in [-1, 1]$$

S consists of a sine curve that oscillates faster and faster, infinitely often, as it approaches the y-axis from the right, together with the line segment from $\langle 0, -1 \rangle$ to $\langle 1, 0 \rangle$. Then it can be shown that S is connected; note that any open set containing the vertical segment on the left must also contain some chunk of the sine curve. But S is not path-connected; there cannot be a path connecting points in the curve to points in the segment because it would not have a unique limit as it moves down the curve toward the segment.

In the AI literature, the two concepts are sometimes conflated. As they are identical on all sets that arise in practice, no great harm is done.

Appendix C: Proofs

C.1. Proofs for section 1

Theorem 1.4: The dual-Hausdorff distance from **B** to **A** is greater than or equal to the zone-tolerance measure from **B** to **A**.

Proof: Let Z be the zone-tolerance measure from **B** to **A**. Let **p** be a point in Bd(**B**) such that d(**p**,Bd(**A**)) = Z. There are two cases to consider: Either **p** is in Int(**A**) or **p** is in Ext(**A**).

Suppose first that **p** is in Int(**A**). Chose any $\epsilon > 0$. Since **p** ∈Bd(**B**) there is a point **q** ∈Ext(**B**) such that $\mathrm{d}(\mathbf{p}, \mathbf{q}) < \epsilon$. Let **r** be any point in Ext(**A**). Since the line segment **pr** goes from Int(**A**) to Ext(**A**), it must cross Bd(**A**) at some point **s** between **p** and **r**. Since **s** ∈Bd(**A**) and d(**p**,Bd(**A**)) = Z, we have $Z \leq \mathrm{d}(\mathbf{p}, \mathbf{s}) < \mathrm{d}(\mathbf{p}, \mathbf{r}) \leq \mathrm{d}(\mathbf{q}, \mathbf{r}) + \epsilon$, by the triangle inequality. But since **r** was an arbitrary point in Ext(**A**), this means that d(**q**,Ext(**A**)) $> Z - \epsilon$, and since ϵ was arbitrary, it means that $Z \leq \inf_{\mathbf{q} \in \mathrm{Ext}(\mathbf{B})} \mathrm{d}(\mathbf{q}, \mathrm{Ext}(\mathbf{A})) \leq \mathrm{d}_{Hc}(\mathbf{A}, \mathbf{B}) \leq \mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B})$.

The proof in the case where **p** is in Ext(**A**) is exactly analogous, reversing interiors with exteriors.
∎

C.2. Proofs of theorems in section 2

C.2.1. Proof of theorem 2.1

Lemma C.2.1: In any space with metric μ,

a. If $A \subset B \subset C \subset D$ then $\mu_H(A, D) \geq \mu_H(B, C)$.

b. If $A \subset B \subset D$ and $A \subset C \subset D$ then $\mu_H(B, C) \leq \max(\mu_H(B, A), \mu_H(B, D))$.

Proof: Immediate from the definition. ▮

Lemma C.2.2: For any regions $\mathbf{R}$, $\mathbf{R}'$ $\mathbf{S}$, $\mathbf{S}'$, $\mathrm{d}(\mathbf{R}, \mathbf{S}) \leq \mathrm{d}(\mathbf{R}', \mathbf{S}') + \mathrm{d}_H(\mathbf{R}', \mathbf{R}) + \mathrm{d}_H(\mathbf{S}', \mathbf{S})$.

Proof: Straightforward. ▮

Lemma C.2.3: For any regions $\mathbf{R}$, $\mathbf{R}'$ $\mathbf{S}$, $\mathbf{S}'$,

$$\mathrm{o}(\mathbf{R}, \mathbf{S}) \leq \mathrm{o}(\mathbf{R}', \mathbf{S}') + \max(\mathrm{d}_{Hc}(\mathbf{R}', \mathbf{R}), \mathrm{d}_{Hc}(\mathbf{S}', \mathbf{S}))$$

Proof: Let $\mathbf{O}$ be a sphere of radius $\mathrm{o}(\mathbf{R}, \mathbf{S})$ contained in $\mathbf{R} \cap \mathbf{S}$. Any point in the complement of $\mathbf{R}'$ must lie within $\mathrm{d}_{Hc}(\mathbf{R}', \mathbf{R})$ of the complement of $\mathbf{R}$. Hence any point in $\mathbf{O} - \mathbf{R}'$ must lie within $\mathrm{d}_{Hc}(\mathbf{R}', \mathbf{R})$ of the boundary of $\mathbf{O}$. Similarly, any point in $\mathbf{O} - \mathbf{S}'$ must lie within $\mathrm{d}_{Hc}(\mathbf{S}', \mathbf{S})$ of the boundary of $\mathbf{O}$. Hence the sphere concentric with $\mathbf{O}$ of radius $\mathrm{o}(\mathbf{R}, \mathbf{S}) - \max(\mathrm{d}_{Hc}(\mathbf{R}', \mathbf{R}), \mathrm{d}_{Hc}(\mathbf{S}', \mathbf{S}))$ must contain no points in the complement of $\mathbf{R}' \cap \mathbf{S}'$. So $\mathrm{o}(\mathbf{R}', \mathbf{S}')$ must be at least $\mathrm{o}(\mathbf{R}, \mathbf{S}) - \max(\mathrm{d}_{Hc}(\mathbf{R}', \mathbf{R}), \mathrm{d}_{Hc}(\mathbf{S}', \mathbf{S}))$. ▮

Lemma C.2.4: Let D and D' be two displays. Then for any configuration C,

$$\mathrm{clearance}(D, C) \leq \mathrm{clearance}(D', C) + 2\mathrm{d}_H(D, D').$$
$$\mathrm{overlap}(D, C) \leq \mathrm{overlap}(D', C) + \mathrm{d}_{Hc}(D, D').$$

Proof: Immediate from lemmas C.2.2 and C.2.3. ▮

Lemma C.2.5: The functions "free(D)" and "cfree(D)" are monotonically non-increasing. That is, if D' is an expansion of D, then free(D') is a subset of free(D) and cfree(D') is a subset of cfree(D).

Proof: Immediate. ▮

Lemma C.2.6: Let D be a display, and let U be a compact region of configuration space. For any $\epsilon > 0$ there exists a $\delta > 0$ such that, for every display D', if $\mathrm{d}_{Hc}(D, D') < \delta$ then any configuration in free(D') $\cap\, U$ is within ϵ of some configuration in free(D).

Proof: For any display D and configuration C, let $\mathrm{overlap}^D(C) = \mathrm{overlap}(\langle D, C \rangle)$. By lemma C.2.3, if $C' \in$free(D') and $\mathrm{d}_{Hc}(D, D') < \delta$, then $\mathrm{overlap}^D(C') < \delta$. Applying lemma A.2.2, with $f(x)$ being the function $\mathrm{overlap}^D$ and with μ being the metric p^D, we infer that we can choose δ so that, for any C' in free(D) $\cap\, U$, if $\mathrm{overlap}^D(C') < \delta$ then there exists a configuration C such that $\mathrm{p}^D(C, C') < \epsilon$ and $\mathrm{overlap}^D(C) = 0$, so $C \in$free(D). ▮

If T and C are two configurations, we will write "$T \circ C$" to mean the configuration $\langle T[1] \circ C[1], \ldots, T[k] \circ C[k] \rangle$.

Definition C.2.1: Let f be a function over configuration space; let U be an open region in configuration space; and let T be a tuple of rigid mappings. We say that T preserves f over U if for every configuration $C \in U$, $f(T \circ C) = f(C)$.

Definition C.2.2: Let D be a display in n-dimensional space. Let $\Delta = \sum_{i=1}^{k}$ (diameter($D[i]$) + 3). The *basic configuration region* of D is the set of all configurations C such, for every i, $CD[i]$ lies inside the box $[0, \Delta]^n$.

Lemma C.2.7 The basic configuration region of D is compact.

Proof: A rigid mapping on n-dimensional Euclidean space can be expressed in a standard way as an $(n+1)^2$ matrix. Hence, a configuration on a k-object display D can be viewed as a $k \cdot (n+1)^2$ vector dimensional vector and the configuration space as a whole is a $kn(n-1)/2$ dimensional manifold of such vectors. It is easily verified (a) that the standard topology over the manifold is the same as the topology defined by the metric $\mathrm{p}^D(C_1, C_2)$; (b) that a set of configurations is bounded in the manifold if and only if it is bounded relative to the metric $\mathrm{p}^D(C_1, C_2)$; (c) that the basic configuration region is closed and bounded relative to the metric $\mathrm{p}^D(C_1, C_2)$. Hence, the basic configuration region is closed and bounded in the manifold; hence it is compact. ▮

We now define two functions over configuration space. $\Gamma^D(C)$ maps each configuration C into the basic configuration region, in a way that preserves the relative position of nearby objects. $\Theta^D(C)$ is the tuple of rigid mappings such that $\Theta^D(C) \circ C = \Gamma^D(C)$.

Definition C.2.3: Let C be any configuration over display D. We say that C has a *gap* in direction $\hat{u}$ from A to B, $A < B$ if

- There is an object I such that maximal coordinate of $CD[I]$ in the $\hat{u}$ direction is equal to A;
- There is an object J such that minimal coordinate of $CD[J]$ in the $\hat{u}$ direction is equal to B; and
- There is no object containing any point whose $\hat{u}$ coordinate is between A and B.

The size of this gap is $B - A$.

Definition C.2.4: Let C be any configuration over display D. We define the configuration $\Theta^D(C)$ using the following algorithm: In each coordinate direction $\hat{u}$, we begin by translating the whole configuration so that the lowermost point in the $\hat{u}$ direction in the scenario has coordinate 1. We then look for gaps $[A, B]$ of size greater than 3. We reduce any such gap to being of size exactly 3 by translating all the objects above the gap by a distance $B - (A + 3)$, while leaving all the objects below the gap where they are. We repeat until all such gaps are eliminated. We carry out this procedure in the $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions. When all this is complete, the final configuration is $\Gamma^D(C)$ and the combined transformations give $\Theta^D(C)$. Table 1 displays this algorithm in pseudo-code. It is easily seen that this procedure gives a unique result; in particular, that the result does not depend on the order in which gaps are reduced or coordinate directions are considered.

Definition C.2.5: The function "bclear$^D(C)$" is defined as min(clearance(D, C),1).

Lemma C.2.8: The functions $\Gamma^D(C)$ and $\Theta^D(C)$ have the following properties:

a. $\Gamma^D(C) = \Theta^D(C) \circ C$.

b. Both $\Gamma^D(C)$ and $\Theta^D(C)$ are continuous functions of C.

c. If two objects are within distance 3 in C then their relative position is the same in $\Gamma^D(C)$ as in C.

d. The distance between two objects is greater than or equal to 3 in C if and only if it is greater than or equal to 3 in $\Gamma^D(C)$.

e. $\Gamma^D(C)$ is in the basic configuration region, a distance 1 from the boundaries of that region.

begin T = the k-tuple of identity mappings;
  for $\hat{u}$ in $\{\hat{x}, \hat{y}, \hat{z}\}$ do /* $\hat{u}$ loops over the coordinate directions */
    begin let L be the minimum coordinate over all objects I of the $\hat{u}$ coordinate of $TCD[I]$;
      let W be the configuration all of whose elements are the translation $\lambda(\vec{P})(\vec{P} + (1 - L)\hat{u})$;
      $T := W \circ T$;
      loop let $[A, B]$ be any gap in $T(C)$ in the $\hat{u}$ direction of size > 3;
        if there is no such gap, exitloop
        else let W be the configuration defined as follows:
          for $I = 1 \ldots K$ do
            if $TC[I](D[I])$ has $\hat{u}$ coordinates less than A
              then $W[I]$ is the identity
              else $W[I]$ is the mapping $\lambda(\vec{V})(\vec{V} - (B - (A + 3))\hat{u})$;
          $T := W \circ T$;
      endloop
    end
  $\Theta^D(C) := T$; $\Gamma^D(C) = T \circ C$
end.

Table 1: Computing $\Theta^D(C)$

f. Let U be the open ball of radius 1 around C in configuration space. Then the function $\lambda(C')(\Theta^D(C)) \circ C'$ maps U into the basic configuration region and it preserves the functions overlap^D and bclear^D over U.

g. For any configurations C, C_1, C_2, we have $\text{p}^D(\Theta^D(C) \circ C_1, \Theta^D(C) \circ C_2) = \text{p}^D(C_1, C_2)$

Proof:

a. Immediate by construction.

b. Within a space of configurations all of which have the same system of large gaps (gaps of size greater than 3 in the same coordinates between the same objects), the continuity of Θ^D is immediate. Moreover, in any such space, as the size of a particular gap approaches 3, the transformation associated with closing the gap approaches the identity, which is its value once the size of the gap becomes 3. Thus Θ^D is continuous. The continuity of Γ^D follows immediately.

c. If two objects are within distance 3 in C, then there cannot be any large gaps between them. Hence, all the intermediate transformations W move them together, thus preserving their relative position.

d. If the two objects are originally on opposite sides of some large gap, that gap will be reduced to size 3, so they are still on opposite sides of a gap of size 3, and hence at least 3 apart. If they were distance 3 apart but not on opposite sides of any gap in C, then the transformations will move them together, their relative positions will be unchanged, and their distance will be unchanged. The converse argument hold the same way.

e. The range from the minimal coordinate of any point in any object in the x direction in $\Gamma^D(C)$ to the maximal coordinate is at most the sum of the diameters of the objects plus 3 times $K-1$, where K is the number of objects. Proof: Order the objects by increasing order of their lowest x coordinate. The lowest x coordinate of object J is at most 3 greater than the maximal x coordinate of some preceding object $I < J$, and hence at most 3 + diameter($D[I]$) greater than the minimal x coordinate of I. Moreover, the objects are placed so that the minimal coordinate in each direction is 1. Therefore, all the objects in $\Gamma^D(C)$ lie in the box $[1, \Delta - 1]^3$. Thus, $\Gamma^D(C)$ lies in the basic configuration region, as does any configuration within distance 1 of $\Gamma^D(C)$.

f. The fact that U is in the basic configuration region is immediate from (e). The fact that overlap^D and bclear^D are preserved is immediate from (c) and (d).

g. For each object i, $\Theta^D(C)[i]$ is just a translation. Hence, for any point $\mathbf{p}$, $\mathrm{d}([\Theta^D(C) \circ C_1](\mathbf{p}), [\Theta^D(C) \circ C_2](\mathbf{p})) = \mathrm{d}(C_1(\mathbf{p}), C_2(\mathbf{p}))$.

▮

Lemma C.2.9: Let D be a display. For any $\epsilon > 0$ there exists a $\delta > 0$ such that, for every display D', if $\mathrm{d}_{Hc}(D, D') < \delta$ then any configuration in free(D') is within ϵ of some configuration in free(D). (This is the same as lemma C.2.6, but with the restriction to a compact region dropped.)

Proof: Let Q be the basic configuration region of D. Let $\epsilon > 0$. Let δ_0 be chosen to satisfy lemma C.2.6 for the value $\min(\epsilon, 1)$ over region Q. Let $\delta < \min(\delta_0, 1, \min_i \mathrm{d}_{Hc}(D[i], \emptyset))$. (The complement-Hausdorff distance from a region $D[i]$ to the empty set is well-defined and finite; it is equal to the radius of the largest circle that can be inscribed in $D[i]$.) Let display D' be then chosen so that $\mathrm{d}_{Hd}(D, D') < \delta$ and let $C' \in$free(D'). Let E be the display such that $E[i] = D'[i] \cap D[i]$; then clearly $C' \in$free(E). It is immediate that $\mathrm{d}_{Hc}(E, D) \leq d_{Hc}(D', D) < \delta$. $E[i]$ is non-empty, because of the constraint that $\delta < \mathrm{d}_{Hc}(D[i], \emptyset)$.

By lemma C.2.8, any two objects in E that do not overlap in C' also do not overlap in $\Gamma^D(C')$, so $\Gamma^D(C') \in$free(E). Since $\Gamma^D(C') \in Q$, by lemma C.2.6, there is a configuration $C_1 \in$free(D) such that $p^D(C_1, \Gamma^D(C')) < \min(\epsilon, 1)$. Let $C = (\Theta^D(C'))^{-1} \circ C_1$. By lemma C.2.9.g, $p^D(C, C')$ = $p^D(C_1, \Gamma^D(C')) < \min(\epsilon, 1)$. By lemma C.2.9.f, $\Theta^D(C')$ preserves overlap^D on C, and hence $C \in$free(D). ▮

Theorem 2.1: Let D be a display. For any distance $\epsilon > 0$ there exists a distance $\delta > 0$ such that the following holds: if D' is a contraction of D and the complement-Hausdorff distance $d_{Hc}(D, D') < \delta$ then the Hausdorff distance between the free spaces p^D_H(free(D), free(D')) $< \epsilon$.

Proof: Showing that the Hausdorff distance is small has two parts:

a. Every configuration in free(D) is close to some configuration in free(D'). This is trivial, since free(D) is a subset of free(D') (lemma 2.4).

b. Every configuration in free(D') is close to some configuration in free(D). This is lemma 2.9.

▮

C.2.2: Proof of lemma 2.2

Lemma C.2.10: Let $\mathbf{R}$ and $\mathbf{S}$ be two compact regions. Then there exists a constant $\gamma > 0$ such that, for any rigid motions M_1 and M_2, $\mathrm{p}^{\mathbf{S}}(M_1, M_2) \geq \gamma \cdot \mathrm{p}^{\mathbf{R}}(M_1, M_2)$

Proof: Let s be the diameter of $\mathbf{S}$. Let $\mathbf{a}$ and $\mathbf{b}$ be two points in $\mathbf{S}$ such that $\mathrm{d}(\mathbf{a}, \mathbf{b}) = s$. Let $\mathbf{m}$ be the midpoint of $\mathbf{a}$ and $\mathbf{b}$. It is easily shown that, for any point $\mathbf{x}$, if $\mathbf{x}$ is closer to $\mathbf{a}$ than to $\mathbf{b}$, then $\mathrm{d}(\mathbf{x}, \mathbf{b}) \geq s/2$ and $\mathrm{d}(\mathbf{x}, \mathbf{b}) > \mathrm{d}(\mathbf{x}, \mathbf{m})$; whereas if $\mathbf{x}$ is closer to $\mathbf{b}$ than to $\mathbf{a}$, then $\mathrm{d}(\mathbf{x}, \mathbf{a}) \geq s/2$ and $\mathrm{d}(\mathbf{x}, \mathbf{a}) > \mathrm{d}(\mathbf{x}, \mathbf{m})$.

Let $\mathbf{e}$ be the furthest point in $\mathbf{R}$ from $\mathbf{m}$. We now claim that the lemma holds for the value $\gamma = s/(s + 2\mathrm{d}(\mathbf{e}, \mathbf{m}))$.

To show this, choose values for M_1, M_2, and let T be the rigid transformation that transforms M_1 into M_2; $T = M_2 M_1^{-1}$.

There are now two cases to consider.

I. T is a translation by vector $\vec{v}$. In that case, for any point $\mathbf{x}$, $M_2(\mathbf{x}) - M_1(\mathbf{x}) = \vec{v}$, so $\mathrm{d}(M_2(\mathbf{x}), M_1(\mathbf{x})) = \mid \vec{v} \mid$. Thus $\mathrm{p}^{\mathbf{O}}(M_1, M_2)$ has the same value $\mid \vec{v} \mid$ for any region $\mathbf{O}$, and thus the above inequality is satisfied with $C = 1$. Note that $s/(s + 2\mathrm{d}(\mathbf{e}, \mathbf{m})) < 1$, so the conclusion of the lemma is satisfied.

II. T is a rotation through angle α. In two dimensions, there will be a center of rotation $\mathbf{c}$; in three dimensions, there will be an axis of rotation $\mathbf{c}$. Note that, for any point $\mathbf{x}$, $\mathrm{d}(M_1(\mathbf{x}), M_2(\mathbf{x})) = 2\sin(\alpha/2)\mathrm{d}(\mathbf{x}, \mathbf{c})$. In either case, the following relations hold:

$\mathrm{p}^{\mathbf{S}}(M_1, M_2) \geq \max(\mathrm{d}(M_1(\mathbf{a}), M_2(\mathbf{a})), \mathrm{d}(M_1(\mathbf{b}), M_2(\mathbf{b})) \geq 2\sin(\alpha/2)\max(\mathrm{d}(\mathbf{a}, \mathbf{c}), \mathrm{d}(\mathbf{b}, \mathbf{c}) \geq 2\sin(\alpha/2)\max(s/2, \mathrm{d}(\mathbf{m}, \mathbf{c}))$

On the other hand, let $\mathbf{f}$ be the point in $\mathbf{R}$ furthest from $\mathbf{c}$. Then $\mathrm{p}^{\mathbf{R}}(M_1, M_2) = \mathrm{d}(M_1(\mathbf{f}), M_2(\mathbf{f})) = 2\sin(\alpha/2)\mathrm{d}(\mathbf{f}, \mathbf{c}) \leq 2\sin(\alpha/2)\mathrm{d}(\mathbf{e}, \mathbf{c}) \leq 2\sin(\alpha/2)\mathrm{d}(\mathbf{e}, \mathbf{m}) + \mathrm{d}(\mathbf{m}, \mathbf{c})$.

Thus,

$$\frac{\mathrm{p}^{\mathbf{S}}(M_1, M_2)}{\mathrm{p}^{\mathbf{R}}(M_1, M_2)} \geq \frac{\max(s/2, \mathrm{d}(\mathbf{m}, \mathbf{c}))}{\mathrm{d}(\mathbf{e}, \mathbf{m}) + \mathrm{d}(\mathbf{m}, \mathbf{c})}$$

It is easily shown that, for fixed $s, \mathbf{e}, \mathbf{m}$, the above fraction attains its minimum value of $s/(s + 2\mathrm{d}(\mathbf{e}, \mathbf{m}))$ when $\mathrm{d}(\mathbf{m}, \mathbf{c}) = s/2$. ∎

Lemma 2.2: Let D and E be two displays over k objects. Then the two metrics p^D and p^E induce the same topology on configuration space. Also, the two metrics p^D_H and p^E_H over regions of configuration space induce the same topology.

Proof: Using lemma C.2.10, for $i = 1 \dots k$, choose $\gamma[i]$ such that for any rigid motions M_1 and M_2, $\mathrm{p}^{D[i]}(M_1, M_2) \geq \gamma[i] \cdot \mathrm{p}^{E[i]}(M_1, M_2)$ and choose $\delta[i]$ such that for any rigid motions M_1 and M_2, $\mathrm{p}^{E[i]}(M_1, M_2) \geq \delta[i] \cdot \mathrm{p}^{D[i]}(M_1, M_2)$. Let $\gamma = \min_{i=1\dots k} \gamma[i]$ and let $\delta = \min_{i=1\dots k} \delta[i]$. Then, clearly, for any configurations C_1, C_2, $\mathrm{p}^D(C_1, C_2) \geq \gamma \mathrm{p}^E(C_1, C_2)$ and $\mathrm{p}^E(C_1, C_2) \geq \delta \mathrm{p}^D(C_1, C_2)$. Now, for any configuration C, for any $\epsilon > 0$, and for any display F, let $\mathrm{B}^F(C, \epsilon)$ be the set of all configurations C' such that $\mathrm{p}^F(C, C') < \epsilon$. It follows directly from the definitions that, for any configuration C and $\epsilon > 0$, $\mathrm{B}^E(C, \epsilon/\gamma) \subset \mathrm{B}^D(C, \epsilon)$ and that $\mathrm{B}^D(C, \epsilon/\delta) \subset \mathrm{B}^E(C, \epsilon)$, so p^D_H and p^E_H induce the same topology on regions of configuration space.

Also, for any region U in configuration space, for any $\epsilon > 0$, and for any display F, let $\mathrm{B}^F(U, \epsilon)$ be the set of all regions V in configuration space such that $\mathrm{p}^F_H(V, U) < \epsilon$. It follows directly from the definitions that, for any U and ϵ, $\mathrm{B}^E(U, \epsilon/\gamma) \subset \mathrm{B}^D(U, \epsilon)$ and that $\mathrm{B}^D(U, \epsilon/\delta) \subset \mathrm{B}^E(U, \epsilon)$, so p^D_H and p^E_H induce the same topology on regions of configuration space.

Proof of theorem 2.3

Theorem 2.3: Let D be a display and let $\delta > 0$ be a distance. Let $E = \mathrm{contract}(D, \delta)$. Let $\epsilon = \mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(E))$. Then D, ϵ, and δ satisfy the conditions of Theorem 2.1. That is, if F is a contraction of D and $\mathrm{d}_{Hc}(F, D) \leq \delta$ then $\mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(F)) \leq \epsilon$. (We have changed the strict inequalities in our original statement of theorem 2.1 to non-strict inequalities.)

Proof: Let F be a contraction of D such that $\mathrm{d}_{Hc}(F, D) < \delta$. By lemma 1.2, E is a contraction of F. By lemma C.2.5, $\mathrm{free}(E) \supset \mathrm{free}(F) \supset \mathrm{free}(D)$. Hence $\mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(F)) \leq \mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(E))$

$= \epsilon$.

Proof of theorem 2.4

In all our discussions of algebraic formulas, we will assume that rigid mappings are parameterized in terms of the coefficients of coordinate transforms relative to some fixed coordinate system (rather than, for example, in terms angles of rotations.)

Lemma C.2.11: The application of a rigid mapping to a point is a bilinear function of the mapping and the coordinates of the points, and thus an algebraic function.

Proof: Specifically, in two dimensions, a rigid mapping is specified in terms of four real parameters c_1, c_2, t_1, t_2 satisfying the constraint $c_1^2 + c_2^2 = 1$. The application of the mapping to a point with coordinate x_1, x_2 is given by the matrix operation

$$\begin{bmatrix} c_1 & c_2 \\ -c_2 & c_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

In three dimensions, a rigid mapping is specified using nine real parameters. The constraints and application formula are similar.

Lemma C.2.12: Let $\mathbf{R}, \mathbf{S}$ be semi-algebraic regions; let D, E be a semi-algebraic displays; let P, Q be semi-algebraic regions in configuration space. The following properties can then expressed in algebraic formulas whose form is easily computed from the forms of $\mathbf{R}, \mathbf{S}, D, E, P, Q$.

- The property of point $\mathbf{x}$, "$\mathbf{x}$ is in the interior of $\mathbf{R}$."
- The properties of point $\mathbf{x}$ and real ϵ, "$\mathbf{x}$ is in the uniform contraction/expansion of $\mathbf{R}$ by ϵ."
- The properties of points $\mathbf{x}_1 \ldots \mathbf{x}_k$ and ϵ, "For $i = 1 \ldots k$, $\mathbf{x}_i$ is in the uniform expansion/contraction of $D[i]$ by ϵ."
- The properties of distance m, "$m = \mathrm{d}_H(\mathbf{R}, \mathbf{S})$", "$m = \mathrm{d}_{Hc}(\mathbf{R}, \mathbf{S})$", and "$m = \mathrm{d}_{Hd}(\mathbf{R}, \mathbf{S})$".
- The properties of configuration C, "C is in free(D)/cfree(D)/forbidden(D)."
- The property of distance m, "$m = \mathrm{p}_H^D(P, Q)$."

Proof: Straightforward from the definitions. For example, cfree(D) can be defined as follows: Let $D[1] \ldots D[k]$ be algebraic formulas for the standard shapes of the objects in D. A configuration C is in cfree(D) if $C[1] \ldots C[k]$ satisfy the open formula

$$\wedge_{i \neq j} \forall_{x,y} [\mathbf{x} \in D[i] \wedge \mathbf{y} \in D[j]] \Rightarrow C[i](\mathbf{x}) \neq C[j](\mathbf{y})$$

∎

Abusing notation, we will use formulas like "$C \in$free(D)" to express the algebraic formula over the coordinates of C that expresses this relation for a fixed algebraic formula that expresses D. It should be kept in mind, though, that ultimately everything can be expanded into a first-order formula over real parameters.

Example C.2.1: Let D contain two objects: the circle of radius 2, $w^2 + x^2 \leq 4$ and the unit square $-1 \leq y \leq 1$, $-1 \leq z \leq 1$. Then the relation "$C \in$ free(D)" is an abbreviation for the relation over eight real parameters, $c_1, c_2, t_1, t_2, d_1, d_2, u_1, u_2$, given by

$$c_1^2 + c_2^2 = 1 \wedge d_1^2 + d_2^2 = 1 \wedge$$
$$\forall_{w,x,y,z}[w^2 + x^2 \leq 4 \wedge -1 \leq y \leq 1 \wedge -1 \leq z \leq 1] \Rightarrow$$
$$[c_1 w + c_2 x + t_1 \neq d_1 y + d_2 z + u_1] \vee [-c_2 w + c_1 x + t_2 \neq -d_2 y + d_1 z + u_2]$$

Theorem 2.4: Assume that display D is semi-algebraic. Then there is an algebraic formula $\Phi(\epsilon, \delta)$ which holds if and only if Theorem 2.1 is satisfied for D, ϵ and δ. Moreover, Φ is easily computable, given the form of D. Hence, by Tarski's theorem, for fixed semi-algebraic D, it is possible to compute δ from ϵ or ϵ from δ.

Proof: By lemma 1.2, the desired relation $\Phi(\epsilon, \delta)$ is given by the formula $\epsilon = \mathrm{p}_H^D(\mathrm{free}(D), \mathrm{free}(\mathrm{contract}(D, \delta))$. By lemma C.2.12, this formula is algebraic if D is semi-algebraic. ▮

Sketch of proof of theorem 2.5

Theorem 2.5: There exists a display D and a distance $\epsilon > 0$ with the following property: Let D' be a semi-algebraic display, and let $\delta = \mathrm{d}_{Hc}(D, D')$. Then there exists a display E such that $\mathrm{d}_{Hc}(E, D') \leq \delta$ but $\mathrm{p}_H^D(\mathrm{free}(D), \mathrm{free}(E)) > \epsilon$.

Sketch of proof: Pick an arbitrary value of $\epsilon > 0$. Let $\mathbf{R}$ be any transcendental region; for example $\{\langle x, y \rangle \mid 0 \leq x \leq \pi, 0 \leq y \leq \sin(x).\}$. Let $\mathbf{T}$ be any rectangle containing $\mathbf{R}$ whose boundary is at least 2ϵ from $\mathbf{R}$. Let $\mathbf{S} = \mathbf{T} - \mathbf{R}$. Let display D consist of two regions: Region A of shape $\mathbf{R}$ and region B of shape $\mathbf{S}$. Then free(D) contains a large space of configurations where A is outside B, and a unique (up to identical motions on both objects) configuration where A is placed inside the hole in B.

Choose D' and δ as above. Since D' is semi-algebraic and d_{Hc} is a semi-algebraic relation, it is not difficult to show that, if $\mathrm{d}_{Hc}(D, D') \leq \delta$ then $\mathrm{d}_{Hc}(E, D') \leq \delta$ for some proper expansion E of D. But object A will not fit inside object B if they are expanded. So the configuration in free(D) where A is inside B is at least ϵ from any configuration in free(E).

Proof of theorem 2.6

Lemma C.2.13: Let D be a display, and let U be a compact region of configuration space. For any $\epsilon > 0$ there exists a $\delta > 0$ such that, for every display D', if $\mathrm{d}_H(D, D') < \delta$ then any configuration in cfree(D) $\cap\, U$ is within ϵ of some configuration in cfree(D').

Proof: Applying lemma B.6, with $f(x)$ being the function bclear^D and with μ being the metric p^D, we infer that we can choose δ so that, for any C in cfree(D) $\cap\, U$, since $\mathrm{bclear}^D(C) > 0$, there exists a configuration C' such that $\mathrm{p}^D(C, C') < \epsilon$ and $\mathrm{bclear}^D(C') > \delta$. By lemma C.2.2, if $\mathrm{bclear}^D(C') > \delta$ and $\mathrm{d}_H(D, D') < \delta$, then $\mathrm{bclear}^{D'}(C') > 0$, so $C' \in$cfree(D'). ▮

Lemma C.2.14: Let D be a display. For any $\epsilon > 0$ there exists a $\delta > 0$ such that, for every display D', if $\mathrm{d}_H(D, D') < \delta$ then any configuration in cfree(D) is within ϵ of some configuration in cfree(D'). (This is the same as lemma 2.12, dropping the restriction to a compact region of configuration space.)

Proof: The proof is exactly analogous to the proof of lemma C.2.9 (in fact, slightly simpler): Start with any configuration $C \in$cfree(D). Use lemma C.2.8 to find the mapping $\Theta^D(C)$ taking C into the basic configuration region. Use lemma C.2.13 to find a configuration to find a configuration $C_0 \in$cfree(D') within ϵ of $\Gamma^D(C)$. Then mapping back $C' = (\Theta^D(C))^{-1} \circ C_0$ gives the desired answer. ▮

Theorem 2.6: The function "cfree(D)", mapping a display to a region in configuration space, is continuous under expansion, using the metric d_H on displays and the topology induced by p_H on regions of configuration space.

Proof: Unwrapping the definitions, this says that, if D' is an expansion of D and is close enough to D in the Hausdorff metric then

a. Every configuration in cfree(D) is close to some configuration in cfree(D'). This is lemma C.2.14.

b. Every configuration in cfree(D') is close to some configuration in cfree(D). This is trivial, since cfree(D') is a subset of cfree(D).

▮

Proof of theorem 2.7

Theorem 2.7: Let D be a display and let $\delta > 0$ be a distance. Let $E = \text{expand}(D, \delta)$. Let $\epsilon = \mathrm{p}_H^D(\text{cfree}(D),\text{cfree}(E))$. If F is an expansion of D and $\mathrm{d}_H^D(F, D) < \delta$,
then $\mathrm{p}_H^D(\text{cfree}(D),\text{cfree}(F)) \leq \epsilon$.

Proof: Let F be a expansion of D such that $\mathrm{d}_H(F, D) < \delta$. By lemma 1.2, E is an expansion of F. By lemma C.2.1, free(D) $\supset$ free(F) $\supset$ free(E). Hence $\mathrm{p}_{Hc}^D(\text{free}(D),\text{free}(F)) \leq \mathrm{p}_{Hc}^D(\text{free}(D),\text{free}(E))$ $= \epsilon$.

Proof of lemma 2.8

We first show that blurring the position of an object is essentially equivalent to expanding it.

Lemma C.2.14: Let D be a display, C be a configuration and $\epsilon > 0$ be a distance. Let $E = \text{expand}(D, \epsilon)$. Let $\mathrm{B}^D(C, \epsilon)$ be the set of all C' such that $\mathrm{p}^D(C, C') \leq \epsilon$. Let F be the display such that $F[I]$ is the union of $C'D[I]$ over all $C' \in \mathrm{B}^D(C, \epsilon)$. Then $F = C(E)$.

Proof: Fix an object I. If $\mathbf{x} \in F[I]$, then $\mathbf{x} \in C'D[I]$ for some C' such that $\mathrm{p}^D(C, C') \leq \epsilon$. Let $\mathbf{y} = C'[I]^{-1}(\mathbf{x})$; then $\mathbf{y} \in D[I]$. By definition of p^D, $\mathrm{d}(C[I](\mathbf{y}, \mathbf{x}) \leq \epsilon$. Since $C[I]$ is a rigid mapping, it follows that $\mathrm{d}(\mathbf{y}, C[I]^{-1}(\mathbf{x})) = \mathrm{d}(C[I](\mathbf{y}), \mathbf{x}) \leq \epsilon$. But since $\mathbf{y} \in D[I]$, $C[I]^{-1}(\mathbf{x})$ is in $E[I]$ so $\mathbf{x} \in CE[I]$.

Conversely, if $\mathbf{x} \in E[I]$, then let $\mathbf{y}$ be a point in $D[I]$ such that $\mathrm{d}(\mathbf{x}, \mathbf{y}) \leq \epsilon$. Let T be the translation by vector $\mathbf{x} - \mathbf{y}$, and let $C' = C \circ T$. Then $\mathrm{p}^D(C, C') = \mathrm{d}(\mathbf{x}, \mathbf{y}) \leq \epsilon$, and $C(\mathbf{x}) = C'(\mathbf{y})$. ▮

Lemma 2.8 For any display D, cfree(D) is the interior of free(D).

Proof: We need to show (a) that cfree(D) $\subset$ Int(free(D)) and (b) that Int(free(D)) $\subset$ cfree(D).

a. Suppose $C \in$cfree(D). Then there is a minimal distance $\epsilon > 0$ between the regions $CD[i]$, $CD[j]$ for any two objects $i \neq j$ in D. Therefore for configuration C', if $\mathrm{p}^D(D', D) < \epsilon/2$, then $C'D[i]$ is disjoint from $C'D[j]$ so $C' \in$free(D). Thus $\mathrm{B}^D(C, \epsilon/2) \subset$free($D$) so $C \in$Int(free(D)).

b. Suppose $C \in$Int(free(D)). By definition, there exists an $\epsilon > 0$ such that $\mathrm{B}^D(C, \epsilon) \in$ free(D). From lemma C.2.15 it follows that $C \in$ free(expand(D, ϵ)). But each region $D[I]$ is distance ϵ from the boundaries of expand($D[I], \epsilon$). Thus, since $C[I]$(expand($D[I], \epsilon$) does not overlap $C[J]$(expand($D[J], \epsilon$), it follows that $CD[I]$ and $CD[J]$ are at least 2ϵ apart. Hence $C \in$cfree(D).

∎

Proof of theorem 2.9

Lemma C.2.16: For any display D and $\delta > 0$, free(expand(D, δ)) $\subset$ cfree(D).

Proof: Let C be a configuration in free(expand(D, δ)). For any objects $I \neq J$, since $C[I]$(expand($D[I], \delta$)) does not overlap $C[J]$(expand($D[J], \delta$)), it follows that $CD[I]$ is at least 2δ from $CD[J]$. Hence, $C \in$cfree(D). ∎

Lemma C.2.17: For any display D and $\delta > 0$, cfree(contract(D, δ)) $\supset$ free(D).

Proof: Let C be any configuration in free(D). Since $CD[I]$ does not overlap $CD[J]$, and since every point in C(contract($D[I], \delta$)) is at least δ from the boundary of $CD[I]$, it follows that C(contract($D[I], \delta$)) is separated by at least 2δ from C(contract($D[J], \delta$)). Hence $C \in$ cfree(contract(D, δ)). ∎

Lemma C.2.18: For any normal, compact region $\mathbf{R}$, for any distance $\epsilon > 0$ there exists $\delta > 0$ such that d_H(contract($\mathbf{R}, \delta$),$\mathbf{R}$) $< \epsilon$.

Proof: For any point $\mathbf{x}$, let $\mathrm{B}(\mathbf{x}, \epsilon/2)$ be the open ball of radius $\epsilon/2$ about $\mathbf{x}$. Since the collection $\{ \mathrm{B}(\mathbf{x}, \epsilon/2) \mid \mathbf{x} \in \mathbf{R} \}$ is an open covering of $\mathbf{R}$ and $\mathbf{R}$ is compact, it has a finite subcovering $\{ \mathrm{B}(\mathbf{x}_i, \epsilon/2), i = 1 \ldots n \}$. Since $\mathbf{R}$ is normal, we may choose a point $\mathbf{y}_i \in \mathrm{B}(\mathbf{x}_i, \epsilon/2) \cap \mathrm{Int}(\mathbf{R})$. Thus, every point in $\mathbf{R}$ is within ϵ of one of the $\mathbf{y}_i$. Choose $\delta < \min_i \mathrm{d}(\mathbf{y}_i, \mathbf{R}^c)$; thus $\mathrm{B}(\mathbf{y}_i, \delta) \subset \mathbf{R}$, so $\mathbf{y}_i \in$ contract($\mathbf{R}, \delta$). Since every point of $\mathbf{R}$ is then within ϵ of a point in contract($\mathbf{R}$,δ), it follows that d_H(contract($\mathbf{R}, \delta$),$\mathbf{R}$) $< \epsilon$. ∎

Lemma C.2.19: For any normal, compact region $\mathbf{R}$, for any distance $\epsilon > 0$ there exists $\delta > 0$ such that d_{Hc}(expand($\mathbf{R}, \delta$),$\mathbf{R}$) $< \epsilon$.

Proof: Let $\mathbf{O}$ = expand($\mathbf{R}, 2\epsilon$). Let $\mathbf{U}$ be the closure of $\mathbf{O} - \mathbf{R}$. Then it is easily shown that $\mathbf{U}$ is compact and normal; that expand($\mathbf{R}, \epsilon$) = $\mathbf{O}-$ contract($\mathbf{U}, \epsilon$); and that d_{Hc}(expand($\mathbf{R}, \delta$),$\mathbf{R}$) = d_H(contract($\mathbf{U}, \delta$),$\mathbf{U}$). Applying lemma C.2.18 to the region $\mathbf{U}$ then completes the proof. ∎

Lemma C.2.20: If the function "free(·)" is continuous under expansion at display D, using the dual-Hausdorff metric d_{Hd} on displays, then it is continuous at D (without qualification).

Proof: Choose an arbitrary $\epsilon > 0$. Assume that the function "free(·)" is continuous under expansion at display D, using the metric d_{Hd} on displays. Then there exists a $\delta_1 > 0$ such that, for any expansion D' of D, if $\mathrm{d}_{Hd}(D, D') < \delta_1$, then p_H^D(free(D), free(D')) $< \epsilon$.

By theorem 2.1, there exists a $\delta_2 > 0$ such that, for any contraction D' of D, if $\mathrm{d}_{Hc}(D, D') < \delta_2$, then p_H^D(free(D), free(D')) $< \epsilon$.

Let $\delta = \min(\delta_1, \delta_2)$, and let D' be any display such that $\mathrm{d}_{Hd}(D, D') < \delta$. Since $\mathrm{d}_{Hc}(D, D') \leq \mathrm{d}_{Hd}(D, D') < \delta$, it follows from lemma 1.2 that $D' \supset$ contract(D, δ). Similarly, it follows from lemma 1.2 that $D' \subset$ expand(D, δ). Therefore, free(contract(D, δ)) $\subset$ free(D') $\subset$ free(expand(D, δ)). Hence, by lemma C.2.1, p_H^D(free(D), free(D')) $< \epsilon$. ∎

Lemma C.2.21: If the function "cfree(·)" is continuous under contraction at display D, using the dual-Hausdorff metric d_{Hd} on displays, then it is continuous at D (without qualification).

Proof: The proof is exactly analogous to the proof of lemma C.2.20. ∎

Theorem 2.9: The following three statements are equivalent:

A. The functions "free(·)" is continuous at display D, using the metric d_{Hd} on displays.

B. The function "cfree(·)" is continuous at display D, using the metric d_{Hd} on displays.

C. free(D) is equal to the closure of cfree(D).

Proof: We will show (I) that A implies C; (II) that C implies B; and (III) that B implies A.

I. Suppose that free($\cdot$) is continuous under expansion at display D. Using lemma 1.2, this implies that, for any $\epsilon > 0$, it is possible to choose $\delta > 0$ such that p_H^D(free(D),free(expand(D,δ)) $< \epsilon$. But, by lemma C.2.16 free(expand(D,δ)) $\subset$ cfree(D). Hence p_H^D(free(D),cfree(D)) $= 0$. That is, every point of free(D) is arbitrarily close to a point of cfree(D), so free(D) is a subset of the closure of cfree(D). Since free(D) is also closed and contains cfree(D), it follows that free(D) is equal to the closure of cfree(D).

II. Let $\epsilon > 0$. By theorem 2.1 and lemma 1.2, there exists $\delta > 0$ such that p_H^D(free(D), free(contract(D,δ)) $< \epsilon$. Trivially, cfree(contract(D,δ)) $\subset$ free(contract(D,δ)). Hence, by lemma C.2.1, p_H^D(free(D),cfree(contract(D,δ)) $< \epsilon$.

Now suppose that free(D) is equal to the closure of cfree(D). Then p_H^D(cfree(D),free(D)) $= 0$. By the triangle inequality,
p_H^D(cfree(D),cfree(contract(D,δ)) $\leq$ p_H^D(cfree(D),free(D)) + p_H^D(free(D),cfree(contract(D,δ)) $< \epsilon$.

Let D' be any contraction of D such that $\mathrm{d}_{Hd}(D',D) < \delta$. Then $\mathrm{d}_{Hc}(D',D) < \mathrm{d}_{Hd}(D',D) < \delta$. By lemma 1.2 p_H^D(cfree(D),cfree(D')) $\leq$ p_H^D(cfree(D),cfree(contract(D,δ)) $< \epsilon$. Thus, cfree($\cdot$) is continuous under contraction at D. By lemma C.2.21, cfree($\cdot$) is continuous at D.

III. Suppose that cfree($\cdot$) is continuous at display D, using the metric d_{Hd} on displays. Choose $\epsilon > 0$. Then there exists a δ_1 such that, if $\mathrm{d}_{Hd}(D',D) < \delta_1$, then p_H^D(cfree(D),cfree(D')) $< \epsilon/2$. Let D_1 be a uniform contraction of D such that $\mathrm{d}_{Hd}(D_1,D) < \delta_1$; such a D_1 exists by virtue of lemma C.2.18. Let E be any expansion of D such that $\mathrm{d}_{Hd}(D,E) < \delta_1$. So we have

cfree(D_1) $\supset$ free(D) by lemma C.2.17.
free(D) $\supset$ free(E) since E is an expansion of D.
free(E) $\supset$ cfree(E) trivially.

Hence, by lemma C.2.1, p_H^D(free(D),free(E)) $\leq$ p_H^D(cfree(D_1),cfree(E)) $\leq$
p_H^D(cfree(D_1),cfree(D)) + p_H^D(cfree(D),cfree(E)) $< \epsilon$.

Thus, free($\cdot$) is continuous under expansion. By lemma C.2.20, free(D) is continuous. ▮

Proof of theorem 2.10

Theorem 2.10: For any displays D, D', p_H^D(forbidden(D),forbidden(D')) $\leq \mathrm{d}_H(D,D')$. Consequently, the function "forbidden(D)" is continuous, if the topology of the domain is given by d_H and the topology of the range is given by p_H.

Proof: Let C be a configuration in forbidden(D). Then there exist objects $I \neq J$ and points $\mathbf{p},\mathbf{q}$ such that $\mathbf{p} \in \mathrm{Int}(D[I])$, $\mathbf{q} \in \mathrm{Int}(D[I])$, and $C[I](\mathbf{p}) = C[J](\mathbf{q})$. By definition of the Hausdorff distance between the two displays, there must exists points $\mathbf{p}' \in \mathrm{Int}(D'[I])$ and $\mathbf{q}' \in \mathrm{Int}(D'[J])$ such that $\mathrm{d}(\mathbf{p},\mathbf{p}') \leq \mathrm{d}_H(D,D')$ and $\mathrm{d}(\mathbf{q},\mathbf{q}') \leq \mathrm{d}_H(D,D')$.

Let $T[I]$ be the translation mapping point $\mathbf{x}$ to $\mathbf{x} + C[I](\mathbf{p}) - C[I](\mathbf{p}')$ and let $T[J]$ be the translation mapping point $\mathbf{x}$ to $\mathbf{x} + C[J](\mathbf{q}) - C[J](\mathbf{q}')$. Let C' be the configuration such that $C'[I] = T[I] \circ C[I]$; $C'[J] = T[J] \circ C[J]$; and $C'[K] = C[K]$ for $K \neq I,J$. Then $C'[I](\mathbf{p}') = C'[J](\mathbf{q}')$, so $C' \in$ forbidden(D'). Also, clearly, $\mathrm{p}^D(C,C') = \max(\mathrm{d}(\mathbf{p},\mathbf{p}'), \mathrm{d}(\mathbf{q},\mathbf{q}')) \leq \mathrm{d}_H(D,D')$.

Thus, to calculate the forbidden space of a display to an accuracy ϵ, it is at worst necessary to use shape approximations of accuracy ϵ. The forbidden space is a continuous function of the shapes of the objects involved, and the maximum ratio of the change in space to the change in shape is 1.

C.3: Proofs from section 3

Proof of theorem 3.1

Definition C.3.1: Let ϕ be a path in configuration space and let D be a display. We define overlap$^D(\phi)$ as the maximal value of overlap^D over ϕ and clearance$^D(\phi)$ as the minimal value of clearance^D over ϕ.

Lemma C.3.1: Let O be ordinarily connected, and let C be a connected component of O. Then there is an open set $U \supset C$ which is disjoint from any other connected component of O.

Proof: By definition of connectivity, given any two connected components C and C', there are disjoint open regions $V \supset C$ and $V' \supset C'$. By the definition of ordinary connectivity, there are only finitely many such C'. Therefore, the intersection of all these regions V is the desired open set. ▮

Lemma C.3.2: Let O be a compact space with metric μ which is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any $p, q \in O$, if $\mu(p, q) < \delta$ then there is a path ϕ from p to q through O such that ϕ is always within ϵ of p.

Proof: Suppose not. Then there exists an ϵ such that for any N we may choose points p_N, q_N for which $\mu(p_N, q_N) < 1/N$ but there is no path from p_N to q_N which remains within ϵ of p_N. Consider the sequence of pairs $\langle p_1, q_1\rangle, \langle p_2, q_2\rangle, \langle p_3, q_3\rangle, \ldots$ Since O is compact, there is a subsequence of the points p_k that converges to a limit point $p \in O$. Clearly, p is the limit of the corresponding q_k as well.

Let $U \subset$B(p, ϵ) be a neighborhood of p such that $U \cap O$ is ordinarily connected; let $S = U \cap O$; and let R be the connected component of S containing p. By lemma C.3.1, there is an open set U containing R and disjoint from $S - R$. Since U is open, it must contain all the points p_N and q_N in the subsequence converging to p for all sufficiently large N. Therefore, these points must be in R. But R is path-connected, and it lies in a sphere of radius $\epsilon/2$, so there must be a path from p_N to q_N that stays within ϵ of p_N. This complete the contradiction. ▮

Lemma C.3.3: Let O be a compact space with metric μ which is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ for which the following holds: Let Q be any space within Hausdorff distance δ of O: $\mu_H(O, Q) < \delta$. Let ψ be any path through Q. Then there exists a path ϕ through O such that the path distance $\mu_t(\phi, \psi) < \epsilon$.

Proof: By lemma C.3.2, we can choose a value δ_1 such that, if the distance between two points $q, r \in O$ is less than δ_1, then there is a path from q to r that lies within $\epsilon/3$ of q. I claim that if δ is chosen less than $\min(\epsilon/3, \delta_1/3)$ then the theorem holds.

Let Q and ψ be chosen as above. We can choose a finite sequence of points $\langle p_0, p_1 \ldots p_k\rangle$ on ψ such that p_0 is the starting point of ψ, p_k is the ending point of ψ, and the section of the path ψ between p_i and p_{i+1} is always within $\delta/3$ of p_i. Since the Hausdorff distance from O to Q is less than δ, we may choose points $q_0, q_1 \ldots q_k$ in O such that each q_i is within δ of p_i. Therefore, the distance from q_i to q_{i+1},

$$\mu(q_i, q_{i+1}) \leq \mu(q_i, p_i) + \mu(p_i, p_{i+1}) + \mu(p_{i+1}, q_{i+1}) < \delta_1$$

By definition of δ_1, therefore, there is a path from q_i to q_{i+1} that remains within $\epsilon/3$ of q_i. Therefore, any point on the path ψ between p_i to p_{i+1} lies within ϵ of the corresponding point between q_i and q_{i+1}. We now string together the paths between q_i and q_{i+1}, and the proof is complete. ▮

Let $\mathrm{p}_t(\phi, \phi')$ be the path distance between paths ϕ and ϕ'. Applying the Hausdorff construction over the space of paths, we can define a distance measure p_{tH} between one set of paths and another. This, in turn, defines a topology over the space of sets of paths.

Theorem 3.1: (Analogue of theorem 2.1.) Let D be a display such that free(D) is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that, if E is a contraction of D and the complement-Hausdorff distance $\mathrm{d}_{Hc}(D, E) < \delta$ then $\mathrm{p}^D_{tH}(\text{paths}(\text{free}(D)), \text{paths}(\text{free}(E))) < \epsilon$.

Proof: What this states is, that for any D and ϵ, δ can be chosen such that, if E is a contraction of D and $\mathrm{d}_{Hc}(D, E) < \delta$ then (a) every path in paths(free(D)) is near a path in paths(free(E)) and (b) every path in paths(free(E)) is near a path in paths(free(D)). Part (a) is trivial, since paths(free(D)) is a subset of paths(free(E)).

We establish part (b) in two steps. First, consider only the space of paths that remain in the basic configuration region. Since this region is compact (lemma C.2.7), (b) follows immediately from lemmas C.3.3 and theorem 2.1.

Second, we use lemma C.2.8 to reduce configuration space as a whole to the basic configuration region, as follows: Assume w.l.o.g. that $\epsilon < 1$. Using the first step above, choose a suitable δ for ϵ and D. Using lemma C.2.8, construct a function $\Gamma^D(C)$ from configuration space into the basic configuration region over D and let $\Theta^D(C)$ be the transformation from C to $\Gamma^D(C)$. Let E be any display such that $\mathrm{d}_{Hc}(D, E) < \delta$ and let ϕ be any path in free(E). Let $\phi_1(T) = \Gamma^D(\phi(T))$, and let $\theta(T) = \Theta^D(\phi(T))$. By lemma C.2.8, ϕ_1 stays in the compact basic configuration region. Also by lemma C.2.8, since Γ^D preserves the function overlap^E, overlap$^E(\phi_1)$ = overlap$^E(\phi) = 0$, so ϕ_1 is in paths(free(E)). By lemma C.3.3, there is a path ψ_1 in paths(free(D)) such that $\mathrm{p}^D_t(\psi_1, \phi_1) < \epsilon$. Now, let $\psi(T) = (\theta(T))^{-1}(\psi_1(T))$. It follows immediately from lemma C.2.8 that $\mathrm{p}(\psi, \phi) < \epsilon$ and that $\phi \in$paths(free(D)).∎

Proof of theorem 3.2

Theorem 3.2: (Analogue of theorem 2.2.) Let D be a display and let $\delta > 0$. Let $\epsilon = \mathrm{p}^D_{tH}(\text{paths}(\text{free}(D)),$ paths(free(contract(D, δ))). If F is a contraction of D and $\mathrm{d}_{Hc}(F, D) < \delta$, then p^D_{tH}(paths(free(D)),paths(free(F))) $\leq \epsilon$.

Proof: Immediate from lemma 1.2 and the facts that "paths" is non-decreasing and that "free" is non-increasing, so the composition, "path(free($\cdot$))" is non-increasing.

Proof of Theorem 3.3

We begin by analyzing the tracking of paths in discrete graph structures. We then show that the continuous problem can be mapped to a discrete problem with any desired accuracy.

Definition C.3.2: Let G and H be undirected graphs. Let $\phi = \langle \phi[0] \dots \phi[k] \rangle$ be a path of length k through G; that is, each $\phi[i]$ is a vertex, and there is an edge in G from $\phi[i]$ to $\phi[i + 1]$. (We will count the length of a path as the number of edges traversed, not the number of vertices.) A *tracking* of ϕ through H consists of two parts:

i. A path $\psi = \langle \psi[0] \dots \psi[m] \rangle$ through H, where $m \geq k$.

ii. A non-decreasing function $\sigma(i)$ from the range $0 \dots m$ onto the range $0 \dots k$. The requirement that the function σ is "onto" means that for every $i \in 0 \dots k$ there is a $j \in 0 \dots m$ such that $\sigma(j) = i$.

The function σ thus carves up the range $[0 \dots m]$ into $k + 1$ non-empty intervals, and maps the first range to 0, the second range to 1 ...and the kth range to $k + 1$. For instance, if $k = 3$ and $m = 6$ then one could map [0,1,2] to 0; [3] to 1; [4,5] to 2, and [6] to 3.

Definition C.3.3: Let G and H be undirected graphs. Let $F(U, V)$ be a function mapping vertices $U \in G$ and $V \in H$ to a positive real value, the distance from U to V. Let ϕ be a path through G

and let $\langle \psi, \sigma \rangle$ be a tracking of ϕ through H. Then the *tracking distance* from $\langle \psi, \sigma \rangle$ to ϕ, denoted as $F_t(\phi, \psi, \sigma)$ is the maximum distance from a point $\psi[i]$ to the corresponding point $\phi[\sigma(i)]$.

$$F_t(\phi, \psi, \sigma) = \max_{i=0...m} F(\phi[\sigma(i)], \psi[i])$$

Definition C.3.4: Let ϕ be a path through G. Let W and Z be vertices in H. Then a *best tracker* of ϕ through H from W to Z is a tracking $\langle \psi, \sigma \rangle$ of ϕ through H such that ψ starts at W and ends at Z for which $F_t(\phi, \psi, \sigma)$ is minimal. (Since the minimum must be one of the finitely many values in matrix F, it must be attained, so there is always a best tracker.) This minimal value of $F_t(\phi, \psi, \sigma)$ is called the *minimal tracking distance* of ϕ through H from W to Z and is denoted "min_track(ϕ, H, W, Z, F)."

Definition C.3.5: Let U and V be vertices in G; let W and Z be vertices in H; and let k be a positive integer. We say that ϕ is a path *of length at most k from U to V hardest to track through H from W to Z* if the minimum value of min_track(α, H, W, Z) over all paths α of length at most k from U to V through G is attained for $\alpha = \phi$. This maximal value is denoted "hard_track(G, U, V, H, W, Z, k, F)".

Example C.3.1: Consider the graphs shown in figure 19. Let G be the graph of vertices A through E, in solid lines, and let H be the graph of vertices Q through Z in dotted lines. Assume that there is a self-loop, not shown, from every vertex in H to itself. The numbers on the vertices indicate the coordinates. For ease of calculation, we will take the distance metric F to be the Manhattan distance; that is, $F(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) = \mid x_2 - x_1 \mid + \mid y_2 - y_1 \mid$

Then we have the following:

Let $\phi = \langle$D,E,B$\rangle$, let $\psi = \langle$Y,R,Z,U$\rangle$ and let $\sigma(\langle 1, 2, 3, 4 \rangle) = \langle 1, 2, 2, 3 \rangle$. That is, σ associates Y with D, R and Z with E, and U with B. Then $F_t(\phi, \psi, \sigma) = \max(F$(S,B), F(R,A), F(Q,A), F(U,B)) $= F$(R,E) $= 5$. ψ, σ is, indeed, a best tracker of ϕ in H.

The best tracker through H for the path $\langle$B,A,D$\rangle$ is the path $\langle$S,Q,Y$\rangle$ with tracking distance 2. (The correspondence is the obvious one.) The best tracker through H for the path $\langle$C,B,A,D$\rangle$ is the path $\langle$V,U,T,W$\rangle$ with tracking distance 4.

A hardest path in G to track is the path $\langle$C,B,A,D,E,B$\rangle$. Note that if you try a simple greedy strategy starting at V, you can closely track the beginning of the path $\langle$C,B,A,D,E$\rangle$ with path $\langle$V,U,T,W,X$\rangle$, but then you are stuck; there is no way to continue to track the path closely. In fact, a best tracker is the path $\langle$V,U,T,W,W,T$\rangle$ under the correspondence V→C, U→B, T→A, W→D, W→E, T→B, with tracking distance 8.

Definition C.3.6: The *semi-Hausdorff tracking distance* from graph H to graph G relative to distance F, denoted $F_{tHs}(H, G)$ is the maximum over all paths ϕ through G of the minimum over all trackings $\langle \psi, \sigma \rangle$ through H of the tracking distance $F_t(\phi, \psi, \sigma)$. Equivalently, it is equal to the maximum over all k and all starting and ending vertices $U, V \in H$ of the minimum over all starting and ending vertices $W, Z \in G$ of hard_track(G, U, V, H, W, Z, k, F).

Definition C.3.7: The *Hausdorff tracking distance* between G to H relative to F, denoted "$F_{tH}(G, H)$", is defined as $\max(F_{tHs}(G, H), F_{tHs}(H, G))$. the maximum of the semi-Hausdorff tracking distance from H to G and the semi-Hausdorff tracking distance from G to H.

We define "splicing" two paths, the first of which stops where the second starts, and "splicing" two tracking functions, in the obvious way.

Definition C.3.8: Let $\phi_1[0 \dots p]$ and $\phi_2[0..q]$ be paths. If $\phi_1[p] = \phi_2[0]$, we will say that ϕ_1 meets ϕ_2. In this case, the *splice* of paths ϕ_1 and ϕ_2, denoted $\phi_1; \phi_2$, is the path ϕ of length $p + q$ such that $\phi[i] = \phi_1[i]$ for $i \in [0, p]$ and $\phi[i] = \phi_2[i - p]$ for $i \in [p, p + q]$.

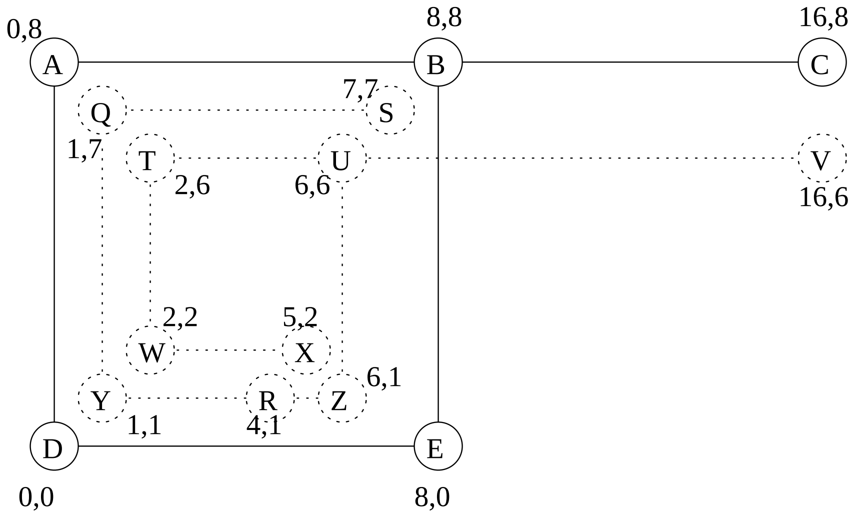


Figure 19: Tracking paths through graphs

Definition C.3.9: Let $\sigma_1(i)$ be a function from $[0, k]$ to $[0, m]$ and let $\sigma_2(i)$ be a function from $[0, p]$ to $[0, n]$. Then the *splice* of σ_1 and σ_2, denoted $\sigma_1; \sigma_2$, is defined to be the function σ from $[0, k+p]$ to $[0, m+n]$ such that $\sigma(i) = \sigma_1(i)$ for $i \in [0, k]$, and $\sigma(i) = \sigma_2(i-k) + m$ for $i \in [k, k+p]$.

Lemma C.3.4: Let $\langle \psi_1, \sigma_1 \rangle$ be a tracker for ϕ_1; let $\langle \psi_2, \sigma_2 \rangle$ be a tracker for ϕ_2; and assume ψ_1 meets ψ_2 and that ϕ_1 meets ϕ_2. Then $F_t(\phi_1; \phi_2, \psi_1; \psi_2, \sigma_1; \sigma_2) = \max(F_t(\phi_1, \psi_1, \sigma_1), F_t(\phi_2, \psi_2, \sigma_2))$

Proof: Immediate from definitions C.3.3, C.3.8, and C.3.9. ∎

Lemma C.3.5: Let ϕ_1 and ϕ_2 be paths through G that meet, of lengths p and q respectively; and let $\phi = \phi_1; \phi_2$. Let W, Z be vertices in H. Let $\langle \psi, \sigma \rangle$ be a best tracker for ϕ through G from W to Z. Let n be the length of ψ, and let $m \in [0, n]$ be an index such that $\sigma[m] = p$. Thus σ associates the path $\psi[0 \ldots m]$ with path ϕ_1 and path $\psi[m \ldots n]$ with path ϕ_2. Let $X = \psi[m]$. Let $\langle \alpha_1, \gamma_1 \rangle$ be a best tracker for ϕ_1 from W to X and let $\langle \alpha_2, \gamma_2 \rangle$ be a best tracker for ϕ_2 from X to Z. Then $\langle \alpha_1; \alpha_2, \gamma_1; \gamma_2 \rangle$ is a best tracker for ϕ through H from W to Z.

Proof: Let $\langle \beta, \lambda \rangle$ be any tracker for ϕ through H from W to Z. Since $\langle \psi, \sigma \rangle$ is the best tracker for ϕ through H from W to Z, we have $F_t(\phi, \psi, \sigma) \leq F_t(\phi, \beta, \lambda)$. By lemma C.3.4, $F_t(\phi, \psi, \sigma) = \max(F_t(\phi_1, \psi_1, \sigma_1), F_t(\phi_2, \psi_2, \sigma_2))$. Since $\langle \alpha_1, \gamma_1 \rangle$ is a best tracker for ϕ_1 from W to X, we have $F_t(\phi_1, \alpha_1, \gamma_1) \leq F_t(\phi_1, \psi_1, \sigma_1)$ and similarly $F_t(\phi_2, \alpha_2, \gamma_2) \leq F_t(\phi_2, \psi_2, \sigma_2)$. By lemma C.3.4, $F_t(\phi, \alpha, \gamma) = \max(F_t(\phi_1, \alpha_1, \gamma_1), F_t(\phi_2, \alpha_2, \gamma_2))$. Putting these together, we get $F_t(\phi, \alpha, \gamma) \leq F_t(\phi, \beta, \lambda)$. ∎

Lemma C.3.6: Let ϕ_1 and ϕ_2 be paths through G that meet, and let $\phi = \phi_1; \phi_2$. Then for any vertices W, Z in H,
$\text{min\_track}(\phi, H, W, Z, F) = \min_{X \in H} \max(\text{min\_track}(\phi_1, H, W, X, F), \text{min\_track}(\phi_2, H, X, Z, F))$.

Proof: Immediate from lemmas C.3.4 and definition C.3.5. ∎

Lemma C.3.7: Let ϕ_1 and ϕ_2 be paths through G that meet, of lengths p and q respectively; and let $\phi = \phi_1; \phi_2$. Let $U = \phi[0]$, $T = \phi[p]$, and $V = \phi[p+q]$. Let W, Z be vertices in H. Suppose that ϕ is a path of length at most $p+q$ from U to V hardest to track by a path from W to Z through H. Let $\langle \psi, \sigma \rangle$ be the best tracker for ϕ from W to Z through H. Let m be an index such that $\sigma(m) = p$; and let $X = \psi(m)$. Let α_1 be a path of length at most p from U to T through G hardest to track by a path from W to X in H; and let α_2 be a path of length at most q from U to T through G hardest to track by a path from X to Z in H. Then $\alpha_1;\alpha_2$ is a path of length at most $p+q$ from U to V hardest to track by a path from W to Z through H.

Proof: Let Δ be any path from U to V through G of length $p+q$. Since ϕ is a path of length $p+q$ from U to V hardest to track by a path from W to Z through H, we know that $\text{min\_track}(\phi, H, W, Z, F) = F_t(\phi, \psi, \sigma) \geq \text{min\_track}(\Delta, H, W, Z, F)$. By lemma C.3.4, $F_t(\phi, \psi, \sigma)$ $= \max(F_t(\phi_1, \psi_1, \sigma_1), F_t(\phi_2, \psi_2, \sigma_2))$. Let $\langle \theta_1, \tau_1 \rangle$ be the best tracker for ϕ_1 from W to X, and let $\langle \theta_2, \tau_2 \rangle$ be the best tracker for ϕ_2 from X to Z. By lemma C.3.5, $F_t(\phi, \psi, \sigma) = \max(F_t(\phi_1, \theta_1, \tau_1), F_t(\phi_2, \theta_2, \tau_2))$

Let $\langle \gamma_1, \eta_1 \rangle$ be the best tracker for α_1 from W to X, and let $\langle \gamma_2, \eta_2 \rangle$ be the best tracker for α_2 from X to Z. Since α_1 and α_2 are the hardest paths to track, we have $F_t(\alpha_1, \gamma_1, \eta_1) \geq F_t(\phi_1, \theta_1, \tau_1)$ and $F_t(\alpha_2, \gamma_2, \eta_2) \geq F_t(\phi_2, \theta_2, \tau_2)$. But by lemma C.3.5, $\langle \gamma_1; \gamma_2, \eta_1; \eta_2 \rangle$ is a best tracker for α from W to Z through H; and by lemma C.3.4 $F_t(\alpha, \gamma_1; \gamma_2, \eta_1; \eta_2) = \max(F_t(\alpha_1, \gamma_1, \eta_1), F_t(\alpha_2, \gamma_2, \eta_2))$. Putting all these together, we get that $\text{min\_track}(\alpha, H, W, Z, F) = F_t(\alpha, \gamma 1; \gamma_2, \eta_1; \eta_2) \geq \text{min\_track}(\Delta, H, W, Z, F)$. Since Δ was an arbitrary path from U to V, this means that α is a hardest path from U to V through G to track from W to Z through H. ∎

Lemma C.3.8: For any vertices $U, V \in G$, vertices $W, Z \in H$ and integers $p, q > 0$,
$\text{hard\_track}(G, U, V, H, W, Z, p+q, F) =$
$\max_{T \in G} \min_{X \in H} \max(\text{hard\_track}(G, U, T, H, W, X, p, F), \text{hard\_track}(G, T, V, H, X, Z, q, F))$

Proof: Straightforward from the proof of lemma C.3.7. ∎

Lemmas C.3.4 through C.3.8 above, which just deal with maxima and minima over lists, are "trivial" in the sense that many mathematical papers would be content just to write down lemma C.3.8, with the comment, "The proof is trivial." Nonetheless, the intermixing of min's and max's can get confusing enough that I thought it worthwhile to spell out the proofs.

Lemma C.3.9: Let g and h be the number of vertices in G and H, and let $M = 2gh+1$. Let ϕ be a path through G from U to V with length is greater than or equal to M. For any vertices $W, Z \in H$, there exists a path α through G from U to V shorter than ϕ such that $\text{min\_track}(\alpha, H, W, Z, F) =$ $\text{min\_track}(\phi, H, W, Z, F)$

Proof: Let $\langle \psi, \sigma \rangle$ be a best tracker for ϕ from W to Z. Let m be the length of ψ; then $m \geq \mid \phi \mid$. We can assume without loss of generality that if $i \neq j$ and $\sigma(i) = \sigma(j)$ then $\psi[i] \neq \psi[j]$; that is, that the tracker does not have a cycle $\psi[i], \psi[i+1] \ldots \psi[j] = \psi[i]$, all of which are mapped to the same $\phi[\sigma(i)]$. Let us now consider the pairs of corresponding vertices $\langle \phi[\sigma(1)], \psi[1] \rangle$ $\langle \phi[\sigma(2)], \psi[2] \rangle$ $\ldots \langle \phi[\sigma(m)], \psi[m] \rangle$. Since there are only gh different pairs of vertices, and since $m \geq \mid \phi \mid \geq 2gh+1$, it follows there must be some pair that appears at least three times in the above list. So let us suppose that for some $p < q < r$, $\psi[p] = \psi[q] = \psi[r]$ and $\phi[\sigma(p)] = \phi[\sigma(q)] = \phi[\sigma(r)]$.

Divide the path ϕ into 4 pieces; $\phi_1 = \phi([1 \ldots \sigma(p)]$, $\phi_2 = \phi[\sigma(p) \ldots \sigma(q)]$, $\phi_3 = \phi[\sigma(q) \ldots \sigma(r)]$, $\phi_4 = \phi[\sigma(r) \ldots \mid \phi \mid]$. (Either the first or the last of these intervals or both may be empty. The middle two are non-empty.) Let $\langle \beta_1, \tau_1 \rangle$, $\langle \beta_2, \tau_2 \rangle$, $\langle \beta_3, \tau_3 \rangle$, $\langle \beta_4, \tau_4 \rangle$ be the best trackers for these between $\psi[1]$ and $\psi[p]$; $\psi[p]$ and $\psi[q]$; $\psi[q]$ and $\psi[r]$; $\psi[r]$ and $\psi[m]$ respectively. Let $\beta = \beta_1; \beta_2; \beta_3; \beta_4$ and $\tau = \tau_1; \tau_2; \tau_3; \tau_4$. Then by the same argument as in lemma C.3.5, $\langle \beta, \tau \rangle$ is a best tracker for ϕ between W and Z.

Suppose that $F_t(\phi_2, \beta_2, \tau_2) \geq F_t(\phi_3, \beta_3, \tau_3)$. Then let $\alpha = \phi_1; \phi_2; \phi_4$. Let $\gamma = \beta_1; \beta_2; \beta_4$, and let $\eta = \tau_1; \tau_2; \tau_4$. Then clearly α is a path shorter than ϕ from U to V through G; γ is a path from W to Z in H; and $F_t(\alpha, \gamma, \eta) = F_t(\phi, \psi, \sigma)$. Moreover, by the same argument as in lemma C.3.5, γ is a best tracker for α from W to Z. Thus $\text{min\_track}(\alpha, H, W, Z, F) = F_t(\alpha, \gamma, \eta) = F_t(\phi, \psi, \sigma) =$ $\text{min\_track}(\phi, H, W, Z, F)$.

If the reverse inequality holds, $F_t(\phi_2, \beta_2, \tau_2) \leq F_t(\phi_3, \beta_3, \tau_3)$, then the argument is exactly analogous, using the path $\alpha = \phi_1; \phi_3; \phi_4$, and likewise for γ and η. ▮.

Now, we can put all this together into an algorithm to compute the semi-Hausdorff distance between two graphs.

Lemma C.3.10: The function "path_trace1(G, H, F)", shown in table 2, computes the semi-Hausdorff distance from H to G relative to F.

Proof: The value of B[U,U,W,Z,1] is supposed to be the value of the hardest path from U to U of length at most 1 to track from W to Z. Since the only such path is just the vertex U, this then is the problem of finding the lowest cost path in H from W to Z, where the "cost" of a path is measured as the maximum distance from U. The code in path_trace1 adapts Floyd's shortest path algorithm to compute this.

If there is an edge U−V, then the value of B[U,U,W,Z,1] is supposed to be the value of the best tracker from W to Z for the edge U−V. This must have the form of a path from W to X mapped into U, followed by an arc from X to Y, followed by a path from Y to Z mapped into V. The code takes the minimum over all edges X−Y in H of the value of such a path.

The main loop of the program then uses lemma C.3.8 to compute the value of B[U,V,W,Z,2^k] from B[U,T,W,X,2^{k-1}] and B[T,V,X,Z,2^{k-1}]. It follows from lemma C.3.9 that the values of hard_track(G,U,V,H,W,Z,m,F) does not change for $m > M$. Therefore it suffices to repeat the main loop up to $k = \lceil log(M) \rceil$. ▮

We can now turn to the continuous problem.

Definition C.3.10: Let R be a region in a space with metric μ. A labelled undirected graph G is a *overlap graph* for R if it satisfies the following:

- Each vertex U of G is labelled by a subset of R containing U, denoted "reg(U)", and by a point in reg(U), denoted "center(U)". The region reg(U) must be path-connected and open in R (i.e. the intersection of R with some open set O.)
- The union of reg(U) over all vertices U in G is equal to R.
- There is an edge in G connecting U and V if and only if reg(U) overlaps reg(V).
- If U and V are distinct vertices of R, then center(U) $\neq$ center(V).

For a vertex U of G, the *radius* of U is defined as $\sup_{x \in \text{reg}(U)} \mu(x, \text{center}(U))$. The *mesh size* of G is defined as the maximum of the radius of U over all vertices U in G.

The gap between continuous paths through region R and paths through the overlap graph G is bridged by functions of time that hop between the centers of vertices.

Definition C.3.11: Let R be a region and let G be an overlap graph for R. A *center-hopping function* is a function $\alpha(t)$ from the interval [0,1] to the centers of the vertices of G such that the following hold:

- There exists a partition of the time interval [0,1] into k subintervals $T_1 = [t_0 = 0, t_1), T_2 = [t_1, t_2), \ldots T_{k-1} = [t_{k-2}, t_{k-1}), T_k = [t_{k-1}, t_k = 1]$, where $t_{i-1} < t_i$ for $i = 1 \ldots k-1$ and $t_{k-1} \leq t_k$.
- Over any interval T_i, α is constant and equal to the center of one of the vertices of G.
- For any two consecutive intervals T_i, T_{i+1}, there is an edge in G from $\alpha(T_i)$ to $\alpha(T_{i+1})$.

```
function path_trace1(in G,H : graphs; F : positive real array [vertex in G, vertex in H])
                    return real;

const M = 2 | G | · | H |;

var B[U,V: vertices in G; W,Z : vertices in H; k : 1 .. M] : real array;

/* B[U,V,W,Z,k] = hard_track(G,U,V,H,W,Z,2^(k−1),F) */

/* Initialize B[U,U,W,Z,1] i.e. the minimum over all paths from W to Z of the maximum
of the distance from U. The algorithm is exactly analogous to Floyd's shortest paths algorithm,
only taking the cost of a path to be the max distance from U rather than the sum of the costs of the edges. */

for U in G
    for W in H do B[U,U,W,W,1] := F[U,W];
    for W in H for Z in H
        if W−Z in H then B[U,U,W,Z,1] := max(F[U,W],F[U,Z])
                    else B[U,U,W,Z,1] := ∞;
        for X in H
            for W in H for Z in H
                B[U,U,W,Z,1] := min(B[U,U,W,Z,1], max(B[U,U,W,X,1], B[U,U,X,Z,1]));

/* Initialize B[U,V,W,Z,1] i.e. the best tracking distance for the arc U−V by a path
through H from W to Z. The best tracking path must consist of a path from W to X, mapped into U;
then an arc X−Y; then a path from Y to Z mapped into V; for the best possible arc X−Y. */

for U in G for V in G for W in H for Z in H
    if U−V in G
      then B[U,V,W,Z,1] = min_{X−Y∈G} max(B[U,U,W,X,1], B[V,V,Y,Z,1])
      else if U≠V then B[U,V,W,Z,1] = 0.

/* Use lemma C.3.10 to compute B[U,B,W,Z,k] for k ≥ 2 */

for k := 2 to ⌈log(M)⌉ + 1 do
    for U in G for V in G for W in H for Z in H
        do B[U,V,W,Z,k] := max_{T∈G} min_{X∈H} max(B[U,T,W,X,k−1], B[T,V,X,Z,k − 1])

return max_{U∈G,V∈G} min_{W∈H,Z∈H} B[U,V,W,Z,M]

end path_trace1.
```

Table 2: Computing the hardest paths to track

The partition $T_1 \dots T_k$ is said to be *induced* by α. Thus, the function ϕ hops through the centers of the vertices of G, moving across edges of G from one vertex to a neighboring vertex, and staying a finite time at each vertex except possibly the last. (The requirement that the intervals in the partition be half-open on the right is, of course, just a matter of notational convenience.)

Definition C.3.12: Let R be a region; let G be an overlap graph for R; let ϕ be a continuous path through R; and let α be a center-hopping function through G. We say that α is a *discretization* of ϕ through G if, for all $t \in [0,1]$ there is a vertex U in G such that $\phi(t) \in$reg(U) and $\alpha(t) =$center(U).

Lemma C.3.11: Let $R, G, \phi(t)$ be as in definition C.3.12, and let $\alpha(t)$ be a discretization of ϕ through G. Let δ be the mesh size of G. Then $\mu_t(\phi, \alpha) \leq \delta$.

Proof: Immediate from the definitions. ∎

Lemma C.3.12: Let R, G, and ϕ be as in Definition C.3.12. Then there exists a discretization of ϕ through G.

Proof: For each $t \in [0,1]$, choose a vertex U_t such that reg(U_t) contains $\phi(t)$, and let $O_t \subset [0,1]$ be the connected component of $\phi^{-1}(\text{reg}(U_t))$ containing t. Thus, O_t is an interval open in [0,1] such that $t \in O_t$ and $\phi(O_t) \subset$reg(U_t). The collection $\{ O_t \}$ is an open covering of the compact interval [0,1] and thus has a finite subcovering. Let $\{ P_1, P_2 \dots P_k \}$ be a minimal subcovering, ordered in increasing order of upper bounds. Since the collection is minimal, there must be a time in each of the P_i that is not in any of the other P_j. Therefore, P_1 must have the form $[s_1 = 0, t_1)$, for some time t_1; P_2 must have the form (s_2, t_2) for some times s_2, t_2 such that $s_1 < s_2 < t_1 < t_2$; in general, for $i = 2 \dots k-1$, P_i must have the form (s_i, t_i) where $s_{i-1} < s_i < t_{i-1} < t_i$; and P_k has the form $(s_k, 1]$ where $s_{k-1} < s_k < t_{k-1} \leq 1$.

We can therefore let the partition be the sequence of intervals $T_1 = [0, t_1), T_1 = [t_1, t_2) \dots T_k = [t_{k-1}, 1]$; and let $\alpha(T_i)$ be the center of the vertex originally associated with P_i. It remains to be shown that $\alpha(T_i)$ and $\alpha(T_{i+1})$ are centers of vertices connected by an edge in G. Let U, V be the regions such that center(U)=$\alpha(T_i)$ and center(V)=$\alpha(T_{i+1})$. Since $\phi(t_i)$ must be an element of both reg(U) and of reg(V), and since all the regions are open, it follows that reg(U) overlaps reg(V), and hence there is an edge from U to V in G. ∎

Lemma C.3.13: Let R be a region and let G be an overlap graph for R. Let $\alpha(t)$ be a center-hopping function in G. Then there exists a continuous path ϕ through R such that α is a discretization of ϕ.

Proof: Let $[0, t_1), [t_1, t_2) \dots [t_{k-1}, 1]$ be the partition induced by α. For $i = 1 \dots k-1$ let p_i be a point in the intersection reg$(\alpha[i]) \cap$ reg$(\alpha[i+1])$; let $p_0 = p_1$ and $p_k = p_{k-1}$. Then over the interval $[t_i, t_{i+1}]$, let $\phi(t)$ be a continuous path through reg$(\alpha[i])$ from p_i to p_{i+1}; such a path exists since reg$(\alpha[i])$ is path-connected. ∎.

Lemma C.3.14: Let R and S be two regions in a space with metric μ. Let G be an overlap graph for R with mesh size δ_G and let H be an overlap graph for S with mesh size δ_H. Let $F[U,V] = \mu(\text{center}(U),\text{center}(V))$. Let $\phi(t)$ be a path through R and let $\psi(t)$ be a path through S. Then there exist a path θ through G, and a tracker $\langle \tau, \sigma \rangle$ for θ through H such that $\mid \mu_t(\phi, \psi) - F_t(\theta, \tau, \sigma) \mid \leq \delta_G + \delta_H$.

Proof: Let $\alpha(t)$ be a discretization of ϕ through G, and let $\beta(t)$ be a discretization of ψ through H. Since μ_t is a metric on paths,

$$\mid \mu_t(\phi, \psi) - \mu_t(\alpha, \beta) \mid \leq \mu_t(\phi, \alpha) + \mu_t(\psi, \beta) \leq \delta_G + \delta_H$$

by lemma C.3.11.

Let $T_1 \dots T_k$ be the partition of [0,1] induced by α and let $S_1 \dots S_m$ be the partition of [0,1] induced by β. Let $P_1 \dots P_n$ be the collection of all non-empty intersections $T_i \cap S_j$ sorted in order; it is easily shown that P_i is a partition of the unit interval. Let θ be the path through G of the successive values of α. Let τ be the path through H of the successive values of $\beta(P_i)$. Let $\sigma(i)$ be the index j such that $P_i \subset T_j$. For any vertices U, V, let $F[U,V] = \mu(\text{center}(U),\text{center}(V))$. It is immediate from definitions C.3.1 and C.3.3 that $\mu_t(\alpha, \beta) = F_t(\theta, \tau, \sigma)$. Substituting in the formula above gives the desired result. ▮

Lemma C.3.15: Let R and S be two regions in a space with metric μ. Let G be an overlap graph for R with mesh size δ_G and let H be an overlap graph for S with mesh size δ_H. Let $F[U,V] = \mu(\text{center}(U),\text{center}(V))$. Let θ be a path through G and let $\langle \tau, \sigma \rangle$ be a tracker for θ through H. Then there exist continuous paths ϕ through R and ψ through S such that $\mid \mu_t(\phi, \psi) - F_t(\theta, \tau, \sigma) \mid \leq \delta_G + \delta_H$.

Proof: Let m be the length of τ, and let $T_1 \ldots T_m$ be a partition of [0,1] into m sub-intervals. Define the center-hopping functions $\alpha(T_i) = \theta[\sigma(i)]$ and $\beta(T_i) = \tau[i]$. Then it is immediate from the definitions that $\mu_t(\alpha, \beta) = F_t(\theta, \tau, \sigma)$. Construct continuous paths ϕ from α and ψ from τ as in lemma C.3.13. As in lemma C.3.14, we have the inequality

$$\mid \mu_t(\phi, \psi) - \mu_t(\alpha, \beta) \mid \leq \mu_t(\phi, \alpha) + \mu_t(\psi, \beta) \leq \delta_G + \delta_H$$

so the result follows immediately. ▮

Lemma C.3.16: Let R and S be two regions in a space with metric μ. Let G be an overlap graph for R with mesh size δ_G and let H be an overlap graph for S with mesh size δ_H. Let $F[U,V] = \mu(\text{center}(U),\text{center}(V))$. Then $\mid \mu_{tH}(R, S) - F_{tH}(G, H) \mid \leq \delta_G + \delta_H$.

Proof: Immediate from definition 3.2 and lemmas C.3.14 and C.3.15. ▮

Lemma C.3.17: Let U be a bounded, algebraic region in R^k and let $\delta > 0$. Then it is possible to compute an overlap graph for U of mesh size at most δ.

Proof: Let $\delta_1 = \delta/\sqrt{k}$, so that δ is equal to the diameter of a k-dimensional cube of side δ_1. Construct all the k-dimensional open cubes of the form

$$[\frac{m_1}{2}\delta_1, \frac{m_1+2}{2}\delta_1] \times [\frac{m_2}{2}\delta_1, \frac{m_2+2}{2}\delta_1] \times \ldots [\frac{m_k}{2}\delta_1, \frac{m_k+2}{2}\delta_1]$$

that intersect U, where $m_1, m_2 \ldots m_k$ are integers. For each such cube C, compute the connected components of $C \cap U$ (there can be only finitely many). For each connected component R, construct a vertex U_R; set $\text{reg}(U_R) = R$, and set $\text{center}(U_R)$ to be any point inside R. For each pair of connected R and S, construct an edge from U_R to U_S if and only if $R \cap S$ is non-empty.

It is shown in (Mishra, 1993) that all the above calculations can be carried out exactly for semi-algebraic regions. ▮

Lemma C.3.18: Any semi-algebraic region is locally ordinarily connected.

Proof: See (Mishra, 1993) for a proof that any semi-algebraic region is ordinarily connected. For any semi-algebraic region O and point $p \in O$, the spherical neighborhood $\mathrm{B}(p, \epsilon)$ is semi-algebraic and therefore ordinarily connected. ▮

Theorem 3.3: (Analogue of theorem 2.3.) Let D be a semi-algebraic display and let $\delta > 0$. Let ϵ be the maximal value of $\mathrm{p}^D_{tH}(\text{free}(D),\text{free}(F))$ where F is a contraction of D and $\mathrm{d}_{Hc}(F, D) \leq \delta$. Then ϵ can be computed to arbitrary precision.

Proof: By theorem 3.2, the desired maximal value is equal to $\mathrm{p}^D_{tH}(\text{free}(D),\text{free}(\text{contract}(D, \delta)))$. By lemma C.2.8, it suffices to consider the intersection of these free spaces with the basic configuration region of D, U. (The basic configuration region of $\text{contract}(D, \delta)$ is a subset.) By lemma C.2.12, the intersections $U \cap \text{free}(D)$ and $U \cap \text{free}(\text{contract}(D, \delta))$ are semi-algebraic. By lemma C.3.17, overlap graphs G and H for $U \cap \text{free}(D)$ and $U \cap \text{free}(\text{contract}(D, \delta))$ can be computed, as can the distance matrix $F[U,V]$. By lemma C.3.10, $B_{tH}(G, H)$ can be computed and by lemma C.3.16, it approximates $\mathrm{p}^D_{tH}(\text{free}(D),\text{free}(\text{contract}(D, \delta))$ to within ϵ. ▮

Proof of Theorem 3.4

Theorem 3.4: (Analogue of theorem 2.4.) Let D be a display such that free(D) is locally ordinarily connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that, if E is a expansion of D and the Hausdorff distance $\mathrm{d}_H(D, E) < \delta$ then p^D(paths(cfree(D)), paths(cfree(E))) $< \epsilon$.

Proof: The proof is essentially identical to that of theorem 3.1 using theorem 2.6 instead of theorem 2.1.

Proof of theorem 3.5

Lemma C.3.19:. Suppose that space U with metric μ is compact and locally internally connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ with the following property: Let ϕ be any path through U. Then there is a path ψ through contract(U, δ) such that $\mu_t(\phi, \psi) < \epsilon$.

Proof: Choose $\epsilon > 0$. For each point $x \in U$, choose an neighborhood O_x of x in U such that $O_x \subset \mathrm{B}(x, \epsilon/2)$ and $O_x \cap \mathrm{Int}(U)$ is connected. Since $\{ O_x \mid x \in U\}$ is an open covering of U and U is compact, there must be a finite subcovering $O_1 \ldots O_k$ of U.

For each pair of regions O_i, O_j that overlap, choose a point $y_{ij} \in O_i \cap O_j \cap \mathrm{Int}(U)$. For each pair of such points y_{ij}, y_{ik} in the same region O_i, choose a path $\eta_{ijk}(t)$ from y_{ij} to y_{ik} through $O_i \cap \mathrm{Int}(U)$. Let δ be the minimum distance from any point on any of the η_{ijk} to the boundary of U. It is clear that $\delta > 0$.

Let $\phi(t)$ be any path through U. By the same argument as in lemma C.3.12, there is a sequence of times $t_0 = 0, t_1, t_2 \ldots t_k = 1$, and a sequence of regions $O_{p_1}, O_{p_2}, \ldots O_{p_k}$, such that $\phi([t_i, t_{i+1})) \subset O_{p_i}$. Construct the path $\phi(t)$ so that $\phi(t_i) = y_{p_{i-1},p_i}$ and for $t \in [t_i, t_{i+1}]$, $\phi(t)$ follows the path $\eta_{p_{i-1}p_ip_{i+1}}$. Since both $\phi(t)$ and $\psi(t)$ are in the same neighborhood O_{p_i} for $t \in [t_i, t_{i+1}]$, and the diameter of O_{p_i} is at most ϵ, it follows that the distance from $\phi(t)$ to $\psi(t)$ is at most ϵ. ∎.

Lemma C.3.20: Suppose that free(D) is locally internally connected. Let U be the basic configuration region of D. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that
p^D_{tH}(paths(free(D) $\cap U$), paths(free(expand(D, δ)) $\cap U$) $< \epsilon$.

Proof: Let $\epsilon > 0$. Using lemma C.3.19, choose a distance $\delta_1 > 0$ such that, for any path ϕ in paths(free(D) $\cap U$) there exists a path ψ in contract^D(free(D) $\cap U$, δ_1) for which $\mathrm{p}^D_t(\phi, \psi) < \epsilon$. Using theorem 2.9, choose a value $\delta > 0$ such that free(expand(D, δ)) $\supset$ contract^D(free(D),δ_1). (Note that if free(D) is locally internally connected, then it is equal to the closure of cfree(D).) ∎.

Lemma C.3.21: Suppose that free(D) is locally internally connected. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that p^D_{tH}(paths(free(D)), paths(free(expand(D, δ)))) $< \epsilon$.

Proof: As in the proof of theorem 3.1, we can mirror the behavior of an arbitrary path through free(D) by a path through free(D) that remains in the basic configuration region.

Theorem 3.5: For any display D, if the configuration space free(D) is locally internally connected, then the function "paths(free($\cdot$))" is continuous at display D, using the metric d_{Hd} on displays and the topology p_{tH} on sets of paths,

Proof: By lemma C.3.21, the function "paths(free($\cdot$))" is continuous under expansion at D, and by theorem 3.1, it is continuous under contraction. Therefore, by an argument analogous to lemma C.2.20, it is continuous at D without qualification. ∎

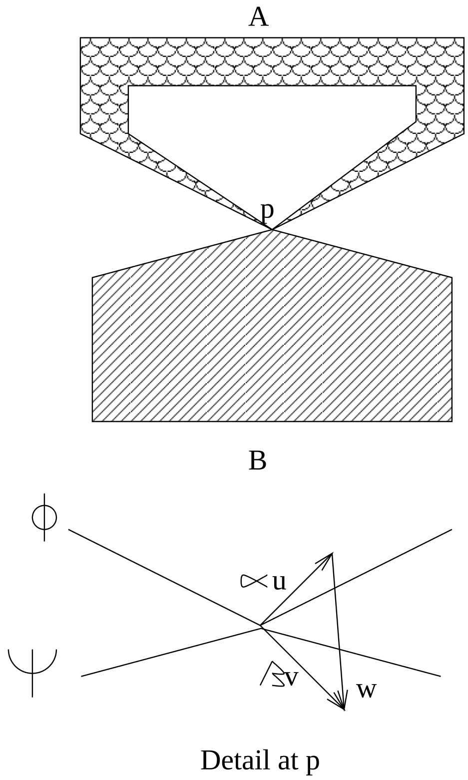


Figure 20: Lemma C.4.1

C.4: Proofs from section 4

Proof of theorem 4.2

Definition C.4.1: A direction $\hat{w}$ is a *positive sum* of directions $\hat{u}$ and $\hat{v}$ if $\hat{w} = \alpha\hat{u} + \beta\hat{v}$ for some $\alpha, \beta > 0$. Direction $\hat{w}$ is a *non-negative sum* of directions $\hat{u}$ and $\hat{v}$ if $\hat{w} = \alpha\hat{u} + \beta\hat{v}$ for some $\alpha, \beta \geq 0$. If $\hat{U}$ and $\hat{V}$ are two sets of directions, then the positive sum of $\hat{U}$ and $\hat{V}$ is the set of all positive sums of all directions $\hat{u} \in \hat{U}$ and $\hat{v} \in \hat{V}$.

Lemma C.4.1: Let $\mathbf{A}, \mathbf{B}$ be non-overlapping PSC regions, and let $\mathbf{p}$ be a point in $\mathbf{A} \cap \mathbf{B}$. Let ϕ=FBd($\mathbf{A}, \mathbf{B}$) and let ψ=FBd($\mathbf{B}, \mathbf{A}$). Let $\hat{u}$ be a direction strictly between forw($\phi, \mathbf{p}$) and back($\phi, \mathbf{p}$) in the counterclockwise direction, and let $\hat{v}$ be a direction strictly between forw($\psi, \mathbf{p}$) and back($\psi, \mathbf{p}$). Let $\hat{w}$ be a positive sum of $-\hat{u}$ and $\hat{v}$. Then $\hat{w}$ is a colliding direction from $\mathbf{A}$ to $\mathbf{B}$ at $\mathbf{p}$.

Proof: Let $\hat{w} = \beta\hat{v} - \alpha\hat{u}$ for some $\beta > 0$, $\alpha > 0$. Since $\hat{u}$ is strictly between forw($\phi, \mathbf{p}$) and back($\phi, \mathbf{p}$), the forward and backward tangents to ϕ, and since ϕ is the boundary of the connected component of $\mathbf{A}^c$ that contains $\mathbf{B}$, it follows that, for sufficiently small $\epsilon_u > 0$, the point $\mathbf{p} + \epsilon_\mathbf{u}\hat{\mathbf{u}}$ is in the interior of the far side of ϕ from $\mathbf{B}$ (Figure 20). Likewise, for sufficiently small $\epsilon_v > 0$, the point $\mathbf{p} + \epsilon_\mathbf{v}\hat{\mathbf{v}}$ is in the interior of the far side of ψ from $\mathbf{A}$. Let $\epsilon_0 = \min(\epsilon_u/\alpha, \epsilon_v/\beta)$. Then it is possible to choose $\epsilon_1 > 0$ such that the open ball, B($\mathbf{p} + \alpha\epsilon_\mathbf{0}\hat{\mathbf{u}}, \epsilon_\mathbf{1}$) is in the interior of the far side of ϕ from $\mathbf{B}$, and the open ball, B($\mathbf{p} + \beta\epsilon_\mathbf{0}\hat{\mathbf{v}}, \epsilon_\mathbf{1}$) is in the interior of the far side of ψ from $\mathbf{A}$.

Let $\hat{x}$ be the perpendicular to $\hat{w}$ ($= \alpha\hat{v} + \beta\hat{u}$). Let t be a parameter in (0,1) and define the functions $\mathbf{q}(t) = \mathbf{p} + \alpha\epsilon_0\hat{u} + t\epsilon_1\hat{x}$ and $\mathbf{r}(t) = \mathbf{p} + \beta\epsilon_0\hat{v} + t\epsilon_1\hat{x}$. Consider the space of all lines $\mathbf{q}(t)$ to $\mathbf{r}(t)$ for $t \in (0,1)$. Every such line crosses ϕ and crosses ψ, and they sweep out a solid parallelogram. Since $\mathbf{A}$ and $\mathbf{B}$ are normal, there must be some value t_0 such that the line from $\mathbf{q}(t_0)$ to $\mathbf{r}(t_0)$ intersects both Int($\mathbf{A}$) and Int($\mathbf{B}$). Let $\mathbf{a}$ and $\mathbf{b}$ be points in the intersection of this line with Int($\mathbf{A}$) and Int($\mathbf{B}$) respectively. Then $\text{dir}(\mathbf{b} - \mathbf{a}) = \text{dir}(\mathbf{r}(t_0) - \mathbf{q}(t_0)) = \text{dir}(\epsilon_0(\beta\hat{v} - \alpha\hat{u})) = \hat{w}$.

Since ϵ_0 and ϵ_1 can be chosen arbitrarily small, we have shown that in every neighborhood of $\mathbf{p}$ there exist points $\mathbf{a} \in$Int($\mathbf{A}$) and $\mathbf{b} \in$Int($\mathbf{B}$) such that $\text{dir}(\mathbf{b} - \mathbf{a}) = \hat{w}$. ▮

Lemma C.4.2: Let $\mathbf{R}$ be a PSC region; let ϕ be a directed boundary curve of $\mathbf{R}$; and let $\mathbf{p}$ be a point on ϕ. Let $\hat{c}$ =forw($\phi, \mathbf{p}$) and $\hat{d}$ =back($\phi, \mathbf{p}$). Then for any $\epsilon > 0$ there exists $\delta > 0$ satisfying the following: Let $\hat{x}$ be a direction counterclockwise between $\hat{d}$ and $\hat{c}$ such that $\text{d}(\hat{x}, \hat{c}) > \epsilon$ and $\text{d}(\hat{x}, \hat{d}) > \epsilon$. Then for any $t \in (0, \delta)$ the point $\mathbf{p} + t\hat{x}$ is outside $\mathbf{R}$.

Proof: Immediate from basic properties of the boundary curve and the tangent.

Lemma C.4.3: Let $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ be directions in non-strict counterclockwise order. If the angle between $\hat{e}$ and $\hat{f}$ is less than or equal to π, and if $\hat{w}$ is strictly between $\hat{e}$ and $\hat{f}$, then $\hat{w}$ is outside the the non-negative sum of $[\hat{g}, \hat{h}]$ and $[-\hat{e}, -\hat{f}]$.

Proof: The non-negative sums of directions $\hat{u}$ and $\hat{v}$ are all the directions on the shorter arc connecting $\hat{u}$ and $\hat{v}$, unless $\hat{u} = \pm\hat{v}$. Let $\hat{u}$ be in $[-\hat{e}, -\hat{f}]$ and $\hat{v}$ be in $[\hat{g}, \hat{h}]$. If an arc goes from $\hat{u}$ to $\hat{v}$, then it must either contain all of $(\hat{e}, \hat{f})$ or none of $(\hat{e}, \hat{f})$, as neither $\hat{u}$ nor $\hat{v}$ is in $(\hat{e}, \hat{f})$. Since $-\hat{u}$ is in $[\hat{e}, \hat{f}]$, it follows that the arc from $\hat{u}$ to $\hat{v}$ containing all of $(\hat{e}, \hat{f})$ must contain $-\hat{u}$ and therefore must have an angle of at least π. Hence, the shorter arc connecting $\hat{u}$ and $\hat{v}$ is disjoint from $(\hat{e}, \hat{f})$. ▮.

Corollary C.4.4: Let $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ be directions in non-strict counterclockwise order, such that $\hat{e} \neq \hat{f}$ and $\hat{g} \neq \hat{h}$. Then there is a direction $\hat{w}$ that is outside the non-negative sum of $[\hat{g}, \hat{h}]$ and $[-\hat{e}, -\hat{f}]$.

Proof: Either the angle from $\hat{e}$ to $\hat{f}$ or the angle from $\hat{g}$, $\hat{h}$ is less than or equal to π. In the former case, choose $\hat{w}$ to be strictly between $\hat{e}$ and $\hat{f}$ and apply lemma C.4.3. In the latter case, choose $\hat{w}$ to be strictly between $-\hat{g}$ and $-\hat{h}$. Then $-\hat{w}$ is strictly between $\hat{g}$ and $\hat{h}$. Applying lemma C.4.4, with the names reversed, $-\hat{w}$ is outside the non-negative sum of $[\hat{e}, \hat{f}]$ and $[-\hat{g}, -\hat{h}]$. Therefore $\hat{w}$ is outside the the non-negative sum of $[\hat{g}, \hat{h}]$ and $[-\hat{e}, -\hat{f}]$. ▮

Lemma C.4.5: Let $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ be directions in non-strict counterclockwise order, such that $\hat{e} \neq \hat{f}$ and $\hat{g} \neq \hat{h}$. Then the positive sum of $(\hat{g}, \hat{h})$ and $(-\hat{e}, -\hat{f})$ contains the interior of the non-negative sum of $[\hat{g}, \hat{h}]$ and $[-\hat{e}, -\hat{f}]$.

Proof: Consider the function $\text{f}(\hat{u}, \hat{v}, \alpha, \beta) = \text{dir}(\beta\hat{v} - \alpha\hat{u})$. Let D be the domain $\hat{u} \in [\hat{e}, \hat{f}]$, $\hat{v} \in [\hat{g}, \hat{h}]$, $\{ \alpha \geq 0, \beta > 0 \} \cup \{ \alpha > 0, \beta \geq 0 \}$. and let O be the open domain $\hat{u} \in (\hat{e}, \hat{f})$, $\hat{v} \in (\hat{g}, \hat{h})$, $\alpha, \beta > 0$. Then $f(D)$ is the non-negative sum of $[\hat{g}, \hat{h}]$ and $[-\hat{e}, -\hat{f}]$ and $f(O)$ the positive sum of $(\hat{g}, \hat{h})$ and $(-\hat{e}, -\hat{f})$. Since f is continuous, and D is contained in the closure of O, $f(D)$ is contained in the closure of $f(O)$. Since D and O are each connected, $f(D)$ and $f(O)$ are each connected. Finally, by corollary C.4.4, D is not the entire unit circle. Hence it is easily shown that D is an arc within the unit circle, and O must contain the entire interior of D. ▮

Lemma C.4.6: Let $\mathbf{A}, \mathbf{B}$ be non-overlapping PSC regions, and let $\mathbf{p}$ be a point in $\mathbf{A} \cap \mathbf{B}$. Let $\hat{w}$ be a colliding direction from $\mathbf{A}$ into $\mathbf{B}$ at $\mathbf{p}$. Let ϕ=FBd($\mathbf{A}, \mathbf{B}$) and let ψ=FBd($\mathbf{B}, \mathbf{A}$). Let $\hat{e}$= forw($\phi, \mathbf{p}$), $\hat{f}$ = back($\phi, \mathbf{p}$) $\hat{g}$= forw($\psi, \mathbf{p}$), $\hat{h}$ = back($\psi, \mathbf{p}$). Then there exist vectors $\hat{u} \in (\hat{e}, \hat{f})$ and $\hat{v} \in (\hat{g}, \hat{h})$ such that $\hat{w}$ is a positive sum of $\hat{v}$ and $-\hat{u}$.

Proof: Since $\hat{w}$ is a colliding direction, for every integer k we may choose points $\mathbf{a}_k$, $\mathbf{b}_k$ such that $\text{d}(\mathbf{a}_k, \mathbf{p}) < 1/k$, $\text{d}(\mathbf{b}_k, \mathbf{p}) < 1/k$ and such that $\text{dir}(\mathbf{b}_k - \mathbf{a}_k) = \hat{w}$. Let $\hat{u}_k$ =dir($\mathbf{a}_k - \mathbf{p}$), $\hat{v}_k$ =dir($\mathbf{b}_k - \mathbf{p}$). Let $\hat{u}$ be a cluster point of the $\hat{u}_k$. Without loss of generality, we may ignore those $\hat{u}_k$ that do not converge on $\hat{u}$, and thus assume that $\hat{u}$ is the limit of the $\hat{u}_k$. Similarly, let $\hat{v}$ be the limits of $\hat{v}_k$. By the contrapositive to lemma C.4.2, $\hat{u} \in [\hat{e}, \hat{f}]$ and $\hat{v} \in [\hat{g}, \hat{h}]$.

Note that, for any k, $\mathrm{d}(\mathbf{a}_k, \mathbf{b}_k)\ \hat{w} = \mathrm{d}(\mathbf{a}_k, \mathbf{b}_k) \cdot \mathrm{dir}(\mathbf{b}_k - \mathbf{a}_k) = \mathbf{b}_k - \mathbf{a}_k =$ $(\mathbf{b}_k - \mathbf{p}) - (\mathbf{a}_k - \mathbf{p}) = \mathrm{d}(\mathbf{p}, \mathbf{b}_k) \cdot \mathrm{dir}(\mathbf{b}_k - \mathbf{p}) - \mathrm{d}(\mathbf{p}, \mathbf{a}_k) \cdot \mathrm{dir}(\mathbf{a}_k - \mathbf{p}) = \mathrm{d}(\mathbf{p}, \mathbf{b}_k) \cdot \hat{v} - \mathrm{d}(\mathbf{p}, \mathbf{a}_k) \cdot \hat{u}$. Thus, if we let $\alpha_k = \mathrm{d}(\mathbf{p}, \mathbf{a}_k)/\mathrm{d}(\mathbf{a}_k, \mathbf{b}_k)$ and $\beta_k = \mathrm{d}(\mathbf{p}, \mathbf{b}_k)/\mathrm{d}(\mathbf{a}_k, \mathbf{b}_k)$, we get $\hat{w} = \beta_k \hat{v}_k - \alpha_k \hat{u}_k$. Therefore, if $\alpha \geq 0$ and $\beta \geq 0$ are the limits of α_k and β_k then we have $\hat{w} = \beta \hat{v} - \alpha \hat{u}$.

Note that α and β may be infinite, but only if $\hat{v} = \hat{u}$, which is only possible if (i) $\hat{v} = \hat{u} = \hat{e} = \hat{h}$, or (ii) $\hat{v} = \hat{u} = \hat{g} = \hat{f}$. In case (i), for sufficiently small $\epsilon > 0$, it must be the case that $\hat{e} - \epsilon\hat{w}$ lies between $\hat{e}$ and $\hat{f}$ and that $\hat{h} + \epsilon\hat{w}$ lies between $\hat{h}$ and $\hat{g}$. (Note that this relies on the fact that the regions are cuspless.) We can then write $\hat{w} = (1/2\epsilon)[(\hat{h} + \epsilon\hat{w}) - (\hat{e} - \epsilon\hat{w}))]$ (since $\hat{e} = \hat{h}$). Thus, we can assume that α and β are in fact finite. Analogously, in case (ii), we can write $\hat{w} = (1/2\epsilon)(\hat{g} + \epsilon\hat{w}) - (\hat{f} - \epsilon\hat{w}))]$, and again assume that α and β are finite.

Thus, every colliding direction is in the non-negative sum of $[-\hat{e}, -\hat{f}]$ and $[\hat{g}, -\hat{h}]$. By definition, every colliding direction is in the interior of the set of colliding directions. Hence, by lemma C.4.5 every colliding direction is in the positive sum of $(-\hat{e}, -\hat{f})$ and $(\hat{g}, -\hat{h})$. ▮

Lemma C.4.7: For any regions $\mathbf{B}, \mathbf{A}$ and point $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$, $\mathrm{sep}(\mathbf{B}, \mathbf{A}, \mathbf{p})$ is an open set in the unit circle.

Proof: Immediate from the definition. ▮.

Lemma C.4.8: Let $\mathbf{A}, \mathbf{B}$ be PSC regions that meet but do not overlap, and let $\mathbf{p}$ be a point in $\mathrm{Bd}(\mathbf{A}) \cap \mathrm{Bd}(\mathbf{B})$. Let $\hat{e} = \mathrm{forw}(\mathrm{FBd}(\mathbf{A}, \mathbf{B}), \mathbf{p})$, $\hat{f} = \mathrm{back}(\mathrm{FBd}(\mathbf{A}, \mathbf{B}), \mathbf{p})$, $\hat{g} = \mathrm{forw}(\mathrm{FBd}(\mathbf{B}, \mathbf{A}), \mathbf{p})$, $\hat{h} = \mathrm{back}(\mathrm{FBd}(\mathbf{B}, \mathbf{A}), \mathbf{p})$, Then $\mathrm{sep}(\mathbf{B}, \mathbf{A}, \mathbf{p}) = \mathrm{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h})$.

Proof: From lemmas C.4.6 and C.4.7, it follows that $\mathrm{sep}(\mathbf{B}, \mathbf{A}, \mathbf{p})$ is the interior of the set difference of the positive sums of $(\hat{g}, \hat{h})$ and $(-\hat{e}, -\hat{f})$ minus the positive sums of $(\hat{e}, \hat{f})$ and $(-\hat{g}, -\hat{h})$. The form of sep1 as a function of $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ is then a straightforward calculation in each of the separate cases. ▮.

Lemma C.4.9:

A. $\mathrm{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) = -\mathrm{sep1}(\hat{g}, \hat{h}, \hat{e}, \hat{f})$

B. If $(\hat{g}, \hat{h}) \subset [-\hat{e}, \hat{e}]$ then $\mathrm{sep1}(\hat{e}, -\hat{e}, \hat{g}, \hat{h}) = (-\hat{e}, \hat{e})$.

C. If $(\hat{e}, \hat{f}) \subset [\hat{q}, -\hat{q}]$ then $\mathrm{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) \subset \mathrm{sep1}(\hat{q}, -\hat{q}, \hat{r}, \hat{s})$, where $\hat{r}$ and $\hat{s}$ are any two vectors in $(-\hat{q}, \hat{q})$.

D. If $(\hat{e}, \hat{f}) \supset [\hat{q}, -\hat{q}]$ then $\mathrm{sep1}(\hat{e}, \hat{f}, \hat{g}, \hat{h}) \subset \mathrm{sep1}(\hat{q}, -\hat{q}, \hat{r}, \hat{s})$, where $\hat{r}$ and $\hat{s}$ are any two vectors in $(-\hat{q}, \hat{q})$.

Proof: Immediate from definition 4.12. ▮

Lemma C.4.10: Let $\theta > 0$, and let $\hat{e}$, $\hat{f}$, $\hat{g}$, $\hat{h}$ and $\hat{e}'$, $\hat{f}'$, $\hat{g}'$, $\hat{h}'$, be two quadruples of vectors, both in positive order, such that $\mathrm{d}(\hat{e}', \hat{e}) < \theta$, $\mathrm{d}(\hat{f}', \hat{f}) < \theta$, $\mathrm{d}(\hat{g}', \hat{g}) < \theta$, and $\mathrm{d}(\hat{h}', \hat{h}) < \theta$. Let $\hat{u}$ be a vector such that $\mathrm{B}(\hat{u}, \theta) \subset \mathrm{sep1}(\hat{e},\ \hat{f},\ \hat{g},\ \hat{h})$. ($\mathrm{B}(\hat{u}, \theta)$ is the set of all directions $\hat{v}$ such that $\mathrm{d}(\hat{v}, \hat{u}) < \theta$.) Then $\hat{u} \in \mathrm{sep1}(\hat{e}',\ \hat{f}',\ \hat{g}',\ \hat{h}')$.

Proof: Immediate from lemma C.4.9. ▮

Definition C.4.2: Let $\langle \hat{e}, \hat{f} \rangle$ and $\langle \hat{g}, \hat{h} \rangle$ be two pairs of directions and let $\epsilon > 0$. We say that $\langle \hat{g}, \hat{h} \rangle$ *resembles* $\langle \hat{e}, \hat{f} \rangle$ *within* ϵ if one of the following holds.

i. $\mathrm{d}(\hat{g}, \hat{e}) < \epsilon$ and $\mathrm{d}(\hat{h}, \hat{f}) < \epsilon$; or

ii. $d(\hat{g}, \hat{e}) < \epsilon$ and $d(\hat{h}, -\hat{e}) < \epsilon$; or

iii. $d(\hat{g}, -\hat{f}) < \epsilon$ and $d(\hat{h}, \hat{f}) < \epsilon$.

Note that this is not a symmetric relation between $\langle \hat{e}, \hat{f} \rangle$ and $\langle \hat{g}, \hat{h} \rangle$.

Lemma C.4.11: Let ϕ be a piecewise smooth cycle, and let $\mathbf{p}$ be a point in ϕ. Then for any $\epsilon > 0$ there exists $\delta > 0$ satisfying the following: Let $\mathbf{q}$ be any point on ϕ such that that $d(\mathbf{p}, \mathbf{q}) < \delta$. Then the pair $\langle$ forw($\phi, \mathbf{q}$), back($\phi, \mathbf{q}$) $\rangle$ resembles $\langle$ forw($\phi, \mathbf{p}$), back($\phi, \mathbf{p}$) $\rangle$ within ϵ.

Proof: It is immediate from basic properties of piecewise smooth curves that, if $\mathbf{q}$ lies ahead of $\mathbf{p}$ on ϕ, then condition (ii) of definition C.4.2 holds, and if $\mathbf{q}$ lies behind $\mathbf{p}$ on ϕ, then condition (iii) of definition C.4.2 holds. If $\mathbf{q} = \mathbf{p}$ then, of course, condition (i) holds.

Definition C.4.3: We say that a quintuple of directions $\langle \hat{e}, \hat{f}, \hat{g}, \hat{h}, \hat{u} \rangle$ is *blocked* if $\langle \hat{e}, \hat{f}, \hat{g}, \hat{h} \rangle$ is in non-strict counterclockwise order, and $\hat{u} \not\in$sep1$(\hat{e}, \hat{f}, \hat{g}, \hat{h})$. Let "blocked" be the set of all blocked quintuples. Measure the distance between two quintuples $\langle \hat{e}, \hat{f}, \hat{g}, \hat{h}, \hat{u} \rangle$ and $\langle \hat{e}', \hat{f}', \hat{g}', \hat{h}', \hat{u}' \rangle$ as $\max(d(\hat{e}, \hat{e}'), d(\hat{f}, \hat{f}'), d(\hat{g}, \hat{g}'), d(\hat{h}, \hat{h}'), d(\hat{u}, \hat{u}'))$ We define the *angle from blockage* of a quintuple $\langle \hat{e}, \hat{f}, \hat{g}, \hat{h}, \hat{u} \rangle$ as the distance from that quintuple to the set "blocked". That is, the angle from blockage is

$$\min_{\langle \hat{e}', \hat{f}', \hat{g}', \hat{h}', \hat{u}' \rangle \in \text{blocked}} \max(d(\hat{e}, \hat{e}'), d(\hat{f}, \hat{f}'), d(\hat{g}, \hat{g}'), d(\hat{h}, \hat{h}'), d(\hat{u}, \hat{u}'))$$

It is easily shown that the set "blocked" is closed; hence this minimum is attained.

That is, the angle from blockage is the minimal change in all these vectors so that the tangent vectors can correspond to those of two non-overlapping objects in contact, and so that $\hat{u}$ does not strongly separate the two objects.

Definition C.4.4: Let $\mathbf{A}$ and $\mathbf{B}$ be PSC regions that meet but do not overlap. Let $\phi = \text{FBd}(\mathbf{A}, \mathbf{B})$, and let $\psi = \text{FBd}(\mathbf{B}, \mathbf{A})$, Let $\mathbf{p}$ be a point in ϕ, and let $\mathbf{q}$ be a point in ψ. Let $\hat{u}$ be any direction. Then the function "free_ang($\mathbf{A}, \mathbf{p}, \mathbf{B}, \mathbf{q}, \hat{u}$)" is defined to be the angle from blockage of $\langle$ forw($\phi, \mathbf{p}$), back($\phi, \mathbf{p}$), forw($\psi, \mathbf{q}$), back($\psi, \mathbf{q}$), $\hat{u}$ $\rangle$.

Lemma C.4.12: Let $\mathbf{A}$ and $\mathbf{B}$ be PSC regions that are strongly separable at point $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$. Let $\hat{u}$ be a direction in sep($\mathbf{A}, \mathbf{B}, \mathbf{p}$). Then there exist $\theta > 0$ and $\delta > 0$ satisfying the following: For any points $\mathbf{q} \in$FBd($\mathbf{A}, \mathbf{B}$) and $\mathbf{r} \in$FBd($\mathbf{B}, \mathbf{A}$), if $d(\mathbf{q}, \mathbf{p}) < \delta$ and $d(\mathbf{r}, \mathbf{p}) < \delta$ then free_ang($\mathbf{A}, \mathbf{q}, \mathbf{B}, \mathbf{r}, \hat{u}$) $> \theta$.

Proof: Let ϕ_1 be a smooth curve along FBd($\mathbf{A}, \mathbf{B}$) with the same orientation such that $\phi_1(0) = \mathbf{p}$. Let ϕ_2 be a smooth curve along FBd($\mathbf{A}, \mathbf{B}$) with the reverse orientation such that $\phi_2(0) = \mathbf{p}$. Let ψ_1 be a smooth curve along FBd($\mathbf{B}, \mathbf{A}$) with the same orientation such that $\phi_1(0) = \mathbf{p}$. Let ψ_2 be a smooth curve along FBd($\mathbf{B}, \mathbf{A}$) with the reverse orientation such that $\phi_2(0) = \mathbf{p}$.

Define the vector functions

$\hat{\alpha}_{1+}(t) = \text{dir}(\dot{\phi}_1(t))$, $\hat{\alpha}_{1-}(t) = -\text{dir}(\dot{\phi}_1(t))$, $\hat{\alpha}_{2+}(t) = -\text{dir}(\dot{\phi}_2(t))$, $\hat{\alpha}_{2-}(t) = \text{dir}(\dot{\phi}_2(t))$,
$\hat{\beta}_{1+}(t) = \text{dir}(\dot{\psi}_1(t))$, $\hat{\beta}_{1-}(t) = -\text{dir}(\dot{\psi}_1(t))$, $\hat{\beta}_{2+}(t) = -\text{dir}(\dot{\psi}_2(t))$, $\hat{\beta}_{2-}(t) = \text{dir}(\dot{\psi}_2(t))$.
and define the constant functions
$\hat{\alpha}_{0+}(t) = \text{dir}(\dot{\phi}_1(0))$, $\hat{\alpha}_{0-}(t) = \text{dir}(\dot{\phi}_2(0))$, $\hat{\beta}_{0+}(t) = \text{dir}(\dot{\psi}_1(0))$, $\hat{\beta}_{0-}(t) = \text{dir}(\dot{\psi}_2(0))$,

That is, as t goes from 0 to 1, $\hat{\alpha}_{1+}(t)$ and $\hat{\alpha}_{1-}(t)$ are the forward and backward tangents along the boundary of $\mathbf{A}$ facing $\mathbf{B}$ ahead of $\mathbf{p}$; $\hat{\alpha}_{2+}(t)$ and $\hat{\alpha}_{2-}(t)$ are the backward and forward tangents along the boundary of $\mathbf{A}$ facing $\mathbf{B}$ behind $\mathbf{p}$; $\hat{\beta}_{1+}(t)$ and $\hat{\beta}_{1-}(t)$ are the forward and backward tangents along the boundary of $\mathbf{B}$ facing $\mathbf{A}$ ahead of $\mathbf{p}$; and $\hat{\beta}_{2+}(t)$ and $\hat{\beta}_{2-}(t)$ are the backward and forward tangents along the boundary of $\mathbf{B}$ facing $\mathbf{A}$ behind $\mathbf{p}$.

Finally define the function $f_{ij}(s,t)$ to be the angle from blockage of $\langle \hat{\alpha}_{i+}(s), \hat{\alpha}_{i-}(s), \hat{\beta}_{j+}(t), \hat{\beta}_{j-}(t), \hat{u} \rangle$.

It is clear that the α and β functions are continuous, and therefore the functions f_{ij} are also continuous. Next, we will show that $f_{i,j}(0,0) > 0$ for each i,j. Let $\hat{e}$ = forw(FBd(A,B),$\mathbf{p}$) = $\alpha_{1+}(0) = \alpha_{0+}(0)$; let $\hat{f}$ = back(FBd(A,B),$\mathbf{p}$) = $\alpha_{2-}(0) = \alpha_{0-}(0)$; let $\hat{g}$ = forw(FBd(B,A),$\mathbf{p}$) = $\beta_{1+}(0) = \beta_{0+}(0)$; and let $\hat{h}$ = back(FBd(B,A),$\mathbf{p}$) = $\beta_{2-}(0) = \beta_{0-}(0)$. We are given that $\hat{u} \in$sep($\mathbf{A}, \mathbf{B}, \mathbf{p}$) = sep1($\hat{e}, \hat{f}, \hat{g}, \hat{h}$)

There are three essentially different cases:

a. $f_{0,0}(s,t)$ has the constant value free_ang($\mathbf{A}, \mathbf{p}, \mathbf{B}, \mathbf{p}, \hat{\mathbf{u}}$) > 0.

b. $f_{1,0}(0,0)$ is equal to the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, \hat{h}, \hat{u}$. If $-\hat{e}$ is not strictly between $\hat{g}$ and $\hat{h}$, then, by lemma C.4.9, sep1($\hat{e}, -\hat{e}, \hat{g}, \hat{h}$) $\supset$ sep1($\hat{e}, \hat{f}, \hat{g}, \hat{h}$). Therefore $\hat{u} \in$sep1($\hat{e}, -\hat{e}, \hat{g}, \hat{h}$), so the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, \hat{h}$ is greater than 0.

   If $-\hat{e}$ is strictly between $\hat{g}$ and $\hat{h}$, then the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, \hat{h}$ is at least (1/2)min(d($-\hat{e}, \hat{g}$), d($-\hat{e}, \hat{h}$)), and so is greater than 0.

c. $f_{1,1}(0,0)$ is equal to the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, -\hat{g}$. If $\hat{e} = -\hat{g}$ then, by lemma C.4.9, sep1($\hat{e}, -\hat{e}, \hat{g}, -\hat{g}$) $\supset$ sep1($\hat{e}, \hat{f}, \hat{g}, \hat{h}$). Therefore $\hat{u} \in$sep1($\hat{e}, -\hat{e}, \hat{g}, -\hat{g}$), so the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, -\hat{g}$ is greater than 0.

   If $\hat{e} \neq -\hat{g}$ then the angle from blockage of $\hat{e}, -\hat{e}, \hat{g}, -\hat{g}$ is at least (1/2)min(d($\hat{e}, \hat{g}$), d($\hat{e}, -\hat{g}$)), and so is greater than 0. (We know that $\hat{e} \neq \hat{g}$ from the facts that $\hat{e}, \hat{f}, \hat{g}, \hat{h}$ are in counterclockwise order, and that $\hat{e} \neq \hat{f}$ and $\hat{g} \neq \hat{h}$.)

Since the configuration is symmetric between $i = 1$ and $i = 2$, between $j = 1$ and $j = 2$, and between α and β, it follows by analogous arguments that $f_{i,j}(0,0)$ is positive for each i,j.

Since $f_{i,j}(0,0)$ is positive and $f_{i,j}$ is continuous, it follows that there exist $\theta_{ij} > 0$ and $\tau_{ij} > 0$, such that, for any s,t, if $0 \leq s < \tau_{ij}$, $0 \leq t < \tau_{ij}$, then $f_{ij}(s,t) > \theta_{ij}$. Choose $\delta_{ij} > 0$ such that, for any s,t, if $i \neq 0$ and d($\phi_i(s), \mathbf{p}$) $< \delta_{ij}$ then $s < \tau_{ij}$ and if $j \neq 0$ and d($\psi_j(t), \mathbf{p}$) $< \delta_{ij}$ then $t < \tau_{ij}$. (Since ϕ_i and ψ_j are one-to-one, their inverses are continuous.) If we choose δ to be the minimum of the δ_{ij} and θ to be the minimum of the θ_{ij}, the conclusion of the lemma is satisfied. ▮

Definition C.4.6: The distance from point $\mathbf{p}$ to motion M, d($\mathbf{p}, M$), is equal to d($\mathbf{p}, \mathbf{o}$) if $M = \langle \mathbf{o}, S \rangle$ and equal to 1 if $M = \hat{v}$. If $\mathbf{R}$ is a region then d($\mathbf{R}, M$) = $\inf_{\mathbf{p} \in \mathbf{R}}$ d($\mathbf{p}, M$).

Lemma C.4.13: For any distinct points $\mathbf{p}, \mathbf{q}$ and motion M, if d($\mathbf{p}, \mathbf{q}$) $<$ (3/5)d($\mathbf{p}, M$) then d(flow($\mathbf{p}, M$), flow($\mathbf{q}, M$)) $<$ (5/4) d($\mathbf{p}, \mathbf{q}$) / d($\mathbf{p}, M$).

Proof: If M is a translation, then flow($\mathbf{p}, M$) = flow($\mathbf{q}, M$), so the statement is trivial. Let $M = \langle \mathbf{o}, S \rangle$. Since flow($\mathbf{p}, M$) is at right angles to $\mathbf{op}$ and flow($\mathbf{q}, M$) is at right angles to $\mathbf{oq}$, we have d(flow($\mathbf{p}, M$), flow($\mathbf{q}, M$)) = d(dir($\mathbf{o}, \mathbf{p}$),dir($\mathbf{o}, \mathbf{q}$)). Note that this is an increasing function of the angle $\angle\mathbf{poq}$. For fixed values of d($\mathbf{p}, \mathbf{q}$) and d($\mathbf{p}, \mathbf{o}$), it is easily shown that the angle $\angle\mathbf{poq}$ is maximal when $\mathbf{pq}$ is perpendicular to $\mathbf{oq}$. Referring to figure 21, it can be seen that d(dir($\mathbf{o}, \mathbf{p}$),dir($\mathbf{o}, \mathbf{q}$)) $<$ y = d($\mathbf{p}, \mathbf{q}$)/d($\mathbf{q}, \mathbf{o}$). Let λ = d($\mathbf{p}, \mathbf{q}$)/d($\mathbf{p}, \mathbf{o}$). Then d($\mathbf{p}, \mathbf{q}$)/d($\mathbf{q}, \mathbf{o}$) = $\lambda/\sqrt{1-\lambda^2} < (5/4)\lambda$ if $\lambda < 3/5$. (The factor "3/5" was chosen just so that the expression in the square root should come out to be rational, for convenience.)

Lemma C.4.14: Let $\mathbf{A}$ and $\mathbf{B}$ be non-overlapping PSC regions, and let M be a motion strongly separating $\mathbf{B}$ from $\mathbf{A}$. For any point $\mathbf{p} \in \mathbf{B}$, there exist $\epsilon > 0, \alpha > 0$ such that, for any $\mathbf{a} \in$ Int($\mathbf{A}$), $\mathbf{b} \in$ Int($\mathbf{B}$) if both $\mathbf{a}$ and $\mathbf{b}$ are within ϵ of $\mathbf{p}$, then d(flow($\mathbf{p}, M$), dir($\mathbf{b}, \mathbf{a}$)) $> \alpha$.

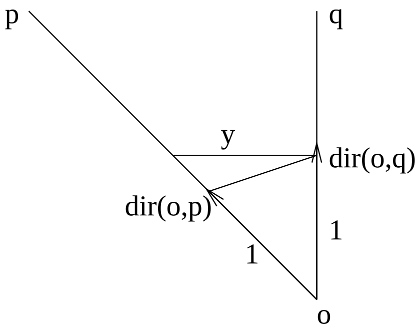


Figure 21: Lemma C.4.13

Proof: If $\mathbf{p}$ is not in $\mathbf{A}$, then let $\epsilon <$ d($\mathbf{p}, \mathbf{A}$). The condition is then satisfied vacuously, as there is no such $\mathbf{a}$. If $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$, then the condition is immediate from definitions 4.9, 4.10, and 4.15.

Lemma C.4.15: Let $\mathbf{A}$ and $\mathbf{B}$ be non-overlapping PSC regions, and let M be a motion strongly separating $\mathbf{B}$ from $\mathbf{A}$. There exist $\epsilon > 0, \alpha > 0$ such that, for any points $\mathbf{p}, \mathbf{b} \in \mathbf{B}$ and $\mathbf{a} \in \mathbf{A}$, if both $\mathbf{a}$ and $\mathbf{b}$ are within ϵ of $\mathbf{p}$, then d(flow($\mathbf{p}, M$), dir($\mathbf{b}, \mathbf{a}$)) $> \alpha$. (This is the same as lemma C.4.14, except here α and ϵ are quantified with larger scope than $\mathbf{p}$.)

Proof: We apply lemma B.8, taking the domain U to be $\mathbf{B}$ and the property $\Phi(\mathbf{p}, \epsilon, \alpha)$ to be the relation, "For any point $\mathbf{b} \in \mathbf{B}$ and $\mathbf{a} \in \mathbf{A}$, if both $\mathbf{a}$ and $\mathbf{b}$ are within ϵ of $\mathbf{p}$, then the angle between flow($\mathbf{p}, M$) and dir($\mathbf{b}, \mathbf{a}$) is at least α." Property (a) of lemma B.8 is immediate. Property (b) requires establishing that, for any point $\mathbf{p} \in \mathbf{B}$ there exist $\delta > 0, \mu > 0, \theta > 0$, such that, if $\mathbf{u} \in \mathbf{B}$ and d($\mathbf{u}, \mathbf{p}$) $< \delta$ then $\Phi(\mathbf{u}, \mu, \theta)$. We do this as follows: For any point $\mathbf{p} \in \mathbf{B}$, use lemma C.4.14 to find $\epsilon > 0, \alpha > 0$ such that $\Phi(\mathbf{p}, \epsilon, \alpha)$. Let $\theta = \alpha/2$ and let $\delta = \mu < \min(\epsilon/2, (3/5)\mathrm{d}(\mathbf{p}, M), (2/5)\mathrm{d}(\mathbf{p}, M)\alpha)$. Now choose $\mathbf{u}$ within δ of $\mathbf{p}$ and $\mathbf{a} \in \mathbf{A}$, $\mathbf{b} \in \mathbf{B}$ such that both $\mathbf{a}$ and $\mathbf{b}$ are within μ of $\mathbf{u}$. Since $\mathbf{a}$ and $\mathbf{b}$ are within ϵ of $\mathbf{p}$, dir($\mathbf{b}, \mathbf{a}$) be at least α from flow($\mathbf{p}, M$). By lemma C.4.13, since d($\mathbf{u}, \mathbf{p}$) $< (3/5)\mathrm{d}(\mathbf{p}, M)$, it follows that d(flow($\mathbf{u}, M$),flow($\mathbf{p}, M$) $< (5/4)(\mathrm{d}(\mathbf{u}, \mathbf{p})/\mathrm{d}(\mathbf{p}, M)) < (5/4)(2/5)$ $\mathrm{d}(\mathbf{p}, M)\alpha)/\mathrm{d}(\mathbf{p}, M) = \alpha/2$.

Thus since d(dir($\mathbf{b}, \mathbf{a}$), flow($\mathbf{p}, M$)) $> \alpha$, and d(flow($\mathbf{u}, M$),flow($\mathbf{p}, M$)) $< \alpha/2$, it follows that d(dir($\mathbf{b}, \mathbf{a}$), flow($\mathbf{u}, M$)) $> \alpha/2$. Therefore we can apply lemma B.8, and conclude that α and ϵ can be chosen uniformly over all $\mathbf{p} \in \mathbf{B}$. ▮

Lemma C.4.16: Let $\mathbf{A}$ and $\mathbf{B}$ be non-overlapping PSC regions, and let M be a motion strongly separating $\mathbf{B}$ from $\mathbf{A}$. There exist $\delta > 0$ and $\theta > 0$ such that, for all $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$, $\mathbf{q} \in$FBd($\mathbf{A}$), $\mathbf{r} \in$FBd($\mathbf{B}$), if $\hat{u}$ =flow($\mathbf{p}, M$), d($\mathbf{p}, \mathbf{q}$) $< \delta$, and d($\mathbf{p}, \mathbf{r}$) $< \delta$, then free_ang($\mathbf{A}, \mathbf{q}, \mathbf{B}, \mathbf{r}, \hat{u}$) $\geq \theta$.

Proof: We apply lemma B.8, taking the domain U to be $\mathbf{A} \cap \mathbf{B}$ and the property $\Phi(\mathbf{p}, \delta, \theta)$ to be the property, "For any $\mathbf{q} \in$FBd($\mathbf{A}$), $\mathbf{r} \in$FBd($\mathbf{B}$), if $\hat{u}$ =flow($\mathbf{p}, M$), d($\mathbf{p}, \mathbf{q}$) $< \delta$, and d($\mathbf{p}, \mathbf{r}$) $< \delta$, then free_ang($\mathbf{A}, \mathbf{q}, \mathbf{B}, \mathbf{r}, \hat{u}$) $\geq \theta$." Property (a) of lemma B.8 is immediate. Property (b) requires establishing that, for any point $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$ there exist a $\mu > 0, \beta > 0, \alpha > 0$, such that, for $\mathbf{x} \in \mathbf{A} \cap \mathbf{B}$, if d($\mathbf{x}, \mathbf{p}$) $< \mu$, then $\Phi(\mathbf{x}, \beta, \alpha)$. We proceed as follows: For any such $\mathbf{p}$, find δ and θ satisfying lemma C.4.12. Let $\alpha = \theta/2$, $\beta = \delta/2$, and $\mu = \min(\delta/2, (3/5)\mathrm{d}(\mathbf{p}, M), (4/5)\alpha \mathrm{d}(\mathbf{p}, M))$. Let $\mathbf{x}$ be any point in $\mathbf{A} \cap \mathbf{B}$ within μ of $\mathbf{p}$ and let $\mathbf{q}, \mathbf{r}$ be points in $\mathbf{A}$ and $\mathbf{B}$ within β of $\mathbf{x}$. Then certainly d($\mathbf{p}, \mathbf{r}$) $<$ δ, so free_ang($\mathbf{A}, \mathbf{q}, \mathbf{B}, \mathbf{r}$,flow($\mathbf{p}$,$M$)) $\geq \theta$. Moreover, by lemma C.4.13, d(flow($\mathbf{x}$,M), flow($\mathbf{p}$,M)) $<$ α. It then follows immediately, from definitions 4.10 and 4.11, that free_ang($\mathbf{A}, \mathbf{q}, \mathbf{B}, \mathbf{r}$,flow($\mathbf{x}$,$M$)) $\geq \theta - \alpha = \alpha$. ▮

Lemma C.4.17: Let $\mathbf{A}$ and $\mathbf{B}$ be compact regions. For any $\epsilon > 0$ there exists a $\gamma > 0$ such that if

$\mathbf{a} \in \mathbf{A}$, $\mathbf{b} \in \mathbf{B}$ and $\mathrm{d}(\mathbf{a}, \mathbf{b}) < \gamma$ then there exists a point $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$ such that $\mathrm{d}(\mathbf{p}, \mathbf{a}) < \epsilon$ and $\mathrm{d}(\mathbf{p}, \mathbf{b}) < \epsilon$.

Proof: Immediate from lemma B.6, with the compact domain $\mathbf{U}$ being the cross product $\mathbf{A} \times \mathbf{B}$, the function $f(\langle \mathbf{a}, \mathbf{b} \rangle) = \mathrm{d}(\mathbf{a}, \mathbf{b})$, and the metric $\mu(\langle \mathbf{a}, \mathbf{b} \rangle, \langle \mathbf{c}, \mathbf{d} \rangle) = \max(\mathrm{d}(\mathbf{a}, \mathbf{c}), \mathrm{d}(\mathbf{b}, \mathbf{d}))$.

Lemma C.4.18: If $\mathbf{B}$ approximates $\mathbf{A}$ in tangent (α, β) and $\mathbf{C}$ approximates $\mathbf{B}$ in tangent (μ, θ) then $\mathbf{C}$ approximates $\mathbf{A}$ in tangent $(\mu + \alpha, \theta + \beta)$.

Proof: If $\mathbf{a} \in \mathbf{A}$ is associated with $\mathbf{b} \in \mathbf{B}$ and $\mathbf{b} \in \mathbf{B}$ is associated with $\mathbf{c} \in \mathbf{C}$ then associate $\mathbf{a}$ with $\mathbf{c}$. ▮

Lemma C.4.19: Let $\mathbf{A}$ and $\mathbf{B}$ be non-overlapping PSC regions, and let M be a motion strongly separating $\mathbf{B}$ from $\mathbf{A}$. There exist $\beta > 0, \psi > 0$ such that, if $\mathbf{X}, \mathbf{Y}$ are non-overlapping PSC regions that respectively approximate $\mathbf{A}, \mathbf{B}$ in tangent (β, ψ), then M strongly separates $\mathbf{Y}$ from $\mathbf{X}$.

Proof: Find δ and θ to satisfy lemma C.4.16. Let $\psi < \theta/2$, and let
$D = (1/2)\min((3/5)\ \mathrm{d}(\mathbf{A} \cap \mathbf{B},\ M),\ (4/5)\ \psi\ \mathrm{d}(\mathbf{A} \cap \mathbf{B},\ M))$. Let $\epsilon = \min(\delta/2, D)$. Find a value of γ to satisfy lemma C.4.17 for this value of ϵ, and let $\beta = \min(\gamma/2, D)$.

Now let $\mathbf{X}, \mathbf{Y}$ be non-overlapping PSC regions that respectively approximate $\mathbf{A}, \mathbf{B}$ in tangent (β, ψ). Let $\mathbf{p} \in \mathbf{X} \cap \mathbf{Y}$. (If there is no such point, then the statement is trivial.) Let $\mathbf{a}$ be a point in $\mathrm{FBd}(\mathbf{A}, \mathbf{B})$ that corresponds to $\mathbf{p}$ and let $\mathbf{b}$ be a point in $\mathrm{FBd}(\mathbf{B}, \mathbf{A})$ that corresponds to $\mathbf{p}$. Then $\mathrm{d}(\mathbf{a}, \mathbf{b}) < 2\beta < \gamma$, so there exists a point $\mathbf{q} \in \mathbf{A} \cap \mathbf{B}$ such that $\mathrm{d}(\mathbf{q}, \mathbf{a}) < \epsilon$, $\mathrm{d}(\mathbf{q}, \mathbf{b}) < \epsilon$. Therefore, $\mathrm{free\_ang}(\mathbf{A}, \mathbf{a}, \mathbf{B}, \mathbf{b}, \mathrm{flow}(\mathbf{q}, M)) \geq \theta$. Now, the tangents to $\mathbf{A}$ at $\mathbf{a}$ are within ψ of the tangents to $\mathbf{X}$ at $\mathbf{p}$, and the tangents to $\mathbf{B}$ are within ψ of the tangents to $\mathbf{Y}$ at $\mathbf{p}$. Also, by lemma C.4.13, $\mathrm{flow}(\mathbf{p}, M)$ is within ψ of $\mathrm{flow}(\mathbf{q}, M)$. Therefore, $\mathrm{free\_ang}(\mathbf{X}, \mathbf{p}, \mathbf{Y}, \mathbf{p}, \mathrm{flow}(\mathbf{p}, M)) \geq \theta - 2\psi > 0$. Since $\mathbf{X}$ and $\mathbf{Y}$ do not overlap, the tangents to $\mathbf{X}$ and $\mathbf{Y}$ at $\mathbf{p}$ are necessarily in proper cyclic order, so $\mathrm{flow}(\mathbf{p}, M) \in \mathrm{sep}(\mathbf{Y}, \mathbf{X}, \mathbf{p})$. ▮

Definition C.4.7: Let C be a configuration over the two objects A and B; let M be a motion; and let $\Delta > 0$. Then "$\mathrm{arc}(C, M, \Delta)$" is the path ϕ that starts in C, leaves A fixed and moves B along M until the distance moved is equal to Δ. Formally,

i. $\phi(0) = C$.

ii. For all $T \in [0, 1]$, $\phi(T)[A] = C[A]$.

iii. For all $T \in (0, 1)$ and for all $\mathbf{p}$, there exists $k > 0$ such that
$d/dT(\phi(T)[B](\mathbf{p})) = k \cdot \mathrm{flow}(\phi(T)[B](\mathbf{p}), M)$,

iv. Let $\mathbf{B} = CD[B]$. Then $\mathrm{p}^D(\phi(0), \phi(1)) = \Delta$.

v. For $T \in (0, 1)$, $\mathrm{p}^D(\phi(0), \phi(T)) < \Delta$. (This, to prevent ϕ from traversing a full rotation before ending up at $\phi(1)$.)

Definition C.4.8: We say that path ϕ *escapes contact* over D if, for all $T \in (0, 1]$ $\langle D, \phi(T) \rangle$ is contact-free.

Lemma C.4.20: Let $\langle D, C \rangle$ be a feasible scenario over objects A and B, where CD[A] and CD[B] are PSC regions. Let M be a motion that strongly separates CD[B] from CD[A]. Then there exists a $\Delta > 0$ such that $\mathrm{arc}(C, M, \Delta)$ escapes contact.

Proof: Find ϵ, α to satisfy lemma C.4.14 for $\mathbf{A} = CD$[A], $\mathbf{B} = CD$[B]. If M is a translation, then choose $\Delta = \epsilon$. Otherwise, if $M = \langle \mathbf{o}, S \rangle$, let R be the maximal value of $\mathrm{d}(\mathbf{o}, \mathbf{p})$ for $\mathbf{p} \in \mathbf{B}$, and let $\Delta = \min(\epsilon, R\alpha)$. Let $\phi = \mathrm{arc}(C, M, \Delta)$. Then for any $\mathbf{p} \in \mathbf{B}$, let $\mathbf{b} = C^{-1}(\mathbf{p}) \in D[\mathrm{B}]$; and let $\mathbf{W}(T)$ be the curve $\phi(T)[B](\mathbf{b})$. We wish to show that $\mathbf{W}(T)$ does not intersect $\mathbf{A}$ for any $T \in (0, 1]$.

If M is a rotation with center $\mathbf{p}$, then $\mathbf{p} \notin \mathbf{A}$, and $\mathbf{W}(T) = \mathbf{p}$ for all T, so the result is immediate. If not, for any $T \in (0,1]$, $\mathrm{d}(\mathbf{W}(T), \mathbf{p}) \leq \Delta$. Let $\dot{\mathbf{W}}(T)$ be the tangent to $\mathbf{W}$ at time T; then, for all T, $\dot{\mathbf{W}}(T)$ =flow($\mathbf{W}(T), M$) which is within α of flow($\mathbf{p}, M$). Then, since $\mathbf{W}(T) - \mathbf{p} = \int_0^T \dot{\mathbf{W}}(s)ds$, it follows that, for any $T \in (0,1]$, d(dir($\mathbf{p}, \mathbf{W}(T)$), flow($\mathbf{p}, M$)) $< \alpha$. By lemma C.4.14 therefore, $\mathbf{W}(T)$ is not in $\mathbf{A}$. ▮

Lemma C.4.20 establishes that, if M strongly separates CD[B] from CD[A], as defined in definition 4.8, then there actually is a path along M that separates B from A.

Lemma C.4.21: Let D be a display and let C and C' be configurations over objects A and B. Let Δ be the minimum of the diameters of D[A] and D[B], and let $\lambda = \mathrm{p}^D(C, C')$. If $2\lambda < \Delta$, then $C'D$ approximates CD in tangent with parameters $(\lambda, 2\lambda/\Delta)$.

Proof: Let $\mathbf{p}$ and $\mathbf{q}$ be two points either both in D[A] or both in D[B] such that $\mathrm{d}(\mathbf{p}, \mathbf{q}) = \Delta$. If there is a rotation between C and C' that moves directional vectors by ψ, then one of $\mathbf{p}$ and $\mathbf{q}$ must be at least $\Delta/2$ from the center of rotation, and therefore must move a distance at least $\Delta\psi/2$. Since neither $\mathbf{p}$ nor $\mathbf{q}$ moves more than λ, ψ must be no more than $2\lambda/\Delta$. ▮

Corollary C.4.22: Let $\langle D, C\rangle$ be a scenario over objects A,B. For any $\epsilon > 0, \psi > 0$, there exist $\mu > 0, \theta > 0, \lambda > 0$ such that, for any scenario $\langle D', C'\rangle$, if $\mathrm{p}^D(C, C') < \lambda$ and D' approximates D in tangent (μ, θ), then $C'D'$ approximates CD in tangent (ϵ, ψ).

Proof: Let Δ be the minimum of the diameters of D[A], D[B]. Let $\theta = \psi/2$, let $\mu = \epsilon/2$, and let $\lambda = \min(\Delta/2, \epsilon/2, \psi\Delta/2)$. By lemma C.4.21, $C'D$ approximates CD in tangent $(\epsilon/2, \psi/2)$. Clearly $C'D'$ approximates $C'D$ in tangent $(\epsilon/2, \psi/2)$. Therefore, by lemma C.4.18, $C'D'$ approximates CD in tangent (ϵ, ψ). ▮.

Lemma C.4.23: Let $\langle D, C\rangle$ be a feasible scenario over objects A,B, and let M be a motion that strongly separates CD[B] from CD[A]. Then there exist $\epsilon > 0, \gamma > 0, \theta > 0$ satisfying the following: Let D_0 be a display that approximates D in tangent (γ, θ) and let C_1 and C_2 be feasible configurations over D_0 such that $C_1[\mathrm{A}] = C_2[\mathrm{A}]$=$C$[A], $\mathrm{p}^D(C, C_1) < \epsilon$ and $\mathrm{p}^D(C, C_2) < \epsilon$. Then there is a path connecting C_1 and C_2 through cfree(D_0). (That is, the path lies in cfree(D_0) except at the endpoints C_1 and C_2.)

Proof: (Figure 22) Find β and ψ to satisfy lemma C.4.19 for $\mathbf{A} = CD$[A], $\mathbf{B} = CD$[B]. Using corollary C.4.22, find $\epsilon_0, \gamma_0, \theta_0$ such that, if D' approximates D in tangent (γ_0, θ_0) and $\mathrm{p}^D(C, C') < \epsilon_0$ then $C'D'$ approximates CD in tangent (β, ψ). Using lemma C.4.20, there exists $\Delta > 0$ such that, for all $T \in (0,1]$, arc(C, M, Δ)(T) $\in$cfree(D). Let ϕ =arc($C, M, \min(\epsilon_0/2, \Delta)$) and let $\alpha =$ (1/2)clearance($\langle D, \phi(1)\rangle$) Using corollary C.4.22 again, find $\epsilon_1, \gamma, \theta$ such that, if D' approximates D (γ, θ) and $\mathrm{p}^D(C, C') < \epsilon_1$ then $C'D'$ approximates CD $(\min(\beta, \alpha), \psi)$. Finally, let $\epsilon = \min(\epsilon_1, \epsilon_0/2)$.

Now, let D_0 be a display that approximates D in tangent (γ, θ) and let C_1 and C_2 be feasible configurations over D_0 within ϵ of C such that $C_1[\mathrm{A}] = C_2[\mathrm{A}]$=$C$[A]. Let ϕ_1, ϕ_2 be paths parallel to ϕ starting at C_1, C_2; that is, $\phi_i(T) = \phi(T)C^{-1}C_i$ for $i = 1, 2$. Let ϕ_M be the uniform translation or rotation taking $\phi_1(1)$ into $\phi_2(1)$ through an angle less than π. We now claim that the path $\phi_0 = \langle \phi_1, \phi_M, \phi_2^{-1}\rangle$ is feasible.

We first note that, by construction C_1D_0 approximates CD in tangent (α, ψ) and therefore $d_H(C_1D_0, CD) < \alpha$. Let $\mathbf{p}$ be any point in D_0[B] and let $\mathbf{q}$ be a corresponding point in $D[B]$. Then $\mathrm{d}(C_1(\mathbf{p}), C(\mathbf{q})) < \alpha$ and $\mathrm{d}(C_2(\mathbf{p}), C(\mathbf{q})) < \alpha$. Since ϕ, ϕ_1, ϕ_2 all move in parallel along the motion M, it follows that, for every $T \in [0,1]$, $\mathrm{d}(\phi_1(T)(\mathbf{p}), \phi(T)(\mathbf{q})) = \mathrm{d}(\phi_1(0)(\mathbf{p}), \phi(0)(\mathbf{q})) = \mathrm{d}(C_1(\mathbf{p}), C(\mathbf{q}))$ $< \alpha$; and by the same token $\mathrm{d}(\phi_2(T)(\mathbf{p}), \phi(T)(\mathbf{q})) < \alpha$. Let $\mathbf{A} = CD$[A] and $\mathbf{A}_0 = CD_0$[A]. Since the position of object A is constant, $\mathbf{A}_0$ is the place of A along the entire path ϕ_1. Since $\mathrm{d}_H(\mathbf{A}, \mathbf{A}_0) < \alpha$, and since clearance($D, \phi(1)$) $> 2\alpha$, it follows that $\mathrm{d}(\phi(1)(\mathbf{q}), \mathbf{A}_0) > \alpha$, and so $\phi_1(1)(\mathbf{p}) \notin \mathbf{A}_0$. Thus $\phi_1(1)$ is feasible, and, by the same token, so is $\phi_2(1)$.

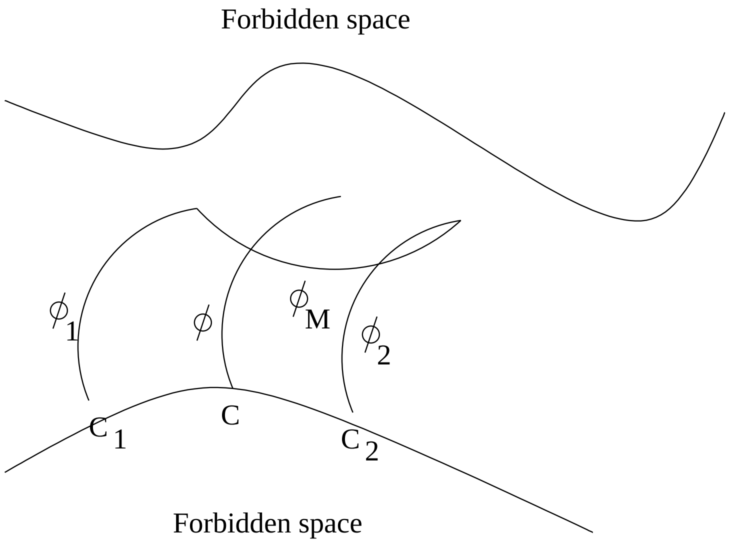


Figure 22: Illustration for lemma C.4.23

Moreover, the trace of point $\mathbf{p}$ along path ϕ_M is either a circle of arc at most π or a line segment. Hence the maximal value over C_M on ϕ_M of $\mathrm{d}(C_M(\mathbf{p}), \phi(1)(\mathbf{q}))$ is attained either at $C_M = \phi_1(1)$ or at $C_M = \phi_2(1)$ and so is always less than α. Hence $C_M(\mathbf{p})$ is never in $\mathbf{A}_0$, so the path ϕ_M is feasible over D_0.

To show that ϕ_1 is feasible over D_0, we use proof by contradiction. Suppose ϕ_1 is not feasible. Let T be the maximal value such that, for all $T' \in [0, T]$, $\phi_1(T')$ is feasible. Since free(D_0) is closed and since $\phi_1(0)$ is feasible, $\phi_1(T)$ is feasible. Now $\mathrm{p}^D(C, \phi_1(T)) \leq \mathrm{p}^D(C, C_1) + \mathrm{p}^D(C_1, \phi_1(T)) \leq \epsilon + \epsilon_0/2 < \epsilon_0$. Hence $\phi_1(T)(D_0)$ approximates CD in tangent (β, ψ). Let $\mathbf{X} = \phi_1(T)D_0[\mathrm{A}]$ and $\mathbf{Y} = \phi_1(T)D_0[\mathrm{B}]$. By lemma C.4.19 M strongly separates $\mathbf{Y}$ from $\mathbf{X}$. By lemma C.4.20, there is a $\Gamma > 0$ such that $\mathrm{arc}(\phi_1(T), M, \Gamma)$ escapes contact over D_0. But this path is just a continuation of the path ϕ_1 past time T, which is a contradiction.

The proof for ϕ_2 is the same as for ϕ_1. ∎.

Corollary C.4.24:. If D is always strongly separable, then free(D) is locally internally connected.

Proof: Let $D_0 = D$. Then lemma C.4.23 asserts that any two configurations in free(D) that are sufficiently close together are connected by a path through cfree(D). ∎.

Lemma C.4.25: Let D be a display over two objects A and B that is always strongly separable. Then there exist $\epsilon > 0, \gamma > 0, \theta > 0$ satisfying the following: Let D_0 be a display that approximates D in tangent (γ, θ); let C be a feasible configuration over D; and let C_1 and C_2 be feasible configurations over D_0 such that $C_1[\mathrm{A}] = C_2[\mathrm{A}] = C[\mathrm{A}]$, $\mathrm{p}^D(C, C_1) < \epsilon$ and $\mathrm{p}^D(C, C_2) < \epsilon$. Then there is a path connecting C_1 and C_2 through cfree(D_0).

Proof: Let $\Delta = \mathrm{diameter(D}[A]) + \mathrm{diameter}(D[\mathrm{B}])$, and let us constrain ϵ and γ to be less than Δ. We can then w.l.o.g. restrict attention to the compact space U of configurations in which the configuration of object A is the same as in C and and the distance from A to B is less than or equal to 2Δ.

We apply lemma B.8 yet again, U being this compact space of configurations, and $\Phi(C, \epsilon, \gamma, \theta)$ being the property "For any display D_0 that approximates D in tangent (γ, θ); and configurations

C_1 and C_2 that are feasible over D_0, if C_1[A] = C_2[A]=C[A], $\mathrm{p}^D(C, C_1) < \epsilon$ and $\mathrm{p}^D(C, C_2) < \epsilon$, then there is a path connecting C_1 and C_2 that is feasible over D_0." We must show that, if $C \in U$, there exists a $\delta > 0$, $\alpha > 0$, $\beta > 0$, $\psi > 0$ such that, for all C' within δ of C, $\Phi(C', \alpha, \beta, \psi)$. By lemma C.4.23, for any C there exists ϵ, γ, θ such that $\Phi(C, \epsilon, \gamma, \theta)$. The result is then immediate if we choose $\alpha = \delta = \epsilon/2$; $\beta = \gamma$, and $\psi = \theta$. ∎

We now have to generalize this to the case where object A is not fixed in C, C_1 and C_2. Intuitively, this is clear, since we can just view everything from A's reference frame. The laborious part of the proof is showing that small distances remain small under this change of reference frame.

Lemma C.4.26: Let D be a display over two objects A and B that is always strongly separable. Then there exist $\epsilon > 0, \gamma > 0, \theta > 0$ satisfying the following: Let D_0 be a display that approximates D in tangent (γ, θ); let C be a feasible configuration over D; and let C_1 and C_2 be feasible configurations over D_0 such that $\mathrm{p}^D(C, C_1) < \epsilon$ and $\mathrm{p}^D(C, C_2) < \epsilon$. Then there is a path connecting C_1 and C_2 through cfree(D_0). (This is the same as lemma C.4.25, but dropping the condition C_1[A] = C_2[A] = C[A].)

Proof: Let $\Delta = 2$ (diameter(D[A]) + diameter(D[B])). As in the proof of lemma C.4.25, we can confine attention to configurations C in which the diameter of CD[A] $\cup$ CD[B] is at most Δ, the lemma being trivial if it is more than Δ. Let E=Δ/diameter(D[A]). Find ϵ_0, γ, θ to satisfy lemma C.4.25. Let $\epsilon = \epsilon_0/(2E+1)$.

Now let D_0 be a display approximating D (γ, θ) and let C_1 and C_2 be feasible configurations over D_0 within ϵ of C. Define the configurations C'_1, C'_2 as follows

$C'_1[A] = C[A] = C[A]C_1^{-1}[A]C_1[A]$.
$C'_1[B] = C[A]C_1^{-1}[A]C_1[B]$.
$C'_2[A] = C[A] = C[A]C_2^{-1}[A]C_2[A]$.
$C'_2[B] = C[A]C_2^{-1}[A]C_2[B]$.

That is, C'_1 is obtained by applying the rigid transformation $C[A]C_1^{-1}[A]$ to both mappings in C_1. Therefore, the relative positions of objects A and B in C'_1 is the same as in C_1. Likewise, the relative positions of objects A and B in C'_2 is the same as in C_2.

Let Λ_1 be the transformation $C[A]C_1^{-1}[A]$ We have $\mathrm{p}^D(C'_1, C) = \max(\mathrm{p}^{D[A]}(C'_1[A], C[A]), \mathrm{p}^{D[B]}(C'_1[B], C[B]))$. The first of these is 0, so this is equal to $\mathrm{p}^{D[B]}(C'_1[B], C[B]))$. Let $\mathbf{a}$ be the point in $D[A]$ such that $\mathrm{d}(C_1[A](\mathbf{a}), C'_1[A](\mathbf{a}))$ is maximal and let $\mathbf{b}$ be the point in $D[B]$ such that $\mathrm{d}(C_1[B](\mathbf{b}), C'_1[B](\mathbf{b}))$ is maximal. Note that $C'_1[A] = \Lambda_1 C_1[A]$ and $C'_1[B] = \Lambda_1 C_1[B]$. Thus, if we define $\mathbf{u} = C_1[A]\mathbf{a}$ and $\mathbf{v} = C_1[B]\mathbf{b}$, then $\mathrm{d}(C_1[A](\mathbf{a}), C'_1[A](\mathbf{a})) = \mathrm{d}(\mathbf{u}, \Lambda_1(\mathbf{u}))$ and $\mathrm{d}(C_1[B](\mathbf{b}), C'_1[B](\mathbf{b})) = \mathrm{d}(\mathbf{v}, \Lambda_1(\mathbf{v}))$.

If Λ is a translation then $\mathrm{d}(\mathbf{v}, \Lambda_1(\mathbf{v})) = \mathrm{d}(\mathbf{u}, \Lambda_1(\mathbf{u}))$. If Λ is a rotation of angle α around center $\mathbf{o}$, then $\mathrm{d}(\mathbf{v}, \Lambda_1(\mathbf{v})) = 2\mathrm{d}(\mathbf{v}, \mathbf{o}) \sin(\alpha/2)$ and $\mathrm{d}(\mathbf{u}, \Lambda_1(\mathbf{u})) = 2\mathrm{d}(\mathbf{u}, \mathbf{o}) \sin(\alpha/2)$, so $\mathrm{d}(\mathbf{v}, \Lambda_1(\mathbf{v}))$ / $\mathrm{d}(\mathbf{u}, \Lambda_1(\mathbf{u})) = \mathrm{d}(\mathbf{v}, \mathbf{o})$ / $\mathrm{d}(\mathbf{u}, \mathbf{o})$. It is easily shown that this cannot be greater than $2E$.

Thus $\mathrm{p}^{D[B]}(C_1[B], C'_1[B]) = \mathrm{d}(C_1[B](\mathbf{b}), C'_1[B](\mathbf{b})) = \mathrm{d}(\mathbf{v}, \Lambda(\mathbf{v})) \leq 2E \cdot \mathrm{d}(\mathbf{u}, \Lambda_1(\mathbf{u})) = 2E \cdot \mathrm{d}(C_1[A](\mathbf{a}), C'_1[A](\mathbf{a})) = 2E \cdot (C_1[A](\mathbf{a}), C[A](\mathbf{a})) \leq 2E \cdot \mathrm{p}^D(C_1, C) = 2E\epsilon$. Also $\mathrm{p}^{D[B]}(C_1[B], C[B]) \leq \mathrm{p}^D(C_1, C) = \epsilon$. Therefore $\mathrm{p}^D(C'_1, C) = \mathrm{p}^{D[B]}(C'_1[B], C[B]) \leq \mathrm{p}^{D[B]}(C_1[B], C'_1[B]) + \mathrm{p}^D(C_1, C) = (2E+1)\epsilon = \epsilon_0$. Likewise $\mathrm{p}^D(C'_2, C) \leq \epsilon_0$.

We can therefore apply lemma C.4.25, and conclude that there is a path $\phi([0,1])$ from C'_1 to C'_2 through cfree(D_0). Let $C_M = \phi(1/2)$. Let ϕ_1 be the path $\Lambda_1^{-1}(\phi([0, 1/2]$ and let ϕ_3 be the path $\Lambda_2^{-1}(\phi([1/2, 1]$. Thus, ϕ_1 is a rigid transformation of the first half of ϕ going from C'_1 to $\Lambda_1^{-1}(\phi(1/2))$. Since it is a rigid transformation of $\phi([0, 1/2])$, the relative positions of A and B are the same as in $\phi([0, 1/2])$, so ϕ_1 is in cfree(D_0) except at the beginning. Likewise, ϕ_3 is a continuous path from $\Lambda_1^{-1}(\phi(1/2))$ to C'_2 which goes through cfree(D_0).

Finally, since the relative positions of A and B is the same in $\phi_1(1)$ and $\phi_3(0)$, let let ϕ_2 be a path between $\phi_1(1)$ and $\phi_3(0)$ that keeps the relative position of the two objects fixed. Then it is immediate that the path ϕ_1, ϕ_2, ϕ_3 connects C_1 to C_2 through cfree(D_0). ▮.

Theorem 4.2: Let D be a display over two objects A and B that is always strongly separable, such that free(D) is locally ordinarily connected. Then for any $\epsilon > 0$ there exist $\alpha > 0$ and $\theta > 0$ such that, if D' is a display that approximates D in tangent (α, θ), p^D_{tS}(paths(free(D)), paths(free(D')) $< \epsilon$.

Proof: Find $\epsilon_0, \gamma, \theta$ to satisfy lemma C.4.26. Let ϵ_1 be the minimal distance between any two path-connected components of free(D); since free(D) is ordinarily connected, $\epsilon_1 > 0$. Let $\epsilon_2 = (1/2)\min(\epsilon_1, \epsilon)$. Since free($D$) is the closure of cfree(D), use theorem 2.9 to find δ_0 such that, if $\mathrm{d}_{Hd}(D', D) < \delta_0$ then p^D_H(free(D),free(D')) $< \epsilon_2$. Since free(D) is locally internally connected, use theorem 3.5 to find δ_1 such that, if $\mathrm{d}_{Hd}(D', D) < \delta_1$ then p^D_{tH}(paths(free(D)),paths(free(D')) $< \epsilon_2$. Let $\alpha = \min(\gamma, \delta_0, \delta_1)$

Let D' be a display that approximates D in tangent (α, θ). We observe the following;

i. Since p^D_H(free(D),free(D')) $< \epsilon_2 \leq (1/2)\epsilon_0$, it follows that any configuration in free(D') is within ϵ_2 of exactly one connected component of free(D).

ii. Let U be any connected component of free(D). Let C'_1 and C'_2 be any configurations in free(D') within ϵ_2 of U. Let C_1 be a configuration in U within ϵ_2 of C'_1 and let C_2 be a configuration in U within ϵ_2 of C'_2. Since C_1 and C_2 are in U, let ϕ be a path from C_1 to C_2. By theorem 3.5, there is a path ψ through free(D') such that $\mathrm{p}^D_t(\phi, \psi) < \epsilon_2$. In particular, $\mathrm{p}^D(\phi(0), \psi(0)) < \epsilon_2$ and $\mathrm{p}^D(\phi(1), \psi(1)) < \epsilon_2$. Since both ψ_0 and C'_1 are within ϵ_2 of C_1, by lemma C.4.26, there is a path ψ_1 from C'_1 to $\psi(0)$ through free(D'). Similarly, there is a path ψ_2 from $\psi(1)$ to C'_2 through free(D'). The path ψ_1, ψ, ψ_2 thus goes from C'_1 to C'_2 within free(D').

iii. Thus, any two configurations C'_1 and C'_2 in free(D') within ϵ_2 of U are connected by a path through free(D'). We also know from (i) that every configuration in free(D) is within ϵ_2 of some configuration in free(D'), We can therefore conclude that for every connected component of free(D') there is a unique connected component of free(D) within ϵ_2, and vice versa.

iv. By the construction of δ_1, it follows that we have p^D_{tH}(paths(free(D)),free(D')) $< \epsilon_2$. It follows from the above that any path through a connected component of free(D) must be tracked by a path through the corresponding connected component of free(D'), since non-corresponding connected components are at least $\epsilon_1/2$ away. Hence, if U and V are corresponding connected components of free(D) and free(D'), respectively, $\mathrm{p}^D_{tH}(U, V) < \epsilon_2$. ▮

Proof of theorem 4.3

Theorem 4.3: Let D be a semi-algebraic display over two objects A and B that is always strongly separable. Then there exists an algebraic formula $\Phi(\epsilon, \alpha, \psi)$ with the following properties:

- For any $\epsilon > 0$ there exists $\alpha > 0$ and $\psi > 0$ such that $\Phi(\epsilon, \alpha, \psi)$.
- If $\Phi(\epsilon, \alpha, \psi)$ and D' is a display that approximates D in tangent (α, ψ), then p^D_{tS}(paths(free(D)), paths(free(D')) $< \epsilon$.
- The form of Φ can be computed from the forms of the regions in D.

Proof: Naturally, the construction of Φ follows the proof of theorem 4.2. We define the following formulas: (The subscripts on the formulas indicate the corresponding lemma. For example, Φ_{16} corresponds to lemma C.4.16.)

$\Phi_{16}(C, M, \delta, \theta)$ holds iff
for any point $\mathbf{p} \in CD[A] \cap CD[B]$,
for all points $\mathbf{q} \in$FBd$(CD[A], CD[B])$ and $\mathbf{r} \in$FBd$(CD[B], CD[A])$ within δ_{16} of $\mathbf{p}$
free_ang$(CD[A], \mathbf{p}, CD[B], \mathbf{q},$flow$(\mathbf{p}, M) > \theta$.

$\Phi_{19}(C, M, \beta, \psi)$ holds iff
there exist $\delta, \theta, D, \epsilon$, and γ such that
$\Phi_{16}(C, M, \delta, \theta)$ and
$\psi < \theta/2$ and $D = (1/2)\min((3/5)\mathrm{d}(CD[A] \cap CD[B], M),\ (4/5)\psi{\cdot}\mathrm{d}(CD[A] \cap CD[B], M)$, and
$\epsilon = \min(\delta/2, D)$ and
for all points $\mathbf{q} \in$FBd$(CD[A], CD[B])$ and $\mathbf{r} \in$FBd$(CD[A], CD[B])$
if $\mathrm{d}(\mathbf{q}, \mathbf{r}) < \gamma$ then
there exists a point $\mathbf{p} \in CD[A] \cap CD[B]$ within ϵ of both $\mathbf{p}$ and $\mathbf{q}$; and
$\beta = \min(\gamma, D)$.

$\Phi_{20}(C, M, \Delta)$ iff
for any transformation T,
if the motion of T is M and $\mathrm{p}^{D[B]}(C[B], T(C[B])) \leq \Delta$,
then the configuration $\langle C[A], T(C[B]) \rangle$ is contact-free over D.

$\Phi_{22}(C, \epsilon, \psi, \mu, \theta, \lambda)$ holds iff
$\mu > 0$, $\theta > 0$, and $\lambda > 0$, and
$\mu + \lambda \leq \epsilon$, $\lambda \leq \Delta/2$ and $\theta + 2\lambda/\Delta \leq \psi$,
where Δ is the minimum of the diameters of $D[A]$, $D[B]$.

$\Phi_{23}(C, M, \epsilon, \gamma, \theta)$ holds iff
$\epsilon, \gamma, \theta > 0$ and there exist $\beta, \psi, \epsilon_0, \gamma_0, \theta_0, \epsilon_1, \Delta, T, \alpha$ such that
$\Phi_{19}(C, M, \beta, \psi)$ and $\Phi_{22}(C, \beta, \psi, \gamma_0, \theta_0, \epsilon_0)$ and
$\Phi_{20}(C, M, \Delta)$ and
T is the transformation of motion M such that $\mathrm{p}^{D[B]}(C[B], T(C[B])) = \min(\epsilon_0/2, \Delta)$ and
α is the clearance of $\langle D, \langle C[A], T(C[B]) \rangle \rangle$, and
$\Phi_{22}(\min(\beta, \alpha), \psi, \epsilon_1, \gamma, \theta)$ and
$\epsilon = \min(\epsilon_1, \epsilon_0/2)$.

$\Phi_{25}(\epsilon, \gamma, \theta)$ iff
for every configuration C there exists a motion M for which $\Phi_{23}(C, M, \epsilon, \gamma, \theta)$.

$\Phi_{26}(\epsilon, \gamma, \theta)$ iff
$\Phi_{25}(\epsilon_0, \gamma, \theta)$ and
there exist Δ, E such that
$\Delta = 2(\mathrm{diameter}(D[A]) + \mathrm{diameter}(D[B]))$, $E = \Delta/\mathrm{diameter}(D[A])$ and $\epsilon = \epsilon_0/(2E + 1)$.

$\Phi(\epsilon, \alpha, \theta)$ iff
$\Phi_{26}(\epsilon_0, \gamma, \theta)$ and
there exists $\epsilon_1, \epsilon_2, \delta_0$ such that
ϵ_1 is the minimal distance between two path-connected components of free(D) and
$\epsilon_2 = (1/2)\min(\epsilon, \epsilon_1)$ and
$\mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(\mathrm{contract}(D, \delta_0)) \leq \epsilon_2$ and
$\mathrm{p}^D_H(\mathrm{free}(D), \mathrm{free}(\mathrm{expand}(D, \delta_0)) \leq \epsilon_2$ and
$\alpha = \min(\gamma, \delta_0)$.

It is straightforward from the proof of theorem 4.2 that Φ satisfies the conditions of theorem 4.3. A couple of subtle points do need clarification.

First, geometric properties of combinations of D with quantified variables, such as "$\mathbf{q} \in$FBd$(CD[A], CD[B])$" in Φ_{16} require a little care. We need to express the constraint here between $\mathbf{q}$ and C as an algebraic formula. One can show that, since D is semi-algebraic, there are only finitely many cases involved; that each case is defined by a semi-algebraic constraint on C and imposes a semi-algebraic constraint on $\mathbf{q}$.

Second, the form of Φ departs somewhat from the proof of theorem 4.2. The constraints $\mathrm{p}_H^D(\mathrm{free}(D),\ \mathrm{free}(\mathrm{contract}(D, \delta_0)) \leq \epsilon_2$ and $\mathrm{p}_H^D(\mathrm{free}(D),\ \mathrm{free}(\mathrm{expand}(D, \delta_0)) \leq \epsilon_2$ are sufficient to establish the condition $\mathrm{p}_H^D(\mathrm{free}(D),\mathrm{free}(D') < \epsilon_2$; the argument is analogous to the proof of theorem 2.9. As for the definition of δ_1 in the proof of theorem 4.2, it requires only a little more work to show that it is, in fact, unnecessary; the condition $\alpha = \min(\gamma, \delta_0)$ suffices. ▮

Proof of theorem 4.4

Theorem 4.4: Let D be a display over n objects that is always strongly separable, and such that free(D) is locally ordinarily connected. Then there exist $\alpha > 0$ and $\psi > 0$ such that, if D' is a display that approximates D in tangent (α, ψ), then the connected components of free(D') correspond to those of free(D).

Sketch of proof: The proof is exactly analogous to that of theorem 4.23. We have replaced circular motion by general continuous differentiable motion (even if we required each object to move in a circle, their relative motion would be a cycloid). However, the flow field of any differentiable motion at any instant is equal to the flow field of a uniform motion, and, by continuity, stays close to that uniform motion over some time. This is sufficient to carry out the above proof. ▮