Robustness of the Sutherland-Hodgman Algorithm

Conor McCoid[0000−0002−2592−5005]

1 Introduction

The intersection of polytopes, geometric objects in arbitrary dimension with flat sides, is of importance in such applications as computer graphics [2], mortar methods [8], and certain finite element methods [7]. Algorithms to compute those intersections must find the vertices of one polytope that lie in the other and intersections between the various faces. Because these intersections are calculated numerically, there is error associated with the result. This error, however small, can cause large intersections to be destroyed or created, even for consistent algorithms [5].

In computer graphics, one is concerned with a viewing window that clips a subject [2]. This gives rise to the terms subject and clipping. A subject polytope is a polytope that is clipped, i.e. parts of it are removed due to lying outside the viewing window. A clipping polytope represents the viewing window, the area that one wants all visual information to lie in. These terms will be used here to distinguish between the two polytopes, which the algorithm treats differently.

The Sutherland-Hodgman algorithm belongs to a family of intersection algorithms that use bisecting hyperplanes to represent the clipping polytope, as does the Sugihara algorithm [11]. All such algorithms may be written in the form of Algorithm 1. Most other intersection algorithms focus on the intersection of 2D shapes

Algorithm 1 Intersection algorithms that use bisecting hyperplanes

```
Inputs: set of vertices X, set of edges E, set of hyperplanes {P_i}
for i = 0 to N do
    X, E ← Bisect(X, E, P_i)
end for
```

Conor McCoid
Department of Mathematics and Statistics, McMaster University, Canada, e-mail: mccoidc@mcmaster.ca

[13, 2, 3, 9, 5], though they may be generalized to higher dimensions [2, 6, 4]. For a survey of intersection algorithms, see [10].

The original version of the Sutherland-Hodgman algorithm [12] considered only convex polygons. This was extended to higher dimensional polytopes by Broman and Shensa [1]. However, only the convexity of the clipping polytope is required for the algorithm to function.

The purpose of this note is two-fold: develop sufficient conditions for numerical robustness for this family of algorithms, represented by Algorithm 1, and; prove that the Sutherland-Hodgman algorithm can satisfy these conditions under careful implementations. These conditions can be used to formalize the robustness of other intersection algorithms. Proofs of the robustness of these algorithms that consider numerical error are rare, and this is a first for the Sutherland-Hodgman algorithm.

Section 2 describes necessary background information and develops the sufficient conditions for robustness. Sections 3 and 4 discuss the only possible numerical errors that can occur in the Sutherland-Hodgman algorithm and how to avoid violating the sufficient conditions for robustness with proper implementation. Section 5 then combines the arguments in the two preceding sections to prove the robustness of the Sutherland-Hodgman algorithm.

2 Background and framework

Consider two polytopes, a (possibly non-convex) subject polytope U and a convex clipping polytope V, both in $\mathbb{R}^n$. The subject polytope U is defined combinatorially as a collection of M vertices X and edges E, $U = (X, E)$, while the clipping polytope V is defined as the intersection of $N + 1$ half-spaces,

$$V = \bigcap_{i=0}^{N} \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{v}_i \geq b_i\},$$

for some set of normalized vectors $\mathbf{v}_i$ and scalars b_i which define hyperplanes P_i that divide the half-spaces.

For each of these hyperplanes, the set of vertices X of U is split into two: one set on the negative side of the hyperplane, X_-, and one on the non-negative side, X_+. The set X_- are to be discarded, as they are on the clipped side of the hyperplane, but they must first be used to compute intersections.

There are edges in E that have one vertex in X_- and one in X_+. Those edges intersect P_i, and those intersections must be found. The edges are clipped, and so the new edges to be kept are those between the vertices in X_+ and the new intersection points. Together, X_+ and these intersections form a new set of vertices. The new set of edges are the clipped edges, those of E with both vertices in X_+, and a new set of edges with both vertices intersection points. The algorithm then repeats with the next hyperplane, now acting on the new set of vertices and edges.

Remark 1 In what follows, Bisect(X, E, P) represents a subroutine that bisects the polytope with vertices X and edges E by hyperplane P, resulting in a new polytope.

Definition 1 (Property A) Bisect(X, E, P) has Property A if, for every input (X, E) representing the finite union of polytopes, the result of the subroutine is a finite union of polytopes.

Remark 2 For each vertex in $X \subset \mathbb{R}^n$, construct a hypersphere that encompasses all of its neighbours. Let r be the radius of the largest of these hyperspheres. The largest hypersphere has a cross sectional hyperarea

$$a = \frac{\pi^{(n-1)/2}}{\Gamma((n+1)/2)} r^{n-1}, \quad \text{where} \quad \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$$

Remark 3 δX represents a small perturbation on X, where each vertex is perturbed by at most $\|\delta X\|_\infty$ in an arbitrary direction. The change to the hypervolume of X caused by this perturbation is bounded by a collection of M cones with base a and height $\|\delta X\|_\infty$, one cone for each vertex in X, for a total maximum change in hypervolume of $aM \|\delta X\|_\infty / n$.

Definition 2 (Property B) Bisect(X, E, P) has Property B with constant α if the hypervolume of the set difference of Bisect($X + \delta X, E, P$) and Bisect(X, E, P) is at most $\alpha a M \|\delta X\|_\infty / n$ for some constant α.

Lemma 1 *Let (W, G) be the result of Algorithm 1 using subroutine Bisect(X, E, P) applied to a polytope (X, E) using hyperplanes $\{P_i\}_{i=0}^N$, and let $(\tilde{W}, \tilde{G})$ be the result of the same algorithm with the same collection of hyperplanes applied to $(X+\delta X, E)$. If Bisect(X, E, P) has Properties A and B with constant α, then (W, G) and $(\tilde{W}, \tilde{G})$ are the unions of finitely many polytopes and the hypervolume of the set difference of (W, G) and $(\tilde{W}, \tilde{G})$ is at most $\alpha^{N+1} a M \|\delta X\|_\infty / n$.*

Proof. Proceed by induction over the number of bisecting hyperplanes, $N + 1$. The base case, with one bisecting hyperplane, is true by Properties A and B of Bisect(X, E, P).

Suppose the statement is true for k bisecting hyperplanes. When applying Algorithm 1 with $k + 1$ bisecting hyperplanes to both (X, E) and $(X + \delta X, E)$, the first k steps are equivalent to applying Algorithm 1 with k bisecting hyperplanes, resulting in the unions of finitely many polytopes represented by (W_0, G_0) and $(\tilde{W}_0, \tilde{G}_0)$ that differ in hypervolume by $\alpha^k \|\delta X\|_\infty$. The final step in the algorithm is to run Bisect(W_0, G_0, P_k) and Bisect($\tilde{W}_0, \tilde{G}_0, P_k$).

By the induction hypothesis, (W_0, G_0) and $(\tilde{W}_0, \tilde{G}_0)$ differ by only a small hypervolume. Without loss of generality, suppose that $(\tilde{W}_0, \tilde{G}_0)$ can be written as

$$(\tilde{W}_0, \tilde{G}_0) = (W_0 + \delta W_0, G_0) \cup (\hat{W}_0, \hat{G}_0).$$

That is, the polytope represented by $(\tilde{W}_0, \tilde{G}_0)$ is the union of a small perturbation in the union of polytopes (W_0, G_0) and a small additional union of polytopes $(\hat{W}_0, \hat{G}_0)$

which all extend from facets on $(W_0 + \delta W_0, G_0)$. Generality is not lost as this represents a worst case scenario, with all of $\hat{W}_0$ added to $\tilde{W}_0$ rather than divided between W_0 and $\tilde{W}_0$. The union of polytopes $(\hat{W}_0, \hat{G}_0)$ can be written as perturbations on the facets with maximum height off each facet of $\left\|\hat{W}_0\right\|_\infty$. Then the hypervolume of $(\tilde{W}_0, \tilde{G}_0)$ satisfies

$$\frac{aM}{n}\left(\|\delta W_0\|_\infty + \left\|\hat{W}_0\right\|_\infty\right) \leq \alpha^k \frac{aM \|\delta X\|_\infty}{n}.$$

Bisect$(\tilde{W}_0, \tilde{G}_0, P_n)$ is then

$$\text{Bisect}(\tilde{W}_0, \tilde{G}_0, P_n) = \text{Bisect}(W_0 + \delta W_0, G_0, P_n) \cup \text{Bisect}(\hat{W}_0, \hat{G}_0, P_k).$$

By Property A, Bisect(W_0, G_0, P_k), Bisect$(\tilde{W}_0, \tilde{G}_0, P_k)$ and Bisect$(\hat{W}_0, \hat{G}_0, P_k)$ are the unions of finitely many polytopes. Any union or set difference of them is then also the union of finitely many polytopes.

By Property B with constant α, the set difference of Bisect$(W_0 + \delta W_0, G_0, P_k)$ and Bisect(W_0, G_0, P_k) has a hypervolume of at most $\alpha aM \|\delta W_0\|_\infty / n$. Likewise, Bisect$(\hat{W}_0, \hat{G}_0, P_k)$ has a hypervolume of at most $\alpha aM \left\|\hat{W}_0\right\|_\infty / n$. Thus, the set difference of (W, G) and $(\tilde{W}, \tilde{G})$ has a hypervolume of at most

$$\alpha \frac{aM}{n}\left(\|\delta W_0\|_\infty + \left\|\hat{W}_0\right\|_\infty\right) \leq \alpha^{k+1} \frac{aM \|\delta X\|_\infty}{n}.$$

□

Thus, Properties A and B are sufficient for robustness. The purpose of the remainder of this note is to show that good implementations of the Sutherland-Hodgman algorithm satisfy these properties.

Algorithm 2 gives a pseudocode representation of the bisection subroutine that forms the Sutherland-Hodgman algorithm. This algorithm replaces line 3 in Algorithm 1. Numerical error can occur at lines 2, where the vertices X are divided into X_+ and X_-, and 6, where intersection points are calculated. Combinatorial error can occur at line 12, where the new facet between intersection points is constructed.

Combinatorial error affects the consistency of the algorithm. Broman and Shensa [1] describe looping over the face lattice of the polytope to determine which k-faces are bisected and to construct new $(k-1)$-faces out of the $(k-2)$-faces found previously. As shown by the authors there, this preserves consistency of the Sutherland-Hodgman algorithm when it is generalized to higher dimensions.

To ensure robustness, numerical error must be limited at lines 2 and 6. It will be shown that the following implementation choices are sufficient to prove robustness. First, non-negativity should be used to determine inclusion to the half-spaces of V, instead of positivity. Second, coordinates should be transformed to map hyperplanes to cardinal axes.

Theorem 1 *If SH-Bisect(X, E, P) is implemented such that*

- *inclusion in a half space is determined using a binary-valued sign function;*

Algorithm 2 SH-Bisect(X, E, P), the Sutherland-Hodgman bisection subroutine

Inputs: subject polytope vertices X and edges E, clipping polytope hyperplane P
Test each vertex in X for which side of P it lies on, creating X_+ and X_-
Create an empty set of vertices Y and an empty set of edges F
for each edge in E **do**
  if the edge has one vertex in X_+ and one in X_- **then**
    Compute the intersection of the edge with P and add it to Y
    Add the edge between the vertex in X_+ and the intersection to F
  else if the edge has both vertices in X_+ **then**
    Add the edge to F
  end if
end for
Add appropriate edges between vertices of Y to F
$X \leftarrow X_+ \cup Y$, $E \leftarrow F$

- *coordinates are transformed to place bisecting hyperplanes on cardinal axes,*

then SH-Bisect(X, E, P) satisfies Properties A and B with constant $\alpha = 2$.

Because SH-Bisect(X, E, P) is consistent [12, 1], the only reason it can fail to have Properties A and B is due to numerical error. As has been discussed above, numerical error can only occur when testing a vertex of X for inclusion in a half space defined by P and computing the intersection of an edge of E with P. Both sources must be considered in turn, the former in Section 3 and the latter in Section 4. Once these sources have been considered, Theorem 1 can be proven in Section 5.

3 Testing a vertex

Consider a vertex $\mathbf{x} \in X$, a small shift to this vertex by δ, and a bisecting hyperplane P. Suppose that $\mathbf{x}$ is close enough to P such that $\mathbf{x} \in X_+$ but $\mathbf{x} + \delta \in X_-$. It must be shown that both $\mathbf{x}$ and $\mathbf{x} + \delta$ result in polytopes that differ in hypervolume by $O(\delta)$.

First note that by using a binary-valued sign function there are only two possibilities for each vertex in X: either the vertex is placed in X_+ or it is placed in X_-. There is no option for the vertex to lie directly on P, which can cause degenerate cases [5].

The vertex $\mathbf{x}$ has m edges extending from it connecting it to m other vertices, $m \geq n$. Suppose k of those vertices lie in X_+ and $m - k$ of them lie in X_-. Since $\mathbf{x} \in X_+$, $m - k$ intersections with P are calculated and $m - k$ new edges added to F, along with k original edges. This means $\mathbf{x}$ continues to have m edges extending from it. The shifted vertex $\mathbf{x} + \delta \in X_-$, meaning k intersections with P are calculated. These k intersections form part of a new facet embedded in P with edges constructed as in [1]. In both cases, an appropriate number of edges are found so that the results remain polytopes. Thus, Property A remains satisfied.

Note that this process does not necessarily preserve connectivity. The bisecting hyperplane may split a connected polytope into several polytopes. For example, a crescent moon may have its tips sliced off. Since the polytopes have a finite number of flat sides, the number of polytopes that result from a bisection is finite.

To show Property B, consider the hypersphere that contains all neighbours of $\mathbf{x}$. If the hyperplane P bisects this hypersphere, then construct two cones that share the intersection between P and the hypersphere as their bases, one with apex $\mathbf{x}$ and the other with apex $\mathbf{x}+\delta$. The bases of these cones have hyperarea at most a, see Remark 2. The difference in hypervolume of the two resulting polytopes is bounded by the sum of the hypervolume of the two cones. Since $\mathbf{x} + \delta$ differs from $\mathbf{x}$ by a distance δ, the sum of the heights of the cones is at most δ, giving a sum of volumes at most $2a\delta/n$.

If the hyperplane P does not bisect the hypersphere, then construct two cones tangent to the hypersphere, one with apex $\mathbf{x}$ and the other with apex $\mathbf{x} + \delta$. Again, the bases of these cones have hyperarea at most a. The difference in hypervolume of the two resulting polytopes is now bounded by the difference in hypervolume of the two cones, which is at most $a\delta/n$. A 3D example of this can be found in Figure 1.

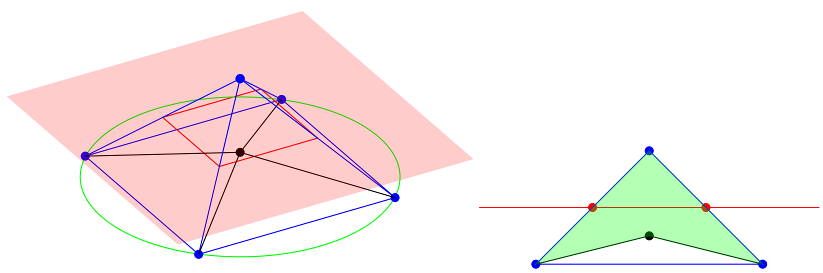

Fig. 1 A 3D example with a square pyramid, in blue. One vertex is perturbed over the bisecting plane, the perturbation in black. Intersections are given as red points. The base of both cones is shown in green. Also shown is the example in profile.

4 Computing an intersection

Consider now numerical error that may be caused by intersection calculations at line 6. Such calculations will not change the graph of the result, meaning they cannot alter Property A. However, they can alter the position of the nodes of the graph, and so it must be confirmed that Property B is maintained.

The following arguments rely on the use of coordinate transformations which place the bisecting hyperplane P onto a coordinate axis. This can be done using Householder reflections and translations, which are numerically stable and do not amplify error. Thus, if a perturbation of size δ is introduced to the original coordinates of X, then the transformed coordinates also have a perturbation of size δ. Let p be the coordinate perpendicular to P in the new transformed coordinates.

The following argument is adapted from [5, Lemma 2]. Consider an edge $(\mathbf{x}_1, \mathbf{x}_2) \in E$ which is bisected by P, which is now defined as the hyperplane where $p = 0$. Let p_1 and p_2 be the p-coordinate of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Without loss of generality, $p_1 \geq 0$ and $p_2 < 0$.

Let q be any other coordinate perpendicular to the p-coordinate, and q_1 and q_2 the q-coordinate of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Then q_0, the q-coordinate of the intersection of the edge with P, is equal to

$$q_0 = \frac{p_1 q_2 - p_2 q_1}{p_1 - p_2}.$$

If the coordinates are shifted by δ, then the change in q_0 is on the order of $r\delta/(p_1 - p_2)$, see Remark 2. This leads to a change in the hyperarea of the facet that lies in P of $a\delta/(p_1 - p_2)$. The change in hypervolume is then bounded by the cone with this hyperarea as a base and $\mathbf{x}_1$ as its apex:

$$\frac{a\delta}{p_1 - p_2} \frac{p_1}{n}.$$

There is no cancellation error in $p_1 - p_2$ since $p_2 < 0$, so $p_1 - p_2 > p_1$ and the change in hypervolume is $O\left(a\delta/n\right)$. Thus, Property B is maintained for $\alpha \geq 1$. See Figure 2 for an example in 2D.

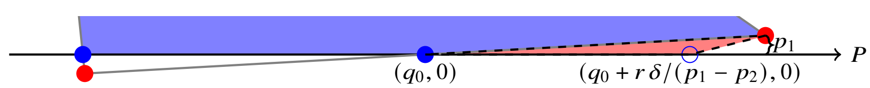

Fig. 2 Change in hypervolume caused by shifts in intersection calculations remains on the order of these shifts. Original hypervolume is shown in blue, with the change shown in red.

5 Proof of Theorem 1

Property A of SH-Bisect(X, E, P) is guaranteed due to the arguments in Section 3 and the construction of new facet edges described in [1].

For each vertex, the previous two sections have shown that the change in hypervolume caused by a shift to the vertex of size δ is bounded by $2a\delta/n$. As there are M vertices, this gives a total maximum change in hypervolume of $2aM\delta/n$. In the statement of Property B, δ is replaced by $\|\delta X\|_\infty$, and it is clear that $\alpha = 2$.

The Sutherland-Hodgman algorithm has been implemented as described by Theorem 1 in [6, Section 5.2], where it performs robustly.

6 Conclusion

The Sutherland-Hodgman algorithm implemented as described by Theorem 1 is a simple but effective algorithm for the intersection of polytopes. It produces shape-consistent results, i.e. having the appropriate number of vertices and edges. Large deletions or creations of hypervolume are not possible, though error can accumulate in high dimensions.

The algorithm is also highly asymmetric, as the order of hyperplanes of the clipping polygon will cause small numerical differences in results. This will not affect the robustness of the algorithm, as Theorem 1 still holds, so all possible results will have nearly the same hypervolume. However, there may be differences in number of vertices and edges and the shapes of facets, which should have little to no impact in practical applications. Symmetric algorithms may be able to take advantage of parallelization or other computational techniques to improve efficiency.

The bisection subroutine of the Sutherland-Hodgman algorithm, see Algorithm 2, cannot be said to preserve convexity. If convexity is important for a given application, one may consider modifications, such as by Sugihara [11]. The bisection subroutine there preserves the connectivity of the intersection, which is a necessary condition for convexity.

References

1. Broman, V., Shensa, M.: A compact algorithm for the intersection and approximation of N-dimensional polytopes. Math. Comput. Simul. **32**(5), 469–480 (1990)
2. Cyrus, M., Beck, J.: Generalized two-and three-dimensional clipping. Computers & Graphics **3**(1), 23–28 (1978)
3. Liang, Y.D., Barsky, B.A.: A new concept and method for line clipping. ACM Trans. Graph. **3**(1), 1–22 (1984)
4. McCoid, C.: Parsimonious simplicial intersection algorithm. ACM Trans. Math. Softw. **51**(4) (2025)
5. McCoid, C., Gander, M.J.: A provably robust algorithm for triangle-triangle intersections in floating-point arithmetic. ACM Trans. Math. Softw. **48**(2), 1–30 (2022)
6. McCoid, C., Gander, M.J.: Intersection of tetrahedra. ACM Trans. Math. Softw. **51**(2), 1–26 (2025)
7. Munjiza, A., Owen, D., Bicanic, N.: A combined finite-discrete element method in transient dynamics of fracturing solids. Eng. Comput. (1995)
8. Puso, M.A.: A 3D mortar method for solid mechanics. Int. J. Numer. Methods Eng. **59**(3), 315–336 (2004)
9. Skala, V.: An efficient algorithm for line clipping by convex polygon. Computers & Graphics **17**(4), 417–421 (1993)
10. Skala, V.: A brief survey of clipping and intersection algorithms with a list of references (including triangle-triangle intersections). Informatica **34**(1), 169–198 (2023)
11. Sugihara, K.: A robust and consistent algorithm for intersecting convex polyhedra. In: Computer Graphics Forum, vol. 13, pp. 45–54. Wiley Online Library (1994)
12. Sutherland, I.E., Hodgman, G.W.: Reentrant polygon clipping. Communications of the ACM **17**(1), 32–42 (1974)
13. Weiler, K., Atherton, P.: Hidden surface removal using polygon area sorting. In: Computer Graphics, vol. 11, pp. 214–222. ACM, New York (1977)