

Growth Distances: New Measures for Object Separation and Penetration

Chong Jin Ong and Elmer G. Gilbert, *Fellow, IEEE*

Abstract—Quantitative measures for the separation and penetration of two convex objects are formulated. These measures, called separation and penetration growth distances, are closely related to traditional translational distance measures and share many of their desirable properties. The solution of a single optimization problem yields both the separation and penetration distances. For three-dimensional polytopes the optimization problem is a linear program in four variables whose asymptotic computational time is $O(m)$, where m is the number of linear inequalities required to specify the two polytopes. This equals or far better the known times required to compute translational distances: $O(m)$ for separation and $O(m^2 \log m)$ for penetration. When the positioning of the two objects depends on configuration variables, the partial derivatives of the growth distances with respect to the configuration variables exist almost everywhere. Moreover, for polytopes they can be evaluated with little numerical effort. The large speed advantage for penetration growth distance creates new opportunities for the algorithmic separation of intersecting objects. Specifically, derivatives of the penetration growth distance can be used to construct motions which separate the objects. An application to path finding is described.

I. INTRODUCTION

THE PROXIMAL relationship between two objects is of interest for many practical reasons. The most obvious is interference detection in robot motion planning. Other areas of interest include collision avoidance in machine-tool path planning, computer-aided design of mechanisms, and computer graphics. Knowing whether or not the mathematical representations of two objects intersect is a basic requirement in such areas, and algorithms for intersection detection have received considerable attention. However, in motion planning intersection detection alone is not sufficient. It is important to have *quantitative* measures of object separation or penetration. Separation and penetration distances serve this need. They show where path adjustment is needed and, more importantly, they give insight into how the adjustments should be carried out. Efficient computational procedures for the distances are crucial because in the planning process it is necessary to determine the distances for many potential placements of the objects. This paper and its predecessors [27], [16], [29] consider new distances, called growth distances, which measure

separation and penetration. They have desirable properties and can be computed very efficiently.

When two objects, represented by point sets A and B , are separated the most natural measure of separation is the shortest Euclidean distance between a pair of points contained in them: $d_S^2(A, B) = \min \|a - b\|_2, a \in A, b \in B$. The dependence of this separation distance on configuration variables which cause the objects to move is well studied [13]. Furthermore, algorithms for its computation are available for convex polytopes (see, e.g., [2], [9], [10], [14], [21], [22], and [34]) as well as for a variety of smooth convex objects [15].

Unlike $d_S^2(A, B)$, which has a clear physical meaning, a penetration distance is inherently conceptual in nature. After all, objects do not penetrate each other in the real world; they collide and are deformed or broken. The general objective of a penetration distance is to quantify *depth* of intersection for object *models*. There is an obvious motivation for studying penetration distances. They provide information on changes in object position which can reduce or eliminate intersection. Separation distances are not useful for this purpose, since they are identically zero on the set of positions which cause intersection. Past research on penetration distance is relatively sparse. To the best of our knowledge, substantial contributions are limited to [4]–[6], [8], [17], [26], and [31]–[33]. In most of these papers, the objects are modeled by convex polytopes and the penetration distance is defined in essentially the same way. Roughly speaking, it is the shortest relative translation of the two objects, measured by the Euclidean norm, that causes them to have no interior points in common. For objects in three-dimensional (3-D) space the computation of this penetration distance, $d_P^2(A, B)$, is very expensive [5], [17], [31], [33]. Consequently, its potential advantages in solving mechanical interference problems and in path planning [4]–[6] have not been realized. Our interest in growth distances grew out of a desire to obtain computationally efficient measures of penetration.

Growth distances are also defined for a pair of convex objects. They are a measure of how much each of the objects must be grown, outward from fixed seed points in their interiors, so that they just touch. When the grown objects are larger than the actual objects, the growth measures separation; when the grown objects are smaller than the actual objects, the growth measures penetration. Thus, the growth for touching determines *both* the separation and penetration distances. The main reason for introducing growth distances is their computational efficiency. Indeed, when the objects are polytopes in 3-D space the computation becomes the

Manuscript received May 6, 1995; revised November 8, 1995. This paper was recommended for publication by Editor R. A. Volz upon evaluation of reviewers' comments.

C. J. Ong is with the Department of Mechanical and Production Engineering, National University of Singapore, Singapore 0511, Republic of Singapore.

E. G. Gilbert is with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109-2118 USA.

Publisher Item Identifier S 1042-296X(96)05738-2.

solution of a linear program in only four variables whose computational time is $O(m)$, where m is the total number of faces in the two objects. It is also possible to compute growth distances efficiently for a general class of smooth convex objects [27]. Desirable properties of growth distances include invariance with respect to the choice of the origin and orientation of the coordinate system in which A and B are represented, continuous dependence of the distances on configuration variables describing the position and orientation of the objects, and under appropriate conditions, a simple characterization of the derivatives of the distances with respect to the configuration variables. The derivatives are important because they determine directions of object motion which either increase separation or decrease penetration. Separation and penetration growth distances are related by bounding expressions to the corresponding translational distances. When the objects are rotund, these bounds show that the growth distances closely approximate the translational distances.

Our separation and penetration growth distances appear to be new concepts. A related idea was discussed by Leven and Sharir [20]. However, they grow only one of the two objects and are mainly concerned with the construction of a generalized Voronoi diagram which they use for planning the translational motion of a convex object in a two-dimensional (2-D) space. Another approach based on growing one object is given in [23]. It has a serious weakness: in some situations the distance is discontinuous with respect to the relative motion of the two objects. Growing both objects rather than one is crucial to most of the desirable results that we obtain. Penetration measures, closer to the one defined in this paper, have appeared recently [26], [32], [35]. They are also obtained by shrinking both object models until they touch. However, the shrinking process is very different: each object boundary point is moved inward, either normal to a polyhedral face [26] or approximately normal to a smooth boundary [32], by an equal distance. This eliminates the need for seed points, and for small penetrations, the resulting penetration measure (twice the movement of the boundary points) closely approximates the translational penetration distance. Unlike penetration translational distance and penetration growth distance, the distance is not defined for all object pairs. The definition fails, for example, when one of the objects is slender. As the objects shrink, the slender object may become empty before the pair of objects just touch.

The following is an outline of the paper and its contents. Section II reviews in a general way prior results on separation and penetration distances. It also motivates the need for a computationally efficient penetration distance and allows contributions of this paper to be placed in a clear perspective. Object models and their growth representations are introduced in Section III. The fractional growth needed for the grown objects to just touch is the growth function, $g(A, B)$. Separation and penetration growth distances, which are defined in terms of $g(A, B)$, are considered in Section IV. Illustrative examples are given, and connections between growth distances and the translational distances are established. Linear programming (LP) approaches to polytopal objects are formulated in Section V. Numerical issues, includ-

ing some computational results, are considered in Section VI. When the polytopes are in a so-called regular configuration, partial derivatives of the growth function with respect to the configuration variables of the objects exist and may be obtained numerically with little effort. These results and their application to the separation of objects which intersect are developed in Section VII. Section VIII describes briefly an application of growth distance to path finding. Section IX summarizes key issues raised by the results of the paper. Proofs of the theorems are contained in the appendixes. Readers who desire an introductory treatment of the main ideas may omit the following material: the last seven paragraphs of Section II, the last four paragraphs of Section III, the last two paragraphs of Section V, all but the last two paragraphs of Section VII, and the Appendixes.

The following is a summary of needed notations. The empty set is \emptyset ; the nonnegative real numbers are R^+ ; $x \in R^n$ is a column vector with real components x^i ; x^T is the transpose of x ; $M \in R^{n \times n}$ is a real $n \times n$ matrix. Generic norms and p -norms on R^n are denoted, respectively, by $\|\cdot\|$ and $\|\cdot\|_p$; the Frobenius norm on $R^{n \times n}$ is $\|\cdot\|_F$. Origin-centered closed balls of radius r in R^n corresponding to the different norms are $N(r) = \{x \mid \|x\| \leq r\}$ and $N_p(r) = \{x \mid \|x\|_p \leq r\}$. The interior, boundary, closure, and complement of a set $A \subset R^n$ are, respectively, $\text{int } A$, $\text{bd } A$, $\text{cl } A$, and A' . The Minkowski sum or difference of $A, B \subset R^n$ is $A \pm B = \{a \pm b \mid a \in A, b \in B\}$. The multiplication of A by a scalar $\sigma \in R$ is $\sigma A = \{\sigma a \mid a \in A\}$. $SO(n)$ is the special orthogonal group of real $n \times n$ orthogonal matrices with determinant one, i.e., $T \in SO(n)$ is a rotation matrix. The hyperplane in R^n passing through p with normal $\eta \neq 0$ is $P(p, \eta) = \{x \mid \eta^T(x - p) = 0\}$; the half-space bounded by $P(p, \eta)$ with outward normal η is $H(p, \eta) = \{x \mid \eta^T(x - p) \leq 0\}$.

The class of sets in R^n that are compact, convex, and have a nonempty interior is denoted by O . It is easy to confirm that $A, B \in O$ and $\sigma \neq 0$ imply $A \pm B \in O$ and $\sigma A \in O$. It is also known [30] that $\text{cl}(\text{int } A) = A$. Thus, A is fully determined by its interior, a property which is expected of solid physical objects. Moreover, physical objects are finite in extent and include their surfaces (boundary points). Hence, O is the natural class of sets for modeling convex objects. In the representation of objects $n = 2$ or 3 . However, in what follows, no special use is made of this fact. If the object A undergoes a change of configuration, i.e., a rotation $T \in SO(n)$ followed by a translation $s \in R^n$, it is represented by $TA + \{s\}$. For a detailed discussion of these matters see [18]. Obviously, $A \in O$ implies $TA + \{s\} \in O$.

II. SEPARATION AND PENETRATION DISTANCES

A pair of objects has three intuitive, qualitative states of proximity: separation, penetration, and touching. For objects modeled by sets A and B in O it is possible to give these states a precise meaning: *separation* $\Leftrightarrow A \cap B = \emptyset$, *penetration* $\Leftrightarrow \text{int } A \cap \text{int } B \neq \emptyset$, *touching* $\Leftrightarrow A \cap B \neq \emptyset$, and $\text{int } A \cap \text{int } B = \emptyset$. Clearly, the three states of proximity are exhaustive and mutually exclusive.

Quantitative measures of separation and penetration should conform to the above definitions. Specifically, a separation

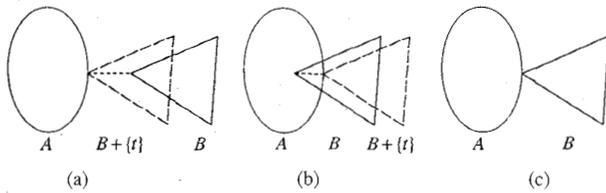


Fig. 1. Determination of t : (a) separation, (b) penetration, and (c) touching.

distance, $d_S(A, B)$, should be positive for separation and zero otherwise, and a penetration distance, $d_P(A, B)$, should be positive for penetration and zero otherwise. It should be emphasized that such distances are *not* metrics on O . Instead, they conform to our physical concept of distance between objects, i.e., they measure the amount of relative motion which is needed to bring A and B into a touching position.

We now develop these ideas more fully. The purpose is to review in a common setting results and different points of view which have appeared in the literature. It is understood throughout that $A, B \in O$.

The preceding remarks suggest the following definitions for d_S and d_P :

$$d_S(A, B) = \inf\{\|t\| \mid A \cap (B + \{t\}) \neq \emptyset\} \quad (1)$$

$$d_P(A, B) = \inf\{\|t\| \mid A \cap (B + \{t\}) \neq \emptyset\}. \quad (2)$$

For obvious reasons we call these distances *translational* distances. It is also possible, with a significant increase in complexity, to develop distances which are based on rotations as well as translations of the objects [28]. Definitions (1) and (2) are similar to those of Cameron and Culley [6] but avoid their need for explicitly using a concept of touching (contact, in their terminology). It is easy to confirm that $d_S(A, B)$ and $d_P(A, B)$ are defined for all A and B in O . Also, they are nonnegative and satisfy the desired positivity conditions for separation and penetration. Note that a generic norm is used to measure the translation and that the actual values of the distances depend on a particular choice of the norm. When p -norms are chosen, the distances will be denoted by $d_S^p(A, B)$ and $d_P^p(A, B)$. Geometric interpretations for the t 's which characterize the infima are illustrated in Fig. 1 for $p = 2$.

The distances $d_S(A, B)$ and $d_P(A, B)$ have simple properties. It is easily confirmed that they are symmetric in their arguments and are invariant with respect to common translations of A and B : $d(A, B) = d(B, A)$, $d(A + \{s\}, B + \{s\}) = d(A, B)$ for all $s \in R^n$. Ideally, the distances should be independent of changes in the position and orientation of the coordinate system. This means that they should be invariant with respect to both rotations and translations, i.e., $d(TA + \{s\}, TB + \{s\}) = d(A, B)$ for all $T \in SO(n)$ and all $s \in R^n$. This property holds *only* for the $p = 2$ norm, and it is undoubtedly the underlying reason for the almost universal adoption of d_S^2 and d_P^2 in the literature.

Both d_S and d_P can be computed as solutions of optimization problems. Note that $A \cap (B + \{t\}) \neq \emptyset$ if and only if $t \in A - B$. Thus, (1) yields

$$\begin{aligned} d_S(A, B) &= \min\{\|t\| \mid t \in A - B\} \\ &= \min\{\|a - b\| \mid a \in A, b \in B\}. \end{aligned} \quad (3)$$

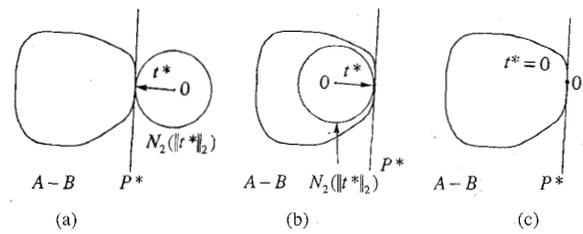


Fig. 2. Determination of t^* from $A - B$: (a) separation, (b) penetration, and (c) touching.

It is acceptable to replace the “inf” by “min” because A, B and $A - B$ are compact, and norms are continuous functions. Similarly, but with a few more details for the proof, it can be shown that

$$\begin{aligned} d_P(A, B) &= \min\{\|t\| \mid t \in \text{cl}(A - B)\} \\ &= \max\{r \mid N(r) \subset (A - B)\}, \quad 0 \in A - B \\ &= 0, \quad 0 \notin A - B. \end{aligned} \quad (4)$$

Unlike the situation for d_S , there is no simple way of expressing d_P directly in terms of A and B . The characterization in terms of $N(r) \subset A - B$ appears in [17] (with sup instead of max) and is implicit in the expressions of [6]. It states that $d_P(A, B)$ is the radius of the largest origin-centered ball in $A - B$. The characterization of penetration distance given by Buckley [5] is basically a variational principle for the minimization problem in (4).

Geometric interpretations for $d_S^2(A, B)$ and $d_P^2(A, B)$ in terms of $A - B$ are shown in Fig. 2 for the cases of separation, penetration, and touching. The minimizers in (3) and (4) are denoted by t^* . In all three cases $t^* \in \text{bd}(A - B)$. It is of interest to note that the sets $N(\|t^*\|)$ and $A - B$ have a common support hyperplane P^* passing through t^* .

The distances d_S and d_P have special properties which justify their application in iterative algorithms for moving object models apart (when they intersect) and farther apart (when they are separated). Specifically, separation (penetration) implies the existence of an incremental motion which increases d_S (decreases d_P). To be precise, $d_S(A, B) > 0$ implies $d_S(A, B)$ increases locally, i.e., there exist arbitrarily small $s \in R^n$ such that $d_S(A + \{s\}, B) > d_S(A, B)$. Similarly, $d_P^2(A, B) > 0$ implies $d_P(A, B)$ decreases locally. Fig. 2(a) and (b) illustrates the simple reasoning behind the proof of these properties. Clearly, $(A + \{s\}) - B = (A - B) + \{s\}$. Choosing $s = \alpha t^*$, where α is small and positive, causes the size of N_2 to change appropriately when $A - B$ is replaced by $A - B + \{s\}$.

The local increasing (decreasing) properties do not hold for *all* separation (penetration) distances [17], [33]. Consider, for example, penetration measured by intersection volume: $d_P^V(A, B) = \text{volume } A \cap B$. While $d_P^V(A, B)$ satisfies all of the other properties of a penetration distance mentioned above, it is not always locally decreasing. Suppose $A \subset \text{int } B$; then $d_P^V(A + \{s\}, B) = \text{volume of } A$ for all $s, \|s\|$ sufficiently small.

The computation of d_S^2 for the case of convex polytopes has been studied at length. Dobkin and Kirkpatrick [9] give an algorithm for $n = 3$ which exploits detailed structural models

of A and B , whose run time is $O(m)$, where $m = m_A + m_B$ and m_A and m_B represent the complexity of A and B (either the number intersecting half-spaces or the number of vertices). In [10] they show that a related procedure has an even faster asymptotic time, $O(\log m_A \times \log m_B)$. However, it assumes that a preprocessing of the representations for A and B has taken place, where the preprocessing times are, respectively, $O(m_A)$ and $O(m_B)$. Other procedures for the $n = 3$ case have been described by Gilbert *et al.* [14], Bobrow [2], [34], and Lin *et al.* [21], [22]. While good asymptotic complexity bounds do not exist for the algorithms in [14], [2], and [34], they get by with simple structural representations (A and B are described either by vertices or by faces) and have efficient, empirically demonstrated performance. Moreover, the algorithmic idea in [14] extends to a wide class of objects with curved boundaries [15]. While the algorithms in [21], [22] have a large speed advantage when the objects move continuously and d_S^2 must be evaluated on a closely spaced sequence of object configurations, they too demand structurally more complex representations of A and B . For polytopes the computation d_S^1 and d_S^∞ can be reduced to the solution of LP problems in $2n+1$ variables and roughly m inequalities. Even though a computational time $O(m)$ is possible [17], it is doubtful that there is a significant speed advantage over the better 2-norm algorithms.

The computation of d_P for convex polytopes is a much more difficult problem and has received relatively little attention. Procedures for computing d_P^2 have been described in [4]–[6], [8], [17], and [33]. The basic idea is to work with $A - B$ represented by an intersection of half-spaces and solve (4) by finding the half-space which is nearest to the origin. The approaches differ in determining the half-space representation of $A - B$, the most time-consuming part of the computation. For $n = 2$, algorithms have been described which have complexity $O(m)$. For $n = 3$, no algorithms are known [31] which have asymptotic performance better than $O(m^2 \log m)$. Keerthi and Sridharan [17] emphasize the striking advantage of working with d_P^1 and d_P^∞ . Specifically, they show that d_P^1 or d_P^∞ can be computed by solving several linear programs in $n + 1$ variables and m inequality constraints. When $n = 3$ the number of LP problems is either six, for the 1-norm, or eight for the ∞ -norm. In both cases the computational time is $O(m)$ [24]. Sridharan *et al.* [33] discuss briefly the case where $A, B \subset R^2$ are nonconvex polygons.

When the position and orientation of objects A and B depend on configuration variables, the question of numerically evaluating the gradients of d_S and d_P with respect to the configuration variables arises. Complete results are given in [13] for d_S^2 . It is shown that the gradient exists for almost all configurations and is given by a simple formula. Where the gradient fails to exist, a generalized gradient [7] exists, and it too is characterized by a simple formula. Similarly complete results are lacking for d_S^p , when $p \neq 2$, and d_P^p , for all p . An approximate approach for evaluating the gradient of d_P^2 , when it exists, is described in [5].

III. GROWTH MODELS AND THE GROWTH FUNCTION

The formulation of the growth distances has three main features: a modeling of the objects, which endows them with a

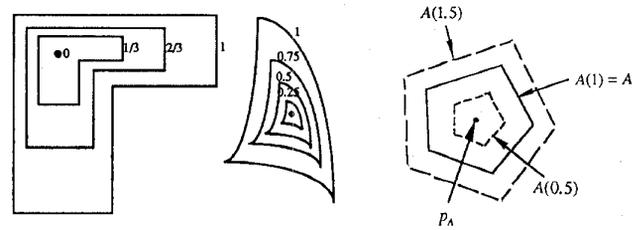


Fig. 3. Growth models.

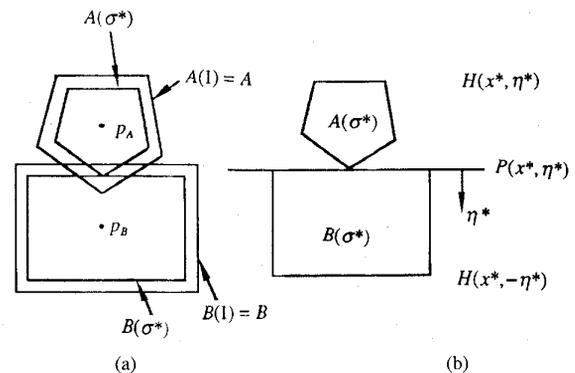


Fig. 4. (a) Determination of σ^* . (b) Geometric configuration of $A(\sigma^*)$ and $B(\sigma^*)$.

fractional growth σ ; the growth, σ^* , which causes them to just touch; and an appropriate scaling of $\sigma^* - 1$, which generates the separation and penetration distances. In this section the first two features are addressed. Since σ^* depends on the choice of the two objects, it is called the *growth function* and is given a special notation: $g(A, B) = \sigma^*$.

A growth model for an object $A \subset R^n$ is a family of compact, nonempty sets $A(\sigma) \subset R^n$, defined for $\sigma \in R^+$, which satisfies the following, somewhat loosely stated, conditions: a) $A(\sigma)$ varies continuously with σ ; b) $A(\sigma)$ is strictly increasing in σ ; c) $A(0) = \{p_A\}$, the seed point; d) $A(1) = A$, the object; e) $A(\sigma)$ covers all of R^n as $\sigma \rightarrow +\infty$. Fig. 3 shows some examples of growth models in R^2 . The growth function for an object pair, $A, B \subset R^n$, is

$$g(A, B) = \sigma^* = \min\{\sigma \in R^+ | A(\sigma) \cap B(\sigma) \neq \emptyset\}. \quad (5)$$

The touching of $A(\sigma^*)$ and $B(\sigma^*)$ required by (5) is illustrated by the example in Fig. 4(a). Conditions a)–e) have the following motivations: e) guarantees $A(\sigma) \cap B(\sigma) \neq \emptyset$ for some $\sigma \in R^+$; a) implies the existence of the minimum in (5); d) determines the relative size of the two growth models so that the sign of $\sigma^* - 1$ determines separation or penetration of A and B ; and b) and c) are required so that $g(A + \{s\}, B)$ increases locally in s (so that the corresponding separation and penetration growth distances respectively increase and decrease locally).

Objects possess many growth models which satisfy the conditions. To establish mathematically precise results and to provide effective computational procedures, it is necessary to introduce specific assumptions about the class of objects and the growth model. As in Section II and in much of the literature on object distances, it is assumed hereafter that

$A, B \in O$. It is also possible to define growth distances for convex objects which have empty interiors. This generalization is not pursued here because it introduces uninteresting complications and has limited value in representing physical objects. As Fig. 3 suggests, it is possible to develop growth models for nonconvex objects. However, this brings in serious mathematical and computational complications. Instead, in what follows, we shall treat nonconvex objects by the familiar idea of representing each of them by a union of convex sets and then defining the distance as the least distance (or the sum of the distances) between all pairs of the convex sets in which one set is taken from each of the objects. Of course, this requires additional effort. The union must be created (for most physical objects a simple inspection suffices), and the number of object pairs is increased.

Our growth model for $A \in O$ is

$$A(\sigma) = \{p_A\} + \sigma\bar{A}, \quad \sigma \in R^+ \quad (6)$$

where p_A and \bar{A} satisfy the conditions

$$p_A \in \text{int } A \quad \text{and} \quad \bar{A} = A - \{p_A\}. \quad (7)$$

Fig. 3(c) is an example of (6). Note \bar{A} is the object A with the origin shifted to p_A . Clearly, $\bar{A} \in O$ and $0 \in \text{int } \bar{A}$.

Requirements a)–e) are satisfied. Conditions a), c), and d) are direct consequences of (6) and (7). Since $0 \in \text{int } \bar{A}$, there exists $r > 0$ such that $N(r) \subset \bar{A}$. Thus, $A(\sigma) \supset \{p_A\} + \sigma N(r)$, which shows that condition e) holds. Let σ_1 and σ_2 satisfy $0 \leq \sigma_1 \leq \sigma_2$. By (6), $A(\sigma_2) = A(\sigma_1) + (\sigma_2 - \sigma_1)\bar{A}$. This implies $A(\sigma_2) \supset A(\sigma_1) + (\sigma_2 - \sigma_1)N(r)$, which confirms condition b).

The growth model (6) is not unique; it is defined for any seed point $p_A \in \text{int } A$. There are, however, guidelines for choosing p_A . As will be seen in the next section, the growth distances more closely approximate the translational distances when seed points are placed well inside the objects. Hereafter, we do make one additional assumption about seed points. When an object undergoes a rigid-body motion, it is assumed that its seed point maintains an object-fixed position. Thus, if A is given a rotation $T_A \in SO(n)$ and a translation $s_A \in R^n$, its growth model is

$$\begin{aligned} (T_A A + \{s_A\})(\sigma) &= \{T_A p_A + s_A\} + \sigma(T_A \bar{A} + \{s_A\} - \{T_A p_A + s_A\}) \\ &= \{T_A p_A + s_A\} + \sigma T_A \bar{A}. \end{aligned} \quad (8)$$

Precise statements of key results concerning the growth function are contained in the following theorem.

Theorem 3.1: Assume $A, B \subset O$ have growth models of the form (6), (7). Then:

- i) $g(A, B)$ is defined by (5), i.e., the minimum exists;
- ii) separation $\Leftrightarrow g(A, B) > 1$; penetration $\Leftrightarrow g(A, B) < 1$;
touching $\Leftrightarrow g(A, B) = 1$;
- iii) $g(A, B) = 0 \Leftrightarrow p_A = p_B$;
- iv) $g(A + \{s\}, B)$ increases locally in s .

Because of i), g is indeed a function $g: O \times O \rightarrow R^+$. From ii) it follows that $g(A, B) - 1$ is a numerical measure of separation and penetration whose sign distinguishes separation and penetration. By iii), maximum penetration corresponds to

$g(A, B) - 1 = -1$. From (5) and (8) it is easy to confirm that g is symmetric in its arguments and is invariant with respect to common rigid body motions of the two objects: $g(A, B) = g(B, A)$ and $g(TA + \{s\}, TB + \{s\}) = g(A, B)$.

The convexity of $A(\sigma^*)$ and $B(\sigma^*)$ gives additional meaning to their geometric configuration. In particular, they are separated by a hyperplane. See Fig. 4(b) and the last paragraph of Appendix A.

While (5) has a nice geometric interpretation, its set theoretic nature sheds little light on the computation of $g(A, B)$. Actually, $g(A, B)$ is often determined by the solution of a simple, convex optimization problem. Consider an example where $A \subset R^3$ is an ellipsoid and $B \subset R^3$ is a truncated circular cylinder: $A = \{x | x^T Q x \leq 1\}$, $Q \in R^{3 \times 3}$ is positive definite; $B = \{x | 0 \leq x^1 \leq 2, 4(x^2)^2 + 4(x^3)^2 \leq 1\}$. Then suitable, centered choices for the seed points are $p_A = 0$ and $p_B = \{1 \ 0 \ 0\}^T$. Hence, $A(\sigma) = \sigma\bar{A} = \{x | x = \sigma\bar{x}, \bar{x}^T Q \bar{x} \leq 1\} = \{x | x^T Q x \leq \sigma^2\}$, and $B(\sigma) = p_B + \sigma\bar{B} = \{x | x = p_B + \sigma\bar{x}, (\bar{x}^1)^2 \leq 1, 4(\bar{x}^2)^2 + 4(\bar{x}^3)^2 \leq 1\} = \{x | (x^1 - 1)^2 \leq \sigma^2, 4(x^2)^2 + 4(x^3)^2 \leq \sigma^2\}$. The condition $A(\sigma) \cap B(\sigma) \neq \emptyset$ is satisfied if and only if there exists an $x \in R^3$ such that $x \in A(\sigma)$ and $x \in B(\sigma)$. Thus, $g(A, B) = \sigma^*$ is determined by solving the convex programming problem as follows:

$$\begin{aligned} &\text{minimize } \sigma \text{ with respect to } (\sigma, x) \in R^4 \text{ satisfying} \\ &x^T Q x \leq \sigma^2, \quad (x^1 - 1)^2 \leq \sigma^2, \quad 4(x^2)^2 + 4(x^3)^2 \leq \sigma^2. \end{aligned} \quad (9)$$

Obviously, the simple ideas expressed in the example can be extended to a great variety of objects in O . The resulting optimization problems are generally convex and can be solved by well-developed, efficient algorithms. The special case of polytopes is treated in detail in Section V.

A completely general approach for setting up an optimization problem is to exploit the Minkowski distance functions of \bar{A} and \bar{B} . Technical details can be found in Appendix A. Given \bar{A} and $x \in R^n$, the Minkowski distance function, $\mu_{\bar{A}}(x)$, is the smallest value of $\lambda \in R^+$ such that $x \in \lambda\bar{A}$. It is shown in Appendix A that $A(\sigma) = \{x | \mu_{\bar{A}}(x - p_A) \leq \sigma\}$. Thus, $g(A, B) = \sigma^*$ is obtained by minimizing σ with respect to $(\sigma, x) \in R^{n+1}$, subject to the constraints $\mu_{\bar{A}}(x - p_A) \leq \sigma$, $\mu_{\bar{B}}(x - p_B) \leq \sigma$. While the Minkowski functions are not particularly useful for computations, they are valuable tools in the proofs of our basic theorems.

Finally, it is important to determine how g changes when A and B undergo general rigid-body motions. For this purpose let the quadruple $P = (T_A, s_A, T_B, s_B) \in (SO(n) \times R^n)^2$ denote the rotations and translations assigned to the objects A and B . Then the dependence of the growth function on P is expressed by

$$G(P) = g(T_A A + \{s_A\}, T_B B + \{s_B\}). \quad (10)$$

A key result is that G is Lipschitz continuous. To state this result precisely, a metric on $(SO(n) \times R^n)^2$ is needed. It is obtained [18] by imbedding $(SO(n) \times R^n)^2$ in $(R^{n \times n} \times R^n)^2$. Since

$$\begin{aligned} \rho(P, \tilde{P}) &= \|T_A - \tilde{T}_A\|_F + \|s_A - \tilde{s}_A\|_2 + \|T_B - \tilde{T}_B\|_F \\ &\quad + \|s_B - \tilde{s}_B\|_2 \end{aligned} \quad (11)$$

is a metric on $(R^{n \times n} \times R^n)^2$, it becomes a metric on $(SO(n) \times R^n)^2$.

Theorem 3.2: Assume $A, B \subset O$ have growth models of the form (6), (7). Let $Q = \{P \in (SO(n) \times R^n)^2 \mid \|s_A\|_2 \leq c_A, \|s_B\|_2 \leq c_B\}$, where c_A and c_B are arbitrary positive constants. Then there exists a $k > 0$, which depends on c_A and c_B , such that $P, \tilde{P} \in Q$ implies $|G(P) - G(\tilde{P})| \leq k\rho(P, \tilde{P})$.

The Lipschitz continuity implies that $G(P)$ is (Fréchet) differentiable almost everywhere in Q [7]. This property is crucial in developing algorithmic procedures for the separation of intersecting objects, a subject which is discussed in Sections VII and VIII.

IV. THE GROWTH DISTANCES

It is evident from Theorem 3.1 that $g(A, B) - 1$ provides a distance-like measure of separation and penetration. However, $g(A, B) - 1$ is a dimensionless quantity related to the *fractional growth* of A and B . As such it does not conform to our intuitive concept of a distance measure; it must be scaled in some way according to the *size* of A and B . Let $S_{AB} > 0$ be the scaling coefficient. Then the separation and penetration growth distances are defined by

$$d_S^G(A, B) = S_{AB}(g(A, B) - 1), \quad g(A, B) \geq 1 \\ = 0, \quad g(A, B) < 1 \quad (12)$$

$$d_P^G(A, B) = S_{AB}(g(A, B) - 1), \quad g(A, B) \geq 1 \\ = 0, \quad g(A, B) > 1. \quad (13)$$

The precise value of S_{AB} is not too important as long as the distances between objects of different size and shape are handled in a consistent, reasonable way. As will be seen, Theorem 4.1 quantifies the notion of “reasonable” by establishing bounds between growth distances and translational distances.

Two measures for the size of a set are the radii of inner and outer spheres. Specifically, $\{x_A\} + N_2(r_A)$ and $\{x_A\} + N_2(R_A)$ are, respectively, inner and outer spheres for $A \in O$ if

$$\{x_A\} + N_2(r_A) \subset A \subset \{x_A\} + N_2(R_A). \quad (14)$$

For a given $x_A \in \text{int } A$, it is clear that values of r_A and R_A exist such that $0 < r_A \leq R_A$. In fact, it is possible [19] to maximize r_A and minimize R_A subject to (14). The resulting extreme values for r_A and R_A depend on the choice of x_A , and there are best choices in the sense that r_A is maximized or R_A is minimized. Simple examples of $A \subset R^2$ show that best choices of x_A for each of the two objectives need not be the same. In addition, the optimum choices may be difficult to compute. However, for most physical objects it is possible by inspection to choose x_A so that the ratio r_A/R_A is at or near its maximum value. If A is slender, the ratio r_A/R_A is necessarily small; if A is rotund, $r_A/R_A \cong 1$; if A is a sphere, choosing x_A as its center gives $r_A = R_A$.

If A and B are spheres it is clear how the seed points and scalings of $g(A, B) - 1$ should be selected. Let $A = \{x_A\} + N_2(R_A)$ and $B = \{x_B\} + N_2(R_B)$. Obviously, the translational distances between A and B are $d_S^2(A, B) = \|x_A - x_B\|_2 - (R_A + R_B)$ for $\|x_A - x_B\|_2 - (R_A + R_B) > 0$ and $d_P^2(A, B) = -\|x_A - x_B\|_2 + (R_A + R_B)$ for $\|x_A - x_B\|_2 -$

$(R_A + R_B) < 0$. It is easy to see that $d_S^G(A, B)$ and $d_P^G(A, B)$ match these distances if and only if $p_A = x_A, p_B = x_B$ and $S_{AB} = R_A + R_B = r_A + r_B$. For nonspherical objects there exist bounds relating the translational and growth distances.

Theorem 4.1: Assume that $A, B \in O$ have seed points which satisfy the conditions

$$\{p_A\} + N_2(r_A) \subset A \subset \{p_A\} + N_2(R_A) \quad \text{and} \\ \{p_B\} + N_2(r_B) \subset B \subset \{p_B\} + N_2(R_B) \quad (15)$$

where $0 < r_A \leq R_A$ and $0 < r_B \leq R_B$. Let $S_{AB} = R_A + R_B$ and define $\alpha = (r_A + r_B)/(R_A + R_B)$. Then

$$\alpha d_S^G(A, B) \leq d_S^2(A, B) \leq d_S^G(A, B) \quad \text{and} \\ d_P^2(A, B) \leq d_P^G(A, B). \quad (16)$$

Assume further that $d_P^2(A, B) \leq r_A + r_B$. Then

$$\alpha d_P^G(A, B) \leq d_P^2(A, B). \quad (17)$$

Several points deserve emphasis. The assumption $d_P^2(A, B) \leq r_A + r_B$, added for (17), appears to be overly strong. Actually, (17) rarely fails when $d_P^2(A, B) > r_A + r_B$. The bounds (16) and (17) are tightest when the coefficient α is at its maximum value. This condition is approximated when the seed points p_A and p_B are picked so that the ratios r_A/r_B and r_B/R_B are near their maximum values. Thus, Theorem 4.1 provides a guideline for centering the seed points. When the objects are both rotund, it is possible to have $\alpha \cong 1$ so that the translational and growth distances are approximately equal. For a pair of slender objects, α is necessarily small, and the translational and growth distances can be expected to differ significantly. Choices for S_{AB} in the range $r_A + r_B \leq S_{AB} < R_A + R_B$ are also appropriate. They modify (16) and (17) in obvious ways. For example, $S_{AB} = r_A + r_B$ gives $d_S^G(A, B) \leq d_S^2(A, B) \leq \alpha^{-1} d_S^G(A, B)$.

It is clear from the results of Section III that $d_S^G(A, B)$ and $d_P^G(A, B)$ share the key properties of the translational distances $d_S^2(A, B)$ and $d_P^2(A, B)$: $d_S^G(A, B) > 0 \Leftrightarrow$ separation; $d_P^G(A, B) > 0 \Leftrightarrow$ penetration; they are symmetric functions of A and B . They remain invariant with respect to common rotational and translational motions of A and B ; when positive they are, respectively, locally increasing and decreasing with respect to relative translations of A and B . From Theorem 3.2 and the piecewise definition of $d_S^G(A, B)$ and $d_P^G(A, B)$ in terms of $g(A, B) - 1$, the functions

$$D_S^G(P) = d_S^G(T_A A + \{s_A\}, T_B B + \{s_B\}) \\ D_P^G(P) = d_P^G(T_A A + \{s_A\}, T_B B + \{s_B\}) \quad (18)$$

which describe the dependence of the growth distances on changes in the configurations of A and B , are Lipschitz continuous.

Examples can be used to further substantiate the reasonable behavior of the growth distances. One approach is to compare variations in the translational and growth distances as one object is moved relative to the other. Fig. 5 shows such a comparison when A is given a translation s relative to B . Fig. 5(a) defines an object pair in R^2 . For both objects the best inner and outer spheres have common centers at

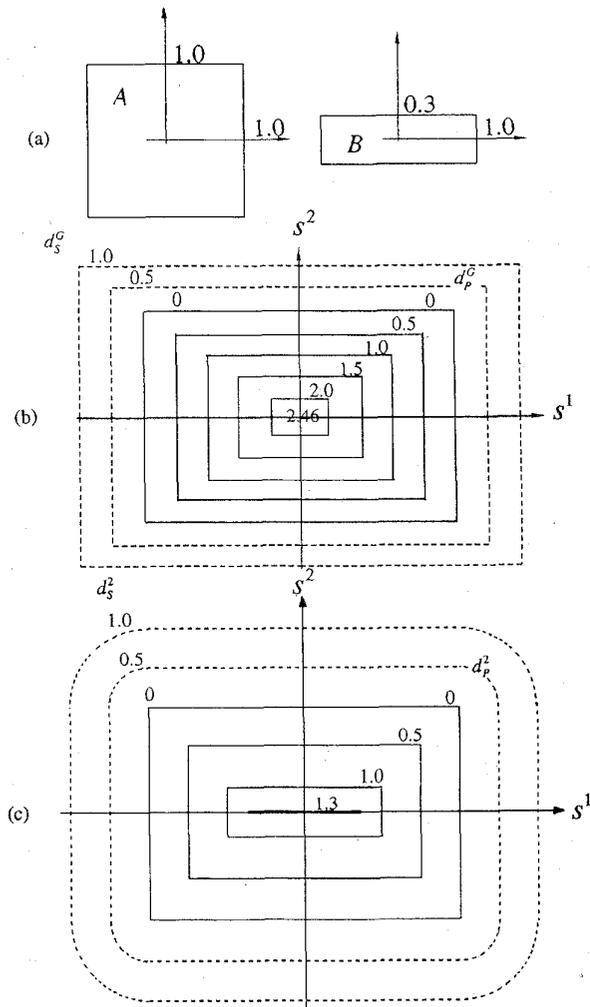


Fig. 5. Level curves for translation of A by s . Dashed curves show separation distances; solid curves show penetration distances. (a) Objects A and B . (b) Growth distances. (c) Translational distances.

the origin so that the natural choice for the seed points is $p_A = p_B = 0$. With this choice, $R_A = \sqrt{2}, r_A = 1, R_B = \sqrt{1.09}, r_B = 0.3, \alpha \cong 0.529$. Assume $S_{AB} = R_A + R_B$. Fig. 5(b) shows how the growth distances change as a function of s ; level curves are shown in the s -plane for the functions $d_S^G(A + \{s\}, B)$ and $d_P^G(A + \{s\}, B)$. As expected, the deepest penetration occurs for $s = 0$ and is equal to $R_A + R_B \cong 2.458$. Fig. 5(c) shows the corresponding level curves for the translational distances $d_S^2(A + \{s\}, B)$ and $d_P^2(A + \{s\}, B)$. Fig. 6 is like Fig. 5, except B is rotated by 45° .

In both Figs. 5 and 6 the level curves for the growth and translational distances are similar. For $d_S^2 = d_P^2 = 0$ and $d_S^G = d_P^G = 0$ they are exactly the same. This is because these conditions correspond to those translations which cause the objects to touch. For $d_P^2(A + \{s\}, B)$ there is more than one translation, s , where the penetration is deepest. This lack of uniqueness is common when at least one of the objects has parallel edges or faces. For d_P^G there is only one translation of B which produces the deepest penetration. This result is true for all $A, B \in O$: by part iii) of Theorem 3.1, $A + \{s\}$

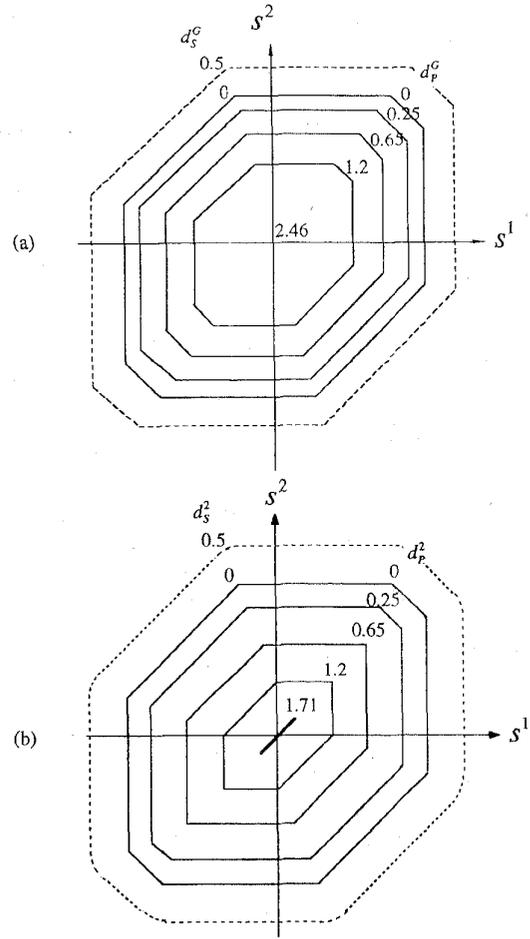


Fig. 6. Level curves resulting from a rotation of B by 45° : (a) growth distances and (b) translational distances.

and B have their deepest penetration, S_{AB} , if and only if $s = p_B - p_A$. For both the translational and the growth distances, the level sets in Figs. 5 and 6 are convex. This result also remains valid for all $A, B \in O$. The proof for the growth distances follows immediately from the convexity of $g(A + \{s\}, B)$ in s ; see Appendix A. Figs. 5 and 6 also confirm the locally decreasing (increasing) property of penetration (separation) growth distance. For example, when $d_P^G(A + \{s\}, B) > 0$ there exist small changes in s which cause $d_P^G(A + \{s\}, B)$ to decrease.

Figs. 5 and 6 confirm the upper and lower bounds in Theorem 4.1. As s takes on all values, the translational distances in Fig. 5 are bounded from above and below by coefficients of 0.814 and 0.529 applied to the corresponding growth distances. For Fig. 6 the coefficients are 0.982 and 0.697. Thus, the coefficients of 1.0 and 0.529 provided by the theorem are quite tight.

V. COMPUTATIONAL FORMULATIONS FOR POLYTOPES

Physical objects are commonly represented by convex polytopes. The resulting object models have simple data structures and, as an abundant literature demonstrates, they are well

suiting to a variety of computational objectives. These advantages also apply to the growth distances. In what follows it is assumed that reasonably placed seed points have been selected and that by shifting the origin we have obtained the polytopes $\bar{A} = A - \{p_A\}$ and $\bar{B} = B - \{p_B\}$.

In our first computational approach, we model \bar{A} and \bar{B} by linear inequalities

$$\begin{aligned} \bar{A} &= \{x | \gamma_i^T x \leq 1, i \in I_A\} \\ \bar{B} &= \{x | \gamma_i^T x \leq 1, i \in I_B\}. \end{aligned} \tag{19}$$

Here, $I_A = \{1, \dots, m_A\}$ and $I_B = \{m_A + 1, \dots, m_A + m_B\}$, m_A and m_B are the number of half-spaces whose intersection defines \bar{A} and \bar{B} , and the γ_i are (nonzero) outward normals to the half-spaces. Note that the representations are consistent with the requirements, $0 \in \text{int } \bar{A}$ and $0 \in \text{int } \bar{B}$, stated in Section III. To obtain the computational procedure for $g(A, B)$ we proceed as in the derivation of (9). From (6) it follows that $x \in A(\sigma)$ if and only if x satisfies $\gamma_i^T(x - p_A) \leq \sigma, i \in I_A$. More generally, $x \in A(\sigma) \cap B(\sigma)$ if and only if

$$\gamma_i^T x - \sigma \leq \delta_i, \quad i \in I = I_A \cup I_B = \{1, \dots, m\} \tag{20}$$

where $\delta_i = \gamma_i^T p_A, i \in I_A, \delta_i = \gamma_i^T p_B, i \in I_B$ and $m = m_A + m_B$ is the total number of half-spaces. Therefore, $g(A, B) = \sigma^*$, where (σ^*, x^*) solves the LP problem as follows:

$$\begin{aligned} &\text{minimize } \sigma \text{ with respect to } (\sigma, x) \in R^{n+1} \text{ satisfying} \\ &\text{the constraints (20)}. \end{aligned} \tag{21}$$

While this LP problem is of low dimension (3 or 4), m depends on the complexity of the object pair and may be large. Algorithms are available [24] which solve (21) in $O(m)$ time. Moreover, the special structure of (21) permits an easy determination of an initial feasible point. Given any $\tilde{x} \in R^n$, compute

$$\tilde{\sigma} = \max\{\gamma_i^T \tilde{x} - \delta_i | i \in I\}. \tag{22}$$

Then $(\tilde{\sigma}, \tilde{x})$ satisfies (20) and $\tilde{\sigma}$ is as small as possible, given the choice of \tilde{x} .

In our second approach, \bar{A} and \bar{B} are represented by their vertices: $\nu_i \in R^n, i \in I_A$ and $\nu_i \in R^n, i \in I_B$. Specifically, \bar{A} and \bar{B} are defined by the convex hulls

$$\begin{aligned} \bar{A} &= \{x = \lambda_1 \nu_1 + \dots + \lambda_{m_A} \nu_{m_A} | \lambda_1 \geq 0, \dots, \lambda_{m_A} \geq 0, \\ &\quad \lambda_1 + \dots + \lambda_{m_A} = 1\} \\ \bar{B} &= \{x = \lambda_{m_A+1} \nu_{m_A+1} + \lambda_{m_A+m_B} \nu_{m_A+m_B} | \lambda_{m_A+1} \geq 0, \\ &\quad \dots, \lambda_{m_A+m_B} \geq 0, \lambda_{m_A+1} + \dots + \lambda_{m_A+m_B} = 1\}. \end{aligned} \tag{23}$$

Clearly, $x \in A(\sigma)$ requires $x = p_A + [\nu_1 \dots \nu_{m_A}] \alpha_A, \alpha_A \in R^{m_A}$, where $\alpha_A^1 + \dots + \alpha_A^{m_A} = \sigma$ and $\alpha_A^i \geq 0, i = 1, \dots, m_A$. Using a similar expression for $x \in B(\sigma)$, it follows that $x \in A(\sigma) \cap B(\sigma)$ if and only if

$$\begin{aligned} &\begin{bmatrix} \nu_1 & \dots & \nu_{m_A} & -\nu_{m_A+1} & \dots & -\nu_{m_A+m_B} \\ 1 & \dots & 1 & -1 & \dots & -1 \end{bmatrix} \begin{bmatrix} \alpha_A \\ \alpha_B \end{bmatrix} \\ &= \begin{bmatrix} p_B - p_A \\ 0 \end{bmatrix}, \quad \alpha_A \geq 0, \quad \alpha_B \geq 0. \end{aligned} \tag{24}$$

Thus, $g(A, B) = \sigma^*$ is obtained by solving the LP problem as follows:

$$\begin{aligned} &\text{minimize } \sigma = [1 \dots 1] \alpha_A \text{ with respect to} \\ &(\alpha_A, \alpha_B) \in R^{m_A+m_B} \text{ satisfying the constraints (24)}. \end{aligned} \tag{25}$$

While this LP is of relatively high dimension, $m_A + m_B$, it has only $n + 1$ constraints. It is of interest to note that the dual [1], [12], [19] of (25), which is an LP obtained by a simple rearrangement of the data in (24), has dimension $n + 1$ and $m = m_A + m_B$ constraints. Thus, (25) has an underlying complexity $O(m)$, where m is the total number of vertices in the two objects. There is no obvious, simple procedure for finding an initial feasible point for either (25) or its dual. Thus, a two-phase LP approach [1] is needed.

VI. NUMERICAL ISSUES

In our numerical work on polytopal objects we have employed the first of the two approaches described above. This choice is in part motivated by the availability of the simple initialization procedure, (22). Stated computational times are for the solution of (21) and do not include the times required to prepare supporting data. These latter times are relatively small. Well-located seed points can usually be located by object inspection, and R_A and R_B are determined by the vertices of \bar{A} and \bar{B} which are farthest from the origin. Moreover, the supporting data need to be determined only once and are not changed as the objects take on different positions.

Of course, many algorithms are available for solving (21). Those designed to guarantee the best asymptotic computational time, $O(m)$, are not likely to be as efficient in practice as traditional algorithms which typically have $O(m)$ performance. We have used the active set approach [1], [12] since it provides several advantages in data manipulation. For instance, it can be applied directly to (21) without introducing the additional artificial variables required by the simplex method. Our computational programs follow closely the plan outlined in [1] and do not exploit in any way the special structure of (21).

The number of steps taken by the LP algorithm is likely to be less if the initial feasible point approximates (σ^*, x^*) . This suggests that \tilde{x} in (22) should approximate x^* . Obvious choices for \tilde{x} lie along the line segment joining p_A and p_B . A simple choice of this type is the halfway point $\tilde{x} = 1/2(p_A + p_B)$. Generally, it causes one constraint to be active, namely the index in (22), which yields the maximum. We have also tried the phase-one, LP approach to generating a feasible initial point. It more than doubles the total time of solution for (21).

Many numerical computations have been performed with two computers. They involved a variety of different objects in R^3 with m ranging between 8 and 400. The characterizations of (19) were minimal in the sense that removal of any inequality forces a change in \bar{A} or \bar{B} . All objects were contained within outer spheres of radius 5. For each object pair the growth distances were computed for 50 random placements. Object rotations were generated by using a Euler angle parameterization of $SO(3)$ and choosing the angles from a uniform distribution on $[-\pi, \pi]$. Translation vectors

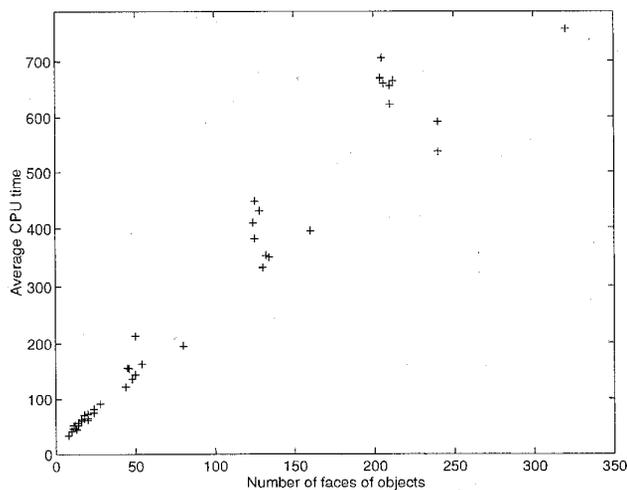


Fig. 7. Computational time in microseconds for polytopal objects. SG computer.

were generated by choosing their components from a uniform distribution on $[-3, 3]$. These choices provided a rich family of separation and penetration configurations.

FORTTRAN code, with an optimal run-time compilation, was run on an HP-Apollo 400 and on a Silicon Graphics Extreme (150 MHz). Respective LINPAC bench marks [11] are 1.9 MFLOPS and 24 MFLOPS. For each of nearly 100 object pairs, run times were averaged over the random placements. Unlike the situation with translational distances, the average times did not differ for separation and penetration cases. Data for the Silicon Graphics machine are shown in Fig. 7. There is some scatter, but the general dependence on m is nearly linear with the average time given approximately by $4m$ microseconds. The corresponding data for the HP-Apollo give $100m$ microseconds.

It is difficult to make precise comparisons with computational times for translational distances. There are many differences in the algorithms, computer hardware and software, representations of the polytopes (vertex, half-space, full geometric structure), and the selection of example problems. We have made comparisons for separation distance using the widely disseminated FORTRAN code by Gilbert *et al.* [14] which we believe is as fast as any of the codes which are based on algorithms that use only vertex [14] or half-space [2], [34] representations of the polytopes. For identical object pairs on the same machine, the average times for d_S^G were about half those required for d_S^2 . Few computational results are available for d_P^2 when $n = 3$ [5], [6]. They are limited to very simple examples, and comparisons are difficult to make because of differences in computer hardware and software. Adjusting as best as is possible for the differences, it appears that the times are at least an order of magnitude greater than those required for d_P^G . In view of the known [31] bound, $O(m^2 \log m)$, the situation should be much worse for large problems. Thus, from the standpoint of computational speed, growth distance is a very much preferred measure of penetration.

In a variety of applications it is necessary to compute the distance between pairs of objects as they move continuously along a specified geometric path. For example, d_S^G has been

used to determine where collision occurs along the path [14], [21], [22] or to assure that the path remains collision free as the path is adjusted for optimality [13]. And, as will be described in Section VIII, we have used d_P^G to solve path-finding problems. In all of these applications it is necessary to carry out distance evaluations on a closely spaced grid of points along the path. It has been noted [14], [21], [22] that the closeness of the grid points may be exploited to reduce the total computational time.

The idea is to use data from the computation of the distance at grid point k to initialize the distance algorithm for the computation at grid point $k + 1$. Suppose the growth function, as given by (7), has been computed at $P = P_k$. Let (σ_k^*, x_k^*) denote the solution of the corresponding LP. The inequality constraints which are active at the solution are known. As equalities they form a linear system which (σ_k^*, x_k^*) satisfies. See the next section for additional details. If P_{k+1} is close to P_k , it is highly likely that the constraints which are active at the optimal solution remain the same. Thus, an excellent candidate for $(\sigma_{k+1}^*, x_{k+1}^*)$ is determined by solving the linear system associated with P_{k+1} and the same index set. The presumed optimality of the candidate is tested by the optimality conditions: positivity of the Lagrange multipliers and satisfaction of the inequality constraints (20). On those infrequent occasions when the optimality conditions are not satisfied, i.e., when the set of constraints active at $k + 1$ is different from the set at k , the LP algorithm is brought into play. We call this scheme maintaining active constraints (MAC).

Fig. 8 confirms the expected pattern of behavior of MAC when it is applied to a simple example run on the HP-Apollo 400. Object A (a cylinder with decagonal ends, $m_A = 12$) moves along a straight line path while rotating at a constant rate; object B (an irregular polytope, $m_B = 14$) is fixed. The need to check the optimality condition increases the overhead time of the LP algorithm from about 1 millisecond to 1.4 milliseconds. For most of the 100 grid points the trial candidate succeeds, and the corresponding time is essentially the overhead time (approximately 1.4 milliseconds). When the number of grid points falls below 50, the speed advantage of MAC begins to decrease significantly.

As it has been described, MAC takes $O(m)$ time because testing optimality requires verification of the constraints (20). Lin and Canny [21] argue that their algorithm for d_S^2 has a more favorable outcome. When the number of grid points is large, the computational time is independent of m . The reason is a more elaborate data structure for representing the polytopes. Geometric adjacency of the constraints is stored so that optimality can be tested using only adjacent constraints. Adjacency information can be used in similar ways for the computation of $g(A, B) = \sigma^*$. We are currently exploring alternative data structures and algorithmic details and will report results in a later paper.

VII. EVALUATION OF MOTION DERIVATIVES

When A and B undergo rigid-body motions, as described by $T_A A + \{s_A\}$ and $T_B B + \{s_B\}$, the resulting growth function is given by $\sigma^* = G(P)$ in (10). In this section a procedure

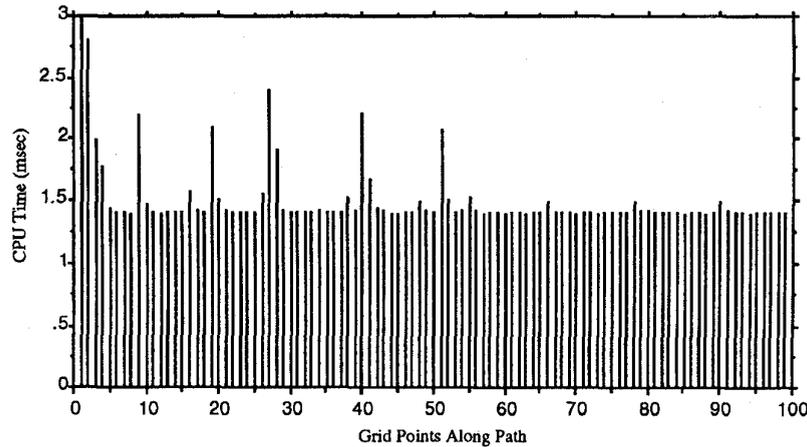


Fig. 8. Computational times at each grid point for MAC procedure. HP computer.

for computing derivatives of $G(P)$ is developed when A and B are polytopes. Similar procedures apply also to more complex object pairs such as the ellipsoid/cylinder example of Section III. Generally, the procedures are more efficient and accurate than finite-difference methods.

For polytopes, $G(P)$ is evaluated by introducing an obvious change of variables in (19) or (23). We limit our attention to the half-space representation (19), although a similar development holds for the vertex representation (23). Clearly, σ^* solves (21) with (20) replaced by

$$\begin{aligned} \gamma_i^T (T_A^T(x - s_A) - \sigma) &\leq \delta_i, & i \in I_A \\ \gamma_i^T T_B(x - s_B) - \sigma &\leq \delta_i, & i \in I_B. \end{aligned} \quad (26)$$

Evaluating the derivative of $G(P) = \sigma^*$ is a more complicated question. It has been noted in Section III that $G(P)$ has an ordinary (Frechet) derivative almost everywhere. There are, however, special points where the ordinary derivative does not exist. At these points more general notions of the derivative are needed [7], [25]. Here we emphasize the important, and technically less difficult, question of characterizing the ordinary derivative where it exists.

To simplify our notations, we focus our attention on the partial derivative with respect to one of the several configuration variables which may determine the relative position and orientation of the object pair. Denote the variable by θ . Then $P = P(\theta) = (T_A(\theta), s_A(\theta), T_B(\theta), s_B(\theta))$. It is assumed that $P(\theta)$ is defined and continuously differentiable on an open interval of R . This is certainly a valid assumption in practical problems where P is a smooth kinematic function of physical variables such as joint angles. Let $\hat{\gamma}_i(\theta) = T_A(\theta)\gamma_i, \delta_i(\theta) = \delta_i + \hat{\gamma}_i^T(\theta)s_A(\theta), i \in I_A$ and $\hat{\gamma}_i(\theta) = T_B(\theta)\gamma_i, \delta_i(\theta) = \delta_i + \hat{\gamma}_i^T(\theta)s_B(\theta), i \in I_B$. Then, $G(P(\theta))$ is obtained by solving the LP problem

$$\min_{\sigma, (x, \sigma) \in R^{n+1}}, \quad \hat{\gamma}_i^T(\theta)x - \sigma \leq \delta_i(\theta), \quad i \in I. \quad (27)$$

Let the solution of (27) be $(\sigma^*(\theta), x^*(\theta))$. Then $G(P(\theta)) = \sigma^*(\theta)$. We wish to obtain $d\sigma^*/d\theta$.

As indicated above, the derivative may fail to exist even though $\hat{\gamma}_i(\theta)$ and $\delta_i(\theta)$ are differentiable. An example is

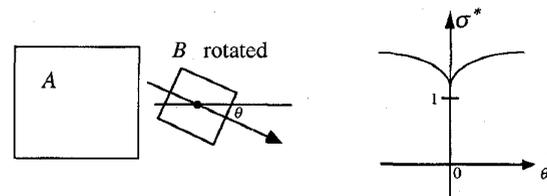


Fig. 9. Example where growth function is not differentiable.

shown in Fig. 9. A is fixed and θ is the rotation angle of B about the fixed center of B . For $-\pi/2 < \theta < \pi/2$ the derivative, $d\sigma^*/d\theta$, exists everywhere except at $\theta = 0$. The derivative fails to exist there because the qualitative nature of the geometric contact between $A(\theta^*(\theta))$ and $B(\sigma^*(\theta))$ is in transition. For $-\pi/2 < \theta < 0, x^*(\theta)$, is uniquely and smoothly determined. It is the upper left vertex of $B(\sigma^*(\theta))$. However, for $0 < \theta < \pi/2, x^*(\theta)$ is the lower left vertex of $B(\sigma^*(\theta))$. It is this transition from one contact vertex to another that causes the trouble. Note that $x^*(0)$ is not unique; it can be any point in the left side of $B(\sigma^*(0))$. The derivative may fail to exist in entirely different geometric situations. Consider, for instance, a similar example in R^2 where the contact between $A(\sigma^*(0))$ and $B(\sigma^*(0))$ consists of the touching of two vertices. Then $x^*(0)$ is unique, but the separating hyperplane between $A(\sigma^*(0))$ and $B(\sigma^*(0))$ is not. This situation again causes a contact transition to occur.

The nature of the geometric contact at $x^*(\theta)$ is determined by the constraints which are active in the LP at the optimum. Assume hereafter that the representations (19) are minimal, i.e., there are no redundant inequalities. Let $I^* = I_A^* \cup I_B^*$ be the index set of constraints active in (27) at $x^*(\theta), \sigma^*(\theta)$, i.e., those i for which the inequalities are satisfied as equalities. The sets I_A^* and I_B^* , which correspond to the active indexes taken from I_A and I_B , are both nonempty because $A(\sigma^*)$ and $B(\sigma^*)$ must have only boundary points in common; see Appendix A.

It is the cardinality of I_A^* and the cardinality of I_B^* that determine the nature of the geometric contact. Consider the case where $n = 2$. Then, the cardinality of I_A^* must be either 1 or 2. One active constraint means $x^*(\theta)$ belongs to an open edge of $A(\sigma^*(\theta))$; two active constraints mean $x^*(\theta)$

is a vertex of $A(\sigma^*(\theta))$. Similarly, the cardinality of I_B^* is either 1 or 2. When $n = 3$ the situation is more complicated. Possible cardinalities of I_A^* and I_B^* are 1, 2, or $l \geq 3$. One active constraint means $x^*(\theta)$ belongs to an open face, two active constraints means $x^*(\theta)$ belongs to an open edge, and l active constraints means $x^*(\theta)$ is a vertex. Clearly, l is the number of faces which meet at the vertex. It can be shown that any three of the l inequalities written as equalities form a linear system which uniquely determines the vertex.

Intuitively, the ordinary derivative exists at those values of θ , where the contact between $A(\sigma^*(\theta))$ and $B(\sigma^*(\theta))$ is *regular*, i.e., where the positioning of $A(\sigma(\theta))$ and $B(\sigma^*(\theta))$ is generic.

When $n = 2$ *nongeneric* contacts are edge-edge (open edge intersecting open edge, as in Fig. 9 at $\theta = 0$) and vertex-vertex. These contacts are delicate; arbitrarily small changes in P can produce several different edge-vertex contacts. Thus, the derivative may fail to exist at such points of contact. Regular contacts are edge-vertex contacts. They occur when there are exactly three active constraints, one from one object and two from the other object. The linear system which is obtained by writing the three active constraints as equalities is nonsingular and remains so for small changes in θ . From the differentiability of $P(\theta)$ it follows that $x^*(\theta)$ and $\sigma^*(\theta)$ are uniquely defined and are differentiable.

When $n = 3$ the regular contacts are face-vertex (open face of one object contains only a vertex of the other object) and edge-edge (open edges intersect at one point). In a face-vertex contact there is one active constraint from one object and $l \geq 3$ active constraints from the other object; in an edge-edge contact there are exactly two active constraints from each object. In both cases the linear system resulting from the active constraints again shows that $x^*(\theta)$ and $\sigma^*(\theta)$ are uniquely defined and are differentiable.

While the above statements are intuitively obvious, their precise justification requires a somewhat lengthy argument which is not pursued here. Note that when (27) has been solved, regularity is easily determined by examining the number of active constraints taken from I_A and I_B .

The computation of the derivative can now be addressed. Suppose the contact is regular at $\theta = \theta_0$. Then the set of active constraints remains the same for small changes in θ . Hence, there exists an $\epsilon > 0$ such that

$$\hat{\gamma}_i^T(\theta)x^*(\theta) - \sigma^*(\theta) = \hat{\delta}_i(\theta), \quad i \in I^*, \quad \text{for all } |\theta - \theta_0| < \epsilon. \quad (28)$$

By the above discussion this system has a unique solution $z(\theta) = (\sigma^*(\theta), x^*(\theta)) \in R^{n+1}$ at $\theta = \theta_0$. Except for a face-vertex contact where $n = 3$ and $l > 3$, there are $n+1$ equations in $n+1$ unknowns. When $l > 3$, three equalities can be selected to determine the vertex. Thus, (28) can be written

$$C(\theta)z(\theta) = c(\theta), \quad |\theta - \theta_0| < \epsilon \quad (29)$$

where $z(\theta) = (\sigma^*(\theta), x(\theta)) \in R^{n+1}$, the components of $c(\theta) \in R^{n+1}$ are $\hat{\delta}_i(\theta)$, and the rows of $C(\theta) \in R^{(n+1) \times (n+1)}$ are $[-1 \ \hat{\gamma}_i(\theta)^T]$. By our assumptions, the elements of $c(\theta)$ and

$C(\theta)$ are differentiable, and $C(\theta_0)$ is nonsingular. Differentiating (29) at $\theta = \theta_0$ gives

$$\frac{dz}{d\theta}(\theta_0) = C^{-1}(\theta_0) \left[\frac{dc}{d\theta}(\theta_0) - \frac{dC}{d\theta}(\theta_0)z(\theta_0) \right]. \quad (30)$$

The derivatives on the right side can be computed from the formulas just above (27). Moreover, the solution of (27) at $\theta = \theta_0$ gives $C(\theta_0)$ and $z(\theta_0)$. In fact, the active set method of solving the LP automatically provides $C^{-1}(\theta_0)$ [1], [12]. Thus, the additional computational effort required to obtain the derivatives of $\sigma^*(\theta)$ and $x^*(\theta)$ is small. Computation of the kinematic derivatives is likely to consume most of the time.

Derivatives of the growth distances, $D_S^G(P(\theta))$ and $D_P^G(\theta)$, are easily determined from the derivative of $G(P(\theta)) = \sigma^*$ by using (12), (13), and (18) and taking into account the piecewise nature of the definitions (12) and (13). Respectively, for penetration and separation

$$\sigma^*(\theta) < 1 \Rightarrow \begin{aligned} \frac{dD_S^G(P(\theta))}{d\theta} &= 0 \\ \frac{dD_P^G(P(\theta))}{d\theta} &= -S_{AB} \frac{d\sigma^*(\theta)}{d\theta} \end{aligned} \quad (31)$$

$$\sigma^*(\theta) > 1 \Rightarrow \begin{aligned} \frac{dD_P^G(P(\theta))}{d\theta} &= 0 \\ \frac{dD_S^G(P(\theta))}{d\theta} &= S_{AB} \frac{d\sigma^*(\theta)}{d\theta}. \end{aligned} \quad (32)$$

If $\sigma^*(\theta) = 1$, the objects touch and the distance derivatives may not exist because of the piecewise character of (12) and (13).

Distance derivatives are valuable when it is necessary to adjust a separation or penetration situation. A common example is in path optimization, where d_S^2 has been used in penalty terms to avoid object collision; see, e.g., [13]. In such applications d_S^2 may be replaced by d_S^G with savings in time for both the computation of the penalty terms and their derivatives. When object models are in collision, it is often useful to generate a move which separates them. For this purpose efficient computation of $D_P^G(P)$ and its derivatives is especially advantageous.

Consider the details for a simple separation problem. Suppose an object $T_A(\theta) + s_A(\theta) \subset R^n$, with a configuration vector $\theta \in R^p$, resides in a field of obstacles $B_1, \dots, B_M \subset R^n$. Penetration with the field of obstacles is measured by

$$f(\theta) = \sum_{i=1}^M d_P^G(T_A(\theta)A + \{s_A(\theta)\}, B_i). \quad (33)$$

This cost is nonnegative and is zero for all collision-free configurations of A . Since the gradient of $f(\theta)$ can be computed using the chain rule and the methods described above, object A can be moved into free space by applying a descent algorithm to $f(\theta)$. There is a possible difficulty. The approach fails if descent converges to a positive local minimum of $f(\theta)$. In practice, the difficulty does not appear to be a serious weakness. It has not occurred in our computations on a variety of separation problems. Another concern is the possible nonsmooth character of $f(\theta)$; conventional descent algorithms

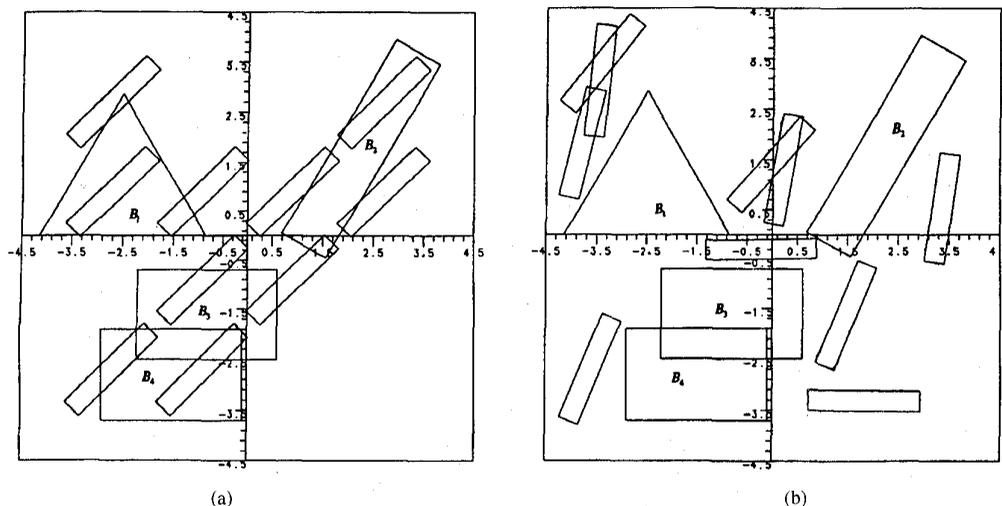


Fig. 10. (a) Placements of A resulting in collisions. (b) Positions of A after minimizing penetration cost.

may fail to converge to the minimum [7]. Again, our actual experience has been favorable. We have employed quasi-Newton minimization algorithms [12] with excellent success. In fact, the exact minimum $f(\theta) = 0$ is usually obtained in a few iterations. There is a simple explanation for this remarkable behavior. Usually, the set of collision-free configurations has a fairly large nonempty interior. Eventually, a descent step crosses the set and produces the global minimum.

Fig. 10 is an example in a two-dimensional square work space [3]. The configuration θ has three components: the horizontal and vertical translations of A and the rotation angle of A . There are $M = 8$ obstacles: the three rectangles, the triangle, and four overlapping squares that represent the boundaries of the square work space. Note that the union of B_3 and B_4 represents a single nonconvex object. Fig. 10(a) shows ten different collision configurations for A . Fig. 10(b) shows the collision-free configurations resulting from descent. Not surprisingly, most of the collision-free positions are relatively close to the originating collision positions.

VIII. PATH PLANNING

The idea of moving a collision configuration into free space generalizes to collision paths in configuration space. In principle, the basic approach is simple: form a cost of collision along a parameterized path and then minimize the cost with respect to the parameters of the path. In this section we outline these steps in the simplified setting described in the last two paragraphs of the preceding section. A more general and complete treatment can be found in [27] and [29].

Let $q(t, \alpha) \in R^p, 0 \leq t \leq 1$ denote a path in configuration space of object A , parameterized by $\alpha \in R^N$. For example, the components of α might represent the coefficients in spline function representations of the components of $q(t, \alpha)$. We assume that the parameterization is arranged so that specified end conditions are met, i.e., $q(0, \alpha) = q_0$ and $q(1, \alpha) = q_1$ for all $\alpha \in R^N$. The cost of collision along the path is measured by adding penetration costs, (33), evaluated on a

grid of configurations along the path

$$J(\alpha) = \sum_{l=1}^{L-1} f(q(l/L, \alpha)). \quad (34)$$

Clearly, $J(\alpha) = 0$ if and only if the path is collision-free at the grid points $q(l/L, \alpha)$. Collision does not occur on the entire path if L is sufficiently large and $J(\alpha) = 0$.

An initial parameterized path with collisions, $q(t, \alpha_0)$, is moved into free space by minimizing $J(\alpha)$. Again, by using the chain rule and results from Section VII, it is possible to compute the derivatives of $J(\alpha)$. Thus, iterative descent algorithms for the minimization of $J(\alpha)$ can be implemented. The numerical problem is much larger than the one considered at end of the previous section because $N = pK$, where K is the number of parameters determining each of the components of $q(t, \alpha)$. If a free-space path exists and the initial path is not too far from free space, local minima will not be encountered, and a collision-free path in the neighborhood of the initial path will be obtained. Usually, if a free space path exists, there are infinitely many free-space paths and the minimizers of $J(\alpha)$ are an open set. As in the object separation problem this makes it likely that the iterative descent will stop in a finite number of iterations with $J(\alpha) = 0$. As with many heuristic methods, there is no guarantee of completeness. The descent may converge to a local minimum where $J(\alpha) > 0$. However, there are possible strategies for avoiding local minima [29].

Fig. 11 shows an example which is based on the same objects and work space that are defined in Fig. 10 [3]. Each of the three components of θ is represented by a cubic, twice continuously differentiable spline with 11 parameters. Thus, $N = 33$. The grid points are spaced by 0.01 so that $L = 100$. The minimization procedure was a quasi-Newton BFGS (Broyden, Fletcher, Goldfarb, and Shanno) algorithm [12]. It proved to be more reliable and efficient than steepest descent and conjugate gradient algorithms.

The initial path appears in Fig. 11(a). It connects the specified end points and is described by a linear translation of the center of A and constant rotational rate. The resulting

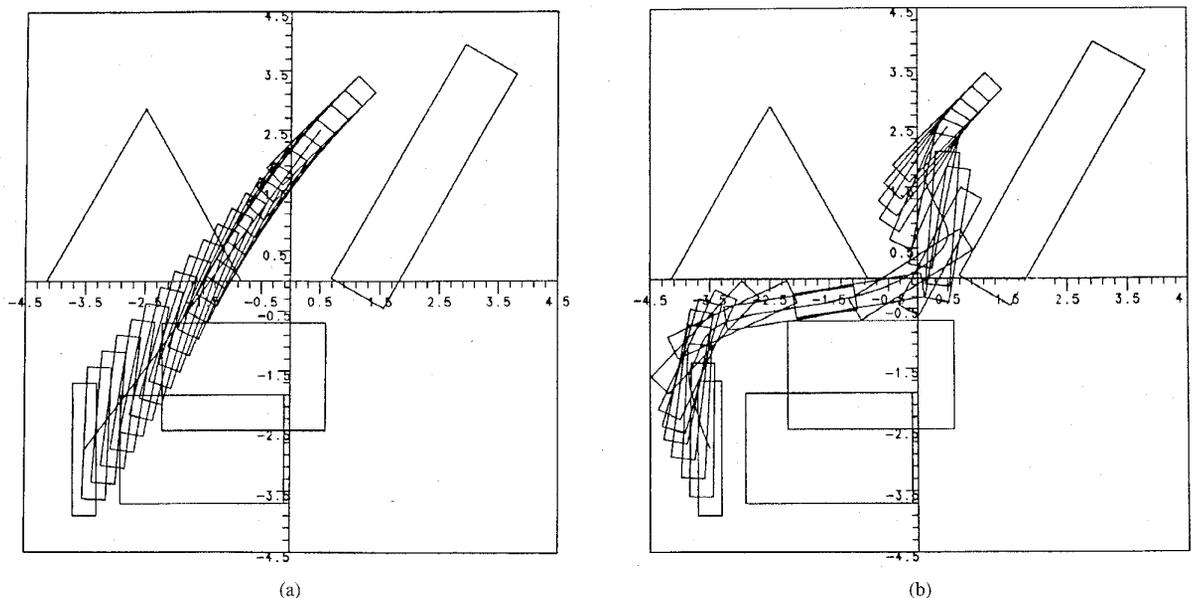


Fig. 11. Example of path finding: (a) initial path and (b) final path.

collision-free path is shown in Fig. 11(b). The function $J(\alpha)$ and its derivatives are evaluated only 40 times, an atypically small number compared with the dimension of the optimization problem, $N = 33$. The most likely explanation for this exceptional performance is the aforementioned nature of the minimum; it is not bowl-like, but instead there is an open set of global minimizers.

We conclude this section with a few general remarks on the basic approach. It extends readily to complex 3-D work spaces and more general kinematics, such as those found in typical industrial manipulators. An example in [29] treats a simple 3-D, 3-degree-of-freedom SCARA arm in a moderately complex work space. The computational time on the HP-Apollo was 29 seconds. Based on the comments in Section VI, one would expect this time to be decreased by a factor of at least ten on the Silicon Graphics machine. In more complex 3-D problems and where attention must be given to the issue of positive local minima [29], there may be sizable increases in the difficulty of the computations. Fortunately, there is a potential for considerable improvement in speed. As has been pointed out in [29], massive parallelization of the computations is achieved easily. The use of simpler parameterized path representations is also likely to prove advantageous. For example, the number of spline parameters in our example could be reduced appreciably, particularly if requirements on path smoothness were relaxed. Since the grid points on the path must be closely spaced, the adjacency ideas described at the end of Section VI should do much to reduce the time in computing the growth distance and its derivative—a major factor in the total time devoted to the evaluation of $J(\alpha)$ and its derivative.

Unlike almost all previously proposed path finding procedures [18], the one sketched here is not restricted to exploration of free space. The principle prior exception is the one proposed by Buckley [5]. Unfortunately, since his procedure depends on

d_P^2 , it leads to excessively large computational times when $n = 3$. As Buckley emphasizes, there are unappreciated advantages in the “flexible-trajectory paradigm” of adjusting collision paths. For instance, they have obvious value when a path with minor collisions is known *a priori*. Such situations occur commonly in applications: obstacles in a work space may be moved, causing collisions along a path which previously had no collisions; or paths generated by an interpolation of collision-free way points often exhibit minor collisions between the way points. Certainly, the computational advantages of penetration growth distances open new opportunities for researching the paradigm.

IX. CONCLUSION

New measures of object separation and penetration, d_S^G and d_P^G , have been introduced. These growth distances have many desirable properties and are closely connected (Theorem 4.1) to the familiar translational distances, d_S^2 and d_P^2 . Their principal advantage is computational simplicity and speed; for a wide variety of object models they can both be evaluated by solving a single mathematical programming problem of simple structure. The speed advantage is especially large for penetration distance. This, together the locally decreasing property and the means for evaluating motion derivatives, opens up new algorithmic approaches for solving difficult collision-elimination problems such as path finding in the presence of obstacles.

When the objects are polytopes, expressed either in terms of their faces or vertices, more can be said. Both the distances and their configuration-variable derivatives are determined by the solution of an LP problem. Asymptotic complexity is $O(m)$, which equals or far better the known times for translational distances, i.e., $O(m)$ for separation and $O(m^2 \log m)$ for penetration. Numerical experiments, based on standard LP techniques, confirm that computational time increases linearly

in m . Moreover, the times compare favorably with those required for translational distances as they are comparable or somewhat better for separation and far better for penetration.

APPENDIX A

GROWTH MODELS AND THE MINKOWSKI FUNCTION

Given an appropriate set, $X \subset R^n$, the corresponding Minkowski function is defined for $x \in R^n$ by

$$\mu_X(x) = \inf\{\lambda \in R^+ | \lambda^{-1}x \in X\}. \quad (35)$$

The following theorem summarizes well-known [19] or easily proved results needed in our developments.

Theorem A.1: Suppose $X \in O$ and $0 \in \text{int } X$. Then μ_X is a function from R^n into R^+ which has the following properties.

- i) It is convex.
- ii) For all $\alpha \in R^+$ and $x \in R^n$, $\mu_X(\alpha x) = \alpha\mu_X(x)$.
- iii) For all $x_1, x_2 \in R^n$, $\mu_X(x_1 + x_2) \leq \mu_X(x_1) + \mu_X(x_2)$.
- iv) For all $\alpha \in R^+$, $\alpha X = \{x | \mu_X(x) \leq \alpha\}$.
- v) $x \in \text{int } X$ if and only if $\mu_X(x) < 1$.
- vi) Let r_X and R_X be chosen so that $0 < r_X \leq R_X$ and $N_2(r_X) \subset X \subset N_2(R_X)$; then $R_X^{-1}\|x\|_2 \leq \mu_X(x) \leq r_X^{-1}\|x\|_2$ for all $x \in R^n$.

The application of the Minkowski function to growth models is evident. Since \bar{A} in (7) satisfies the assumptions of the Theorem, $\mu_{\bar{A}}$ is defined and result iv) yields

$$A(\sigma) = \{x | \mu_{\bar{A}}(x - p_A) \leq \sigma\}. \quad (36)$$

From this representation, and the corresponding one for $B(\sigma)$, it follows from (5) that

$$g(A, B) = \min\{\sigma | \mu_{\bar{A}}(x - p_A) \leq \sigma, \mu_{\bar{B}}(x - p_B) \leq \sigma, x \in R^n\}.$$

Thus

$$g(A, B) = \min F_{AB}(x), \quad x \in R^n \quad (37)$$

where

$$F_{AB}(x) = \max\{\mu_{\bar{A}}(x - p_A), \mu_{\bar{B}}(x - p_B)\}. \quad (38)$$

Since $\mu_{\bar{A}}$ and $\mu_{\bar{B}}$ are convex, F_{AB} is convex. Hence, any local minimizer of F_{AB} determines $g(A, B)$.

There is another characterization of $g(A, B)$. Note that $A(\sigma) \cap B(\sigma) \neq \emptyset$ is equivalent to $0 \in A(\sigma) - B(\sigma)$ which is equivalent to $p_B - p_A \in \sigma(\bar{A} - \bar{B})$. Since $0 \in \bar{A} - \bar{B}$ and $\bar{A} - \bar{B} \in O$, part iv) of Theorem A.1 implies

$$g(A, B) = \min\{\sigma | \mu_{\bar{A}-\bar{B}}(p_B - p_A) \leq \sigma\} = \mu_{\bar{A}-\bar{B}}(p_B - p_A). \quad (39)$$

While this characterization appears to be very simple, it is computationally complex because it is expensive to determine $\bar{A} - \bar{B}$ from \bar{A} and \bar{B} . The characterization is useful in theoretical developments. For example, it shows immediately that $g(A + \{s\}, B) = \mu_{\bar{A}-\bar{B}}(p_B - p_A - s)$ is a convex function of s .

It also provides a simple proof the geometric situation shown in Fig. 4(b). Since $\mu_{\bar{A}-\bar{B}}(p_B - p_A) = \sigma^*$, parts iv) and v) of Theorem A.1 imply $p_B - p_A \in \text{bd } \sigma^*(\bar{A} - \bar{B})$. By the convexity of $\sigma^*(\bar{A} - \bar{B})$ it follows [19] that $\sigma^*(\bar{A} - \bar{B})$ has a support hyperplane passing through $p_B - p_A$, i.e., there exists

$\eta^* \neq 0$ such that $\eta^{*T}(z + p_A - p_B) \leq 0$ for all $z \in \sigma^*(\bar{A} - \bar{B})$. From this it is easy to confirm that $\eta^{*T}(x_A - x_B) \leq 0$ for all $x_A \in A(\sigma^*), x_B \in B(\sigma^*)$. Let $x^* \in A(\sigma^*) \cap B(\sigma^*)$. Then, $\eta^{*T}(x_A - x^*) \leq 0$ for all $x_A \in A(\sigma^*)$. Hence, $A(\sigma^*) \subset H(x^*, \eta^*)$; similarly, $B(\sigma^*) \subset H(x^*, -\eta^*)$. This proves the existence of the separating hyperplane $P(x^*, \eta^*)$. Since $P(x^*, \eta^*)$ is a support hyperplane of both $A(\sigma^*)$ and $B(\sigma^*)$, it includes only boundary points of these sets [19]. Hence $A(\sigma^*)$ and $B(\sigma^*)$ have only boundary points in common, i.e., $A(\sigma^*) \cap B(\sigma^*) = \text{bd } A(\sigma^*) \cap \text{bd } B(\sigma^*)$.

APPENDIX B

PROOF OF THEOREM 3.1

Expression (39) proves part i). Part iii) follows from (39) and part vi) of Theorem A.1. Let $s = \alpha(p_A - p_B)$. Then $g(A + \{s\}, B)$ can be obtained by substituting $p_A + \alpha(p_A - p_B)$ for p_A in (39). Thus, by part ii) of Theorem A.1, $g(A + \{s\}, B) = \mu_{\bar{A}-\bar{B}}((1 + \alpha)(p_B - p_A)) = (1 + \alpha)g(A, B)$. This shows that $g(A + \{s\}, B) > g(A, B)$ for all $\alpha > 0$ which proves part iv).

Now consider part ii). Recall the definitions in the first paragraph of Section II. Because the three states of proximity are exhaustive and mutually exclusive, it is sufficient to prove the right-directed implications in part ii). Clearly, separation implies $A(1) \cap B(1) = \emptyset$ which in turn implies $\sigma^* > 1$. Penetration implies the existence of $x \in \text{int } A(1) \cap \text{int } B(1)$. Thus, by (36) and part v) of Theorem A.1, $\mu_{\bar{A}}(x - p_A) < 1$; similarly $\mu_{\bar{B}}(x - p_B) < 1$. From (37) it follows that $\sigma^* < 1$. Touching implies that $\sigma^* \leq 1$. In fact, $\sigma^* = 1$. Otherwise, there exists $x^* \in R^n$ such that $F_{AB}(x^*) < 1$, and by (38), $\mu_{\bar{A}}(x^* - p_A) < 1$ and $\mu_{\bar{B}}(x^* - p_B) < 1$. Part v) of Theorem A.1 then shows that $x^* \in \text{int } A(1) \cap \text{int } B(1)$, which contradicts the touching condition, $\text{int } A \cap \text{int } B = \emptyset$.

APPENDIX C

PROOF OF THEOREM 3.2

By (36), with A replaced by $T_A A + \{s_A\}$, and part iv) of Theorem A.1

$$\begin{aligned} & (T_A A + \{s_A\})(\sigma) \\ &= \{x | x = T_A p_A + s_A + T_A \bar{x}, \mu_{\bar{A}}(\bar{x}) \leq \sigma\} \\ &= \{x | \mu_{\bar{A}}(T_A^T(x - s_A) - p_A) \leq \sigma\}. \end{aligned} \quad (40)$$

Using this expression, the corresponding expression for $(T_B B + \{s_B\})(\sigma)$, and the obvious generalization of (37) shows that $G(P) = \min F(x, P), x \in R^n$, where $F(x, P) = \max\{\mu_{\bar{A}}(T_A^T(x - s_A) - p_A), \mu_{\bar{B}}(T_B^T(x - s_B) - p_B)\}$. Let \tilde{x}^* be a minimizer of $F(x, P)$. Clearly, $G(P) \leq F(\tilde{x}^*, P)$.

We now obtain a bound for $F(\tilde{x}^*, P)$. By parts iii) and vi) of Theorem A.1

$$\begin{aligned} & \mu_{\bar{A}}(T_A^T(\tilde{x}^* - s_A) - p_A) \\ &= \mu_{\bar{A}}(\tilde{T}_A^T(\tilde{x}^* - \tilde{s}_A) - p_A + T_A^T(\tilde{s}_A - s_A) \\ & \quad + (T_A^T - \tilde{T}_A^T)(\tilde{x} - \tilde{s}_A)) \\ &\leq G(\tilde{P}) + r_A^{-1}(\|T_A^T(\tilde{s}_A - s_A)\|_2 \\ & \quad + \|(\tilde{T}_A^T - T_A^T)(\tilde{x}^* - \tilde{s}_A)\|_2) \\ &\leq G(\tilde{P}) + r_A^{-1}(\|\tilde{s}_A - s_A\|_2 \\ & \quad + \|\tilde{T}_A^T - T_A^T\|_F \|\tilde{x}^* - \tilde{s}_A\|_2). \end{aligned} \quad (41)$$

Note that $\tilde{x}^* \in (\tilde{T}_A A + \{\tilde{s}_A\})(\tilde{\sigma}^*)$. This implies $\tilde{x}^* - \tilde{s}_A \in \tilde{T}_A(p_A + \tilde{\sigma}^* \tilde{A}) \subset N_2(\|p_A\|_2 + \tilde{\sigma}^* R_A)$ so that $\|\tilde{x}^* - \tilde{s}_A\|_2 \leq \|p_A\|_2 + \tilde{\sigma}^* R_A$. Using $T_B p_B + \tilde{s}_B \in \{\tilde{T}_A p_A + \tilde{s}_A\} + \sigma' N_2(r_A) \subset (\tilde{T}_A A + \{\tilde{s}_A\})(\sigma')$ as condition to determine σ' as an upper bound on $\tilde{\sigma}^*$ gives $\tilde{\sigma}^* \leq r_A^{-1}(\|p_A\|_2 + \|p_B\|_2 + \|\tilde{s}_A\|_2 + \|\tilde{s}_B\|_2) \leq r_A^{-1}(\|p_A\|_2 + \|p_B\|_2 + c_A + c_B)$. Choose k_A so that it is the larger of the two numbers, r_A^{-1} or $r_A^{-1}\|p_A\|_2 + r_A^{-2}R_A(\|p_A\|_2 + \|p_B\|_2 + c_A + c_B)$. Then

$$\begin{aligned} \mu_{\tilde{A}}(T_A^T(\tilde{x}^* - s_A) - p_A) \\ \leq G(\tilde{P}) + k_A(\|\tilde{s}_A - s_A\|_2 + \|\tilde{T}_A - T_A\|_F). \end{aligned} \quad (42)$$

Similarly

$$\begin{aligned} \mu_{\tilde{B}}(T_B^T(\tilde{x}^* - s_B) - p_B) \\ \leq G(\tilde{P}) + k_B(\|\tilde{s}_B - s_B\|_2 + \|\tilde{T}_B - T_B\|_F). \end{aligned} \quad (43)$$

From (11), (42), (43), the definition of F , and $k = \max\{k_A, k_B\}$, it follows that $G(P) \leq F(\tilde{x}^*, P) \leq G(\tilde{P}) + k\rho(P, \tilde{P})$.

Because $\rho(\tilde{P}, P) = \rho(P, \tilde{P})$, $G(\tilde{P}) \leq G(P) + k\rho(P, \tilde{P})$. Thus, the proof is complete.

APPENDIX D

PROOF OF THEOREM 4.1

We begin with a useful notation. For $x, y, \eta \in R^n, \eta \neq 0$ define

$$\delta(x, y, \eta) = \|\eta\|^{-1} \eta^T (y - x). \quad (44)$$

When $\delta > 0$, it is the separation distance between the half-spaces $H(x, \eta)$ and $H(y, -\eta)$, i.e., $\delta(x, y, \eta) = d_S^2(H(x, \eta), H(y, -\eta))$. When $\delta < 0$, the half-spaces overlap, i.e., $H(x, \eta) \cap H(y, -\eta)$ is a slab of thickness $-\delta(x, y, \eta)$. In what follows, notations and results in the last paragraph of Appendix A will be used without comment. Without loss of generality, $\|\eta^*\| = 1$.

Consider the separation distance inequalities in (16). They hold as equalities for $\sigma^* \leq 1$, so assume hereafter that $\sigma^* > 1$. Define

$$\begin{aligned} x_A &= p_A + (\sigma^*)^{-1}(x^* - p_A) \\ x_B &= p_B + (\sigma^*)^{-1}(x^* - p_B). \end{aligned} \quad (45)$$

Since $x^* \in \text{bd } A(\sigma^*)$, it follows from (6) that $x_A \in \text{bd } A$. Moreover, since $P(x^*, \eta^*)$ is a support hyperplane for $A(\sigma^*)$, $P(x_A, \eta^*)$ is a support hyperplane for A . From this and the corresponding results for x_B it follows that $A \subset H(x_A, \eta^*)$ and $B \subset H(x_B, -\eta^*)$. These inclusions imply

$$\begin{aligned} d_S^2(A, B) &\geq \delta(x_A, x_B, \eta^*) = \eta^{*T}(x_B - x_A) \\ &= (1 - (\sigma^*)^{-1})\eta^{*T}(p_B - p_A). \end{aligned} \quad (46)$$

By (6) and (15), $N_2(\sigma^* r_A) + \{p_A\} \subset A(\sigma^*) \subset H(x^*, \eta^*)$. Similarly, $N_2(\sigma^* r_B) + \{p_B\} \subset H(x^*, -\eta^*)$. Thus, by obvious geometry, $\eta^{*T}(p_B - p_A) \geq \sigma^*(r_A + r_B) = \alpha\sigma^*(R_A + R_B)$. Substituting this inequality into (46) proves the first inequality in (16).

Since $x_A \in A$ and $x_B \in B$, it follows from (3) and (45) that

$$\begin{aligned} d_S^2(A, B) &\leq \|x_A - x_B\|_2 \leq \|x^* - x_A\|_2 + \|x^* - x_B\|_2 \\ &= (1 - (\sigma^*)^{-1})(\|x^* - p_A\|_2 + \|x^* - p_B\|_2). \end{aligned} \quad (47)$$

By (6) and (15), $\|x^* - p_A\|_2 \leq \sigma^* R_A$. Similarly, $\|x^* - p_B\|_2 \leq \sigma^* R_B$. Substituting these inequalities into (47) proves the second inequality in (16).

Next, consider the third inequality in (16). Since it holds trivially for $\sigma^* \geq 1$, assume $\sigma^* < 1$. Clearly, (15) implies $d_P^2(A, B) \leq R_A + R_B$. For $\sigma^* = 0$, $d_P^2(A, B) = R_A + R_B$. Thus, the inequality holds for $\sigma^* = 0$. Therefore, it is sufficient to consider $0 < \sigma^* < 1$. Since $\eta^{*T}(p_B - p_A) \geq \sigma^*(r_A + r_B) > 0$, (46) shows that $\delta(x_A, x_B, \eta^*) < 0$. Because $H(x_A, \eta^*)$ and $H(x_B, -\eta^*)$ overlap by a distance $-\delta$, $H(x_A, \eta^*)$ and $H(x_B - \delta\eta^*, -\eta^*)$ have a common bounding plane, $P(x_A, \eta^*)$. Thus, $B + \{\delta\eta^*\} \subset H(x_B - \delta\eta^*, -\eta^*) = H(x_A, -\eta^*)$. This and $A \subset H(x_A, \eta^*)$ imply $d_P^2(A, B) \leq -\delta$. Hence

$$\begin{aligned} d_P^2(A, B) &\leq \eta^{*T}((x_A - x^*) - (x_B - x^*)) \\ &\leq ((\sigma^*)^{-1} - 1)(\eta^{*T}(x^* - p_A) - \eta^{*T}(x^* - p_B)). \end{aligned} \quad (48)$$

Because $\eta^{*T}x^* = \max\{\eta^{*T}x | x \in A(\sigma^*)\}$ it follows from (6) and (15) that $\sigma^* r_A < \eta^{*T}(x^* - p_A) \leq \sigma^* R_A$. Also, $\sigma^* r_B < -\eta^{*T}(x^* - p_B) \leq \sigma^* R_B$. Substitution in (48) completes the proof.

In proving (17) it is only necessary to consider the case where $d_P^2(A, B) > 0$. Thus, the characterization shown in Fig. 2(b) applies where $d_P^2(A, B) = \|t^*\| > 0$. Furthermore, $P(t^*, t^*)$ is a support hyperplane for $A - B$ and $A - B \subset H(t^*, t^*)$. It is easy to see that these facts imply the existence of $a^* \in A$ and $b^* \in B$ such that $t^* = a^* - b^*$, $A \subset H(a^*, t^*)$, $B \subset H(b^*, -t^*)$. The inclusions, together with (15), show that $N_2(r_A) + \{p_A\} \subset H(a^*, t^*)$ and $N_2(r_B) + \{p_B + t^*\} \subset B + \{t^*\} \subset H(b^* + t^*, -t^*) = H(a^*, -t^*)$. Hence, $\|t^*\|^{-1} t^{*T}(p_B + t^* - p_A) \geq r_A + r_B$ and

$$r_A + r_B - d_P^2(A, B) \leq \|t^*\|^{-1} t^{*T}(p_B - p_A). \quad (49)$$

From $A \subset H(a^*, t^*)$ it follows that $A(\sigma) \subset H(x_A(\sigma), t^*)$, where $x_A(\sigma) = p_A + \sigma(a^* - p_A)$; also, $B(\sigma) \subset H(x_B(\sigma), -t^*)$, where $x_B(\sigma) = p_B + \sigma(b^* - p_B)$. Clearly, $\delta(x_A(\sigma), x_B(\sigma), t^*) = 0$ implies $\sigma \leq \sigma^*$. By (44) this equality holds if

$$\begin{aligned} 0 &= \|t^*\|^{-1} t^{*T}(\sigma(b^* - a^*) + (1 - \sigma)(p_B - p_A)) \\ &= -\sigma d_P^2(A, B) + (1 - \sigma)\|t^*\|^{-1} t^{*T}(p_B - p_A). \end{aligned} \quad (50)$$

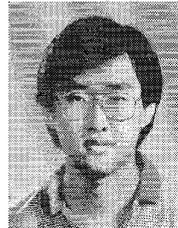
The assumption $d_P^2(A, B) \leq r_A + r_B$ and (49) guarantee that (50) has a unique root, $0 < \sigma < 1$. Solving (50) for $1 - \sigma$ and using (49) gives

$$\begin{aligned} 1 - \sigma^* &\geq 1 - \sigma = (d_P^2(A, B) \\ &\quad + \|t^*\|^{-1} t^{*T}(p_B - p_A))^{-1} d_P^2(A, B) \\ &\geq (r_A + r_B)^{-1} d_P^2(A, B). \end{aligned} \quad (51)$$

Multiplying by $R_A + R_B$ completes the proof. Note that (51) remains valid if $t^{*T}(p_B - p_A) > 0$. It is geometrically evident that this condition is usually satisfied, even when $r_A + r_B \leq d_P^2(A, B)$.

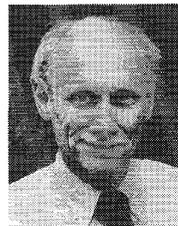
REFERENCES

- [1] M. J. Best and K. Ritter, *Linear Programming: Active Set Analysis and Computer Programs*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] J. E. Bobrow, "A direct minimization approach for obtaining the distance between convex polyhedra," *Int. J. Robot. Res.*, vol. 8, no. 3, pp. 65-76, 1989.
- [3] R. A. Brooks and T. Lozano-Perez, "A subdivision algorithm in configuration space for findpath with rotation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 224-233, 1985.
- [4] C. E. Buckley, "A proximity metric for continuum path planning," in *Proc. 9th Int. Conf. Artificial Intelligence*, 1985, pp. 1096-1102.
- [5] ———, "A foundation for the 'flexible-trajectory' approach to numeric path planning," *Int. J. Robot. Res.*, vol. 8, no. 3, pp. 44-64, 1989.
- [6] S. A. Cameron and R. K. Culley, "Determining the minimum translational distance between two convex polyhedra," in *Proc. IEEE Int. Conf. Robotics Automation*, San Francisco, CA, 1986, pp. 591-596.
- [7] F. H. Clarke, *Optimization and Nonsmooth Analysis*. New York: Wiley, 1983.
- [8] D. P. Dobkin, J. Hershberger, D. G. Kirkpatrick, and S. Suri, "Computing the intersection depth of polyhedra," *Algorithmica*, vol. 9, pp. 518-533, 1993.
- [9] D. P. Dobkin and D. G. Kirkpatrick, "A linear algorithm for determining the separation of convex polyhedra," *J. Algorithms*, vol. 6, pp. 381-392, 1985.
- [10] ———, "Determining the separation of preprocessed polyhedra—A unified approach," in *Proc. 17th ICALP, Lecture Notes Computer Sci.*, no. 443. Berlin: Springer-Verlag, 1990, pp. 400-412.
- [11] J. J. Dongarra, "Performance of various computers using standard linear equations software," *Comput. Sci. Dept., Univ. Tenn., Knoxville, TN, Rep. CS-89-85*, 1995.
- [12] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.
- [13] E. G. Gilbert and D. W. Johnson, "Distance functions and their application to robot path planning in the presence of obstacles," *IEEE J. Robot. Automat.*, vol. RA-1, pp. 21-30, 1985.
- [14] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, "A fast procedure for computing the distance between complex objects in three-dimensional space," *IEEE J. Robot. Automat.*, vol. 4, pp. 193-203, 1988.
- [15] E. G. Gilbert and C. P. Foo, "Computing the distance between general convex objects in three-dimensional space," *IEEE Trans. Robot. Automat.*, vol. 6, pp. 53-61, 1990.
- [16] E. G. Gilbert and C. J. Ong, "New distances for the separation and penetration of objects," in *Proc. IEEE Conf. Robotics Automation*, San Diego, CA, 1994, pp. 579-586.
- [17] S. S. Keerthi and K. Sridharan, "Measures of intensity of collision between convex objects and their efficient computation," in *Proc. Int. Conf. Intelligent Robotics*, Bangalore, India, 1991, pp. 266-275.
- [18] J. Latombe, *Robot Motion Planning*. Boston, MA: Kluwer, 1991.
- [19] S. R. Lay, *Convex Sets and Their Applications*. New York: Wiley, 1982.
- [20] D. Leven and M. Sharir, "Planning a purely translational motion for a convex object in two-dimensional space using generalized Voronoi diagrams," *Discrete Computational Geometry*, vol. 2, pp. 9-31, 1987.
- [21] M. C. Lin and J. F. Canny, "A fast algorithm for incremental distance calculation," in *Proc. IEEE Int. Conf. Robotics Automation*, Sacramento, CA, 1991, pp. 1008-1014.
- [22] M. C. Lin, D. Manocha, and J. F. Canny, "Fast contact determination in dynamic environments," in *Proc. IEEE Conf. Robotics Automation*, San Diego, CA, 1994, pp. 602-608.
- [23] C. Y. Liu and R. Mayne, "Distance calculations in motion planning problems," in *Proc. ASME 16th Design Auto. Conf.*, Chicago, IL, 1990, pp. 145-152.
- [24] N. Megiddo, "Linear programming in linear time when the dimension is fixed," *J. ACM*, vol. 31, pp. 114-127, 1984.
- [25] B. S. Mordukovich, *Approximation Methods in Problems of Optimization and Control*. Moscow: Nauka, 1988 (Engl. transl. in press, New York: Wiley-Interscience).
- [26] M. Nahon, "Determination of interference distance between two objects using optimization techniques," in *Proc. ASME Design Auto. Conf.*, Minneapolis, MN, 1994, pp. 1-6.
- [27] C. J. Ong, "Penetration distances and their applications to robot path planning," Ph.D. dissertation, Dept. Mechanical Eng. Appl. Mechanics, Univ. Michigan, Ann Arbor, MI, 1993.
- [28] ———, "Properties of penetration between general objects," in *Proc. IEEE Conf. Robotics Automation*, Nagoya, Japan, 1995, pp. 2293-2298.
- [29] C. J. Ong and E. G. Gilbert, "Robot path planning with penetration growth distance," in *Proc. IEEE Conf. Robotics Automation*, San Diego, CA, 1994, pp. 2146-2152.
- [30] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
- [31] N. K. Sancheti and S. S. Keerthi, "Computation of certain measures of proximity between convex objects," in *Proc. IEEE Conf. Robotics Automation*, Nice, France, 1992, pp. 2508-2513.
- [32] I. Sharf and M. Nahon, "Interference distance calculation for two objects bounded by quadratic surfaces," in *Proc. ASME Design Auto. Conf.*, 1995, pp. 634-641.
- [33] K. Sridharan, H. E. Stephanou, K. C. Craig, and S. S. Keerthi, "Distance measures on intersecting objects and their applications," *Inform. Proc. Lett.*, vol. 51, pp. 181-188, 1994.
- [34] S. Zeghloul, P. Rambeaud, and J. P. Lallemand, "A fast distance calculation between convex objects by optimization approach," in *Proc. IEEE Conf. Robotics Automation*, Nice, France, 1992, pp. 2520-2525.
- [35] K. Zikan and D. Curtis, "Intersections and separation of polytopes," Boeing Comput. Services, Rep. BCSTech-93-031, Oct. 1993.



Chong Jin Ong received the B.Eng. (Hon.) and M.Eng. degrees from the National University of Singapore in 1986 and 1988, respectively. He obtained the M.S.E. and Ph.D. degrees from the University of Michigan, Ann Arbor, in 1993.

He joined the National University of Singapore in 1993 and is a Lecturer with the Department of Mechanical and Production Engineering. His research interests are in distance computation of geometrical objects, path planning algorithms, control and optimization theories, and their applications.



Elmer G. Gilbert (S'51-A'52-M'57-SM'78-F'79) received the Ph.D. degree in instrumentation engineering from the University of Michigan, Ann Arbor.

Since 1957, he has been a Faculty Member in the Department of Aerospace Engineering at the University of Michigan, where he is now Professor Emeritus. He is also a member of the Department of Electrical Engineering and Computer Science. His current research interests include nonlinear control, optimization, and robotics. He has published numerous papers and holds eight patents.

Dr. Gilbert's special awards include: the O. H. Shuck Award for best paper at the 1978 Joint Automatic Control Conference, a Control Systems Society best paper award in 1979, and Research Excellence and Distinguished Faculty Achievement Awards from the University of Michigan. In 1994 he received the IEEE Control Systems Award. In 1996, he received the Richard E. Bellman Control Heritage Award of the American Automatic Control Council. He is a member of the Johns Hopkins University Society of Scholars, a fellow of the American Association for the Advancement of Science, and a member of the National Academy of Engineering.