Discrete Comput Geom 35:223–240 (2006)
DOI: 10.1007/s00454-005-1206-y

Discrete & Computational Geometry
© 2005 Springer Science+Business Media, Inc.

Minkowski Sums of Monotone and General Simple Polygons*

Eduard Oks[1] and Micha Sharir[1,2]

[1]School of Computer Science, Tel Aviv University,
Tel Aviv 69978, Israel
BEO007@motorola.com

[2]Courant Institute of Mathematical Sciences, New York University,
New York, NY 10012, USA
michas@post.tau.ac.il

Abstract. Let P be a simple polygon with m edges, which is the disjoint union of k simple polygons, all monotone in a common direction u, and let Q be another simple polygon with n edges, which is the disjoint union of ℓ simple polygons, all monotone in a common direction v. We show that the combinatorial complexity of the Minkowski sum $P \oplus Q$ is $O(k\ell mn\alpha(\min\{m, n\}))$, where $\alpha(\cdot)$ is the inverse Ackermann function. Some structural properties of the case $k = \ell = 1$ have been (implicitly) studied in [17]. We rederive these properties using a different proof, apply them to obtain the above complexity bound for $k = \ell = 1$, obtain several additional properties of the sum for this special case, and then use them to derive the general bound.

1. Introduction

Let P and Q be two regions in the plane. Their *Minkowski sum* [2] is defined as

$$P \oplus Q = \{x + y \mid x \in P,\ y \in Q\}.$$

This is a fundamental construct that arises in many applications. One notable application involves placements and translational motion planning of an object in the presence of another object, which acts as a stationary obstacle. Assuming, without loss of generality, that

---

* Work on this paper has been supported by a grant from the U.S.–Israeli Binational Science Foundation, by NSF Grants CCR-97-32101 and CCR-00-98246, by a grant from the Israel Science Fund (for a Center of Excellence in Geometric Computing), and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University. This paper is part of the M.Sc. dissertation of the first author, prepared under the supervision of the second author.

the origin o lies in P, and denoting by $-P$ the reflection of P with respect to o, it follows by definition that $K = (-P) \oplus Q$ is the set of all vectors v such that translating P by v makes it intersect Q. Hence the complement K^c of K is a representation of the space of all free (translational) placements of P (namely, placements disjoint from Q). This observation makes Minkowski sums a central tool in the analysis of translational motion planning (see, e.g., [3], [6], [11], and [14]), and we also use this interpretation in our analysis.

Suppose that P and Q are polygonal regions with m and n edges, respectively. If P and Q are both convex, then $P \oplus Q$ is a convex polygon with at most $m + n$ edges (and can be computed in linear time) [12], [16]. If P is a simple polygon and Q is convex, the complexity of $P \oplus Q$ is $O(mn)$, and if both P and Q are simple, the complexity of $P \oplus Q$ is $O(m^2n^2)$; see [3]. All these bounds are tight in the worst case. The special case where P is a simple n-gon and Q is a line segment has been recently analyzed by Pustylnik and Sharir [13], where the precise worst-case upper bound $2n - 1$ on the complexity of $P \oplus Q$ has been established. A more pragmatic study of Minkowski sums has recently been undertaken by Agarwal et al. [1], [4], [5]. A simple algorithm for constructing the Minkowski sum is to decompose each of P, Q into simpler shapes, say, triangles, compute the Minkowski sums of all pairs of a subtriangle of P and a subtriangle of Q, and compute the union of the resulting sums, each being a convex polygon of at most six sides. Agarwal et al. have noted that triangulating P and Q typically results in a poor performance of the algorithm, and that coarser decompositions are more advantageous.

A simple polygon P is said to be *monotone* in direction u (also referred to as *u-monotone*) if every line orthogonal to u intersects P in a connected (possibly empty) interval. We can decompose any simple polygon P with m edges into simple subpolygons, all monotone in some specified direction u, by drawing a vertical segment through each vertex of P which is a locally u-extremal point of ∂P, and by extending that segment *inside* P till it hits ∂P again. These segments decompose P into $O(m)$ pairwise openly disjoint u-monotone simple polygons, and this bound is tight in the worst case.

Let P be a u-monotone simple polygon with m edges, and let Q be a v-monotone simple polygon with n edges, for two (possibly different) directions u, v. We show (Theorem 3.1) that the complexity of $P \oplus Q$ in this case is only $O(mn\alpha(\min\{m, n\}))$, which is tight in the worst case. (The upper bound was obtained by Hernańdez-Barrera [10] for the special case $u = v$. He also showed that the lower bound can be attained in this case.)

The proof relies on the following separation property, due to Toussaint and El-Gindy [17]: Given *disjoint* monotone polygons P and Q as above, we can translate P to infinity, without colliding with Q, in at least one of the four directions $u \pm \pi/2$, $v \pm \pi/2$. This property implies that $P \oplus Q$ is *simply connected*, from which the complexity bound follows using known bounds on the complexity of a single face in an arrangement of line segments; see, e.g., [15].

We provide, in Theorem 2.1, an alternative proof of the result of [17], and then use it to obtain the asserted complexity bound. Moreover, we derive several additional structural properties of the sum $P \oplus Q$ of two monotone simple polygons. For example, we show that its boundary is the concatenation of four connected portions, two of which are u-monotone and two v-monotone. We also show that the number of *pockets* along $\partial(P \oplus Q)$ is only $O(m + n)$. This notion is defined and analyzed in Section 5. This is roughly equivalent to asserting that the number of points on $\partial(P \oplus Q)$ that are locally x-extremal or y-extremal is $O(m+n)$.

We next use all these properties to prove the main result of the paper, which asserts that if P is a simple polygon with m edges which is the disjoint union of k simple u-monotone subpolygons, and Q is a simple polygon with n edges which is the disjoint union of ℓ simple v-monotone subpolygons, for any (possibly distinct) directions u, v, then the complexity of $P \oplus Q$ is $O(k\ell mn\alpha(\min\{m, n\}))$. This (almost) properly interpolates between the two extreme cases $k = \ell = 1$ (where the bound is worst-case tight), and $k = \Theta(m)$, $\ell = \Theta(n)$ (where we get an extra $\alpha(\cdot)$ factor).

2. Separating Two Monotone Chains

Theorem 2.1 (slightly reformulated) was already proven in [17]. We present here a different proof, using a functional representation of monotone polygonal paths.[1]

Theorem 2.1. *Let $f(x)$: $[a, b] \mapsto \mathbb{R}$, $g(y)$: $[c, d] \mapsto \mathbb{R}$ be (graphs of) continuous real functions defined on the above intervals of the x- and y-axes, respectively, that do not intersect each other. Then $f(x)$ can be translated to infinity along at least one of the four axis directions without colliding with $g(y)$.*

We say that a point p of the plane is *directly to the right* of another point q if the half-line starting at q and pointing to the positive x-direction passes through p. The notions of being directly to the left, directly above, and directly below are defined in an analogous manner.

Lemma 2.2. *Suppose that $g(y)$ has a point directly to the right of the right endpoint of $f(x)$. Then $g(y)$ has no point directly to the left of any point of $f(x)$.*

Proof. It is enough to show, by symmetry, that this holds for every $y \geq f(b)$. If for every $y \geq f(b)$ we have $g(y) > b$, we are done. Otherwise, set $x_0 := g(f(b))$, $y_0 := f(b)$, $x_1 := b$. If there exist $y \geq f(b)$ with $g(y) \leq b$, then denote by y_1 the infimum of all such y. Then, by continuity, $g(y_1) = x_1$, and the statement (that $g(y)$ has no point directly to the left of any point of $f(x)$) holds on the interval $[y_0, y_1]$. If $f(x)$ remains under y_1 in every point to the left of x_0, then the statement holds on the whole interval $[y_0, d]$. Otherwise, there is a largest x_2 where (proceeding from right to left) $f(x)$ first attains y_1. Similarly, now $g(y)$ either remains to the right of x_2 all the way to the end $(d, g(d))$, or there exists y_2 where $g(y)$ first reaches x_2. See Fig. 1.

This alternating construction terminates in finitely many steps, for otherwise we would obtain a bounded sequence (x_0, y_0), (x_1, y_0), (x_1, y_1), (x_2, y_1), (x_2, y_2), …, monotone in both coordinates, and its limit would be a common point of $f(x)$ and $g(y)$. It is easily seen that the termination of the process implies the statement of the lemma over the interval $[y_0, d]$, and a symmetric argument implies it for $[c, y_0]$. □

[1] We are grateful to János Pach for suggesting this proof, which has simplified our earlier analysis.

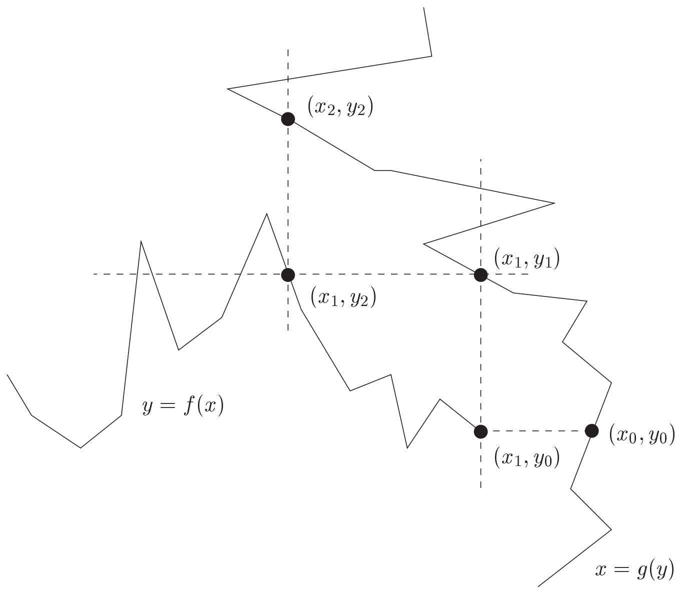

Fig. 1. The "staircase" of critical values $x_0, y_0, x_1, y_1, \ldots$.

Lemma 2.3. *If any point of $g(y)$ is directly to the right of any point of $f(x)$, then the largest x_0 such that $f(x_0)$ is directly to the left of a point of $g(y)$ has the property that $g(y)$ has no point directly to the left of any point of $f([a, x_0])$.*

Proof. Put $x_0 = \sup\{x \in [a, b] \mid f(x) \in [c, d] \text{ and } g(f(x)) > x\}$, and apply Lemma 2.2 to g and to f restricted to $[a, x_0]$. □

Proof of Theorem 2.1. We can assume that $g(y)$ has a point below $f(x)$ and a point above $f(x)$, otherwise we are done. By Lemma 2.3 (and by symmetry), there is a smallest y^- such that $(g(y^-), y^-)$ is above $f(x)$, and then all points of $g(y)$ above y^- are not below $f(x)$, and there is a largest y^+ such that $(g(y^+), y^+)$ is below $f(x)$, and then the points of $g(y)$ below y^+ are not above $f(x)$. Obviously, we have $y^- > y^+$.

By definition of y^-, y^+, the points of $g(y)$ on the interval (y^+, y^-) are neither above nor below $f(x)$, so these $g(y)$ values do not belong to $[a, b]$. By Bolzano's theorem, they must all be smaller than a or all bigger than b. Assume that all $g(y) > b$ on this interval. Then, by the continuity of $g(y)$, we have $g(y^-) = b = g(y^+)$. So $f(b) \in (y^+, y^-)$, hence $b < g(f(b))$. Thus, we can apply Lemma 2.2 to conclude that no point of $g(y)$ is directly to the left of any point of $f(x)$, so $f(x)$ can be translated to the left. □

Remarks. (1) Theorem 2.1 also holds when the graphs $f(x)$, $g(y)$ touch each other without crossing at any finite number of points. We omit details of this extension.

(2) Theorem 2.1 also holds when we replace $f(x)$, $g(y)$ by any pair of bounded connected arcs, each monotone in some direction, and these directions are not required to be orthgonal to each other (as is the case in the theorem). If u, v are the directions of monotonicity, the theorem asserts that one arc can be translated to infinity in one of the four directions $u \pm \pi/2$, $v \pm \pi/2$, without meeting the other arc. Indeed, apply a "shearing" affine transformation which maps the direction $u + \pi/2$ to the positive y-direction, and maps $v - \pi/2$ to the positive x-direction. This transforms the scenario into the one studied above, and an application of Theorem 2.1 in the new scenario, combined with the inverse shearing transformation, establishes the asserted property.

(3) Theorem 2.1 also holds when we replace $f(x)$, $g(y)$ by any pair of simple polygons, monotone in the x- and y-directions, or, as in (2), in any two directions. This extension follows easily from the preceding analysis, and is the one proved in [17] (using a different approach).

3. Minkowski Sum of Two Monotone Polygons

Theorem 3.1. *Let P and Q be two simple monotone polygons in two* (*possibly different*) *directions*, *having m and n edges*, *respectively. Then the complexity of the Minkowski sum $P \oplus Q$ is $O(mn\alpha(\min\{m, n\}))$.*

Proof. Suppose that P is monotone in direction u and that Q is monotone in direction v. Arguing as in Remark (2) of the preceding section, we may assume that u is the x-direction and that v is the y-direction. Let $\widetilde{P} = -P$ denote the reflection of P about the origin. Let t be a vector in the plane such that $t \notin P \oplus Q$. Then, by definition, $\widetilde{P}_t = \widetilde{P} + t$ is disjoint from Q. By Theorem 2.1 and Remark (3) following it, we can translate $\widetilde{P}_t$ in one of the four coordinate directions all the way to infinity, so that it does not intersect Q during the motion. This implies that there is a ray ρ in one of the four axis directions that emanates from t and is disjoint from $P \oplus Q$. This in turn implies that the complement of $P \oplus Q$ has no bounded components ("holes" of $P \oplus Q$), and thus $P \oplus Q$ is simply connected.

In other words, the boundary of $P \oplus Q$ is connected, and coincides with the boundary of the unbounded face of its complement. Let Σ denote the set of all line segments of the form $e + v$, where e is an edge of P and v is a vertex of Q, or e is an edge of Q and v is a vertex of P. Σ consists of $2mn$ segments, and any point on $\partial(P \oplus Q)$ must be contained in one of these segments. As is well known (see, e.g., [15]), the complexity of any single face in an arrangement of $2mn$ segments is $O(mn\alpha(mn))$. To obtain the slightly improved asserted bound, assume, without loss of generality, that $m \le n$. Note that Σ can be represented as the union of $2m$ subsets, each consisting of the sums of all edges of Q with a fixed vertex of P, or of the sums of all vertices of Q with a fixed edge of P. Each subset consists of pairwise (openly) disjoint segments, so the complexity of the subarrangement that they form is $O(n)$. We then apply the *Combination Lemma* of Har-Peled [8], which implies that the complexity of a single face in the overlay of $2m$ arrangements, each of complexity $O(n)$, is $O(mn\alpha(m))$. See also [9] for an alternative proof. □

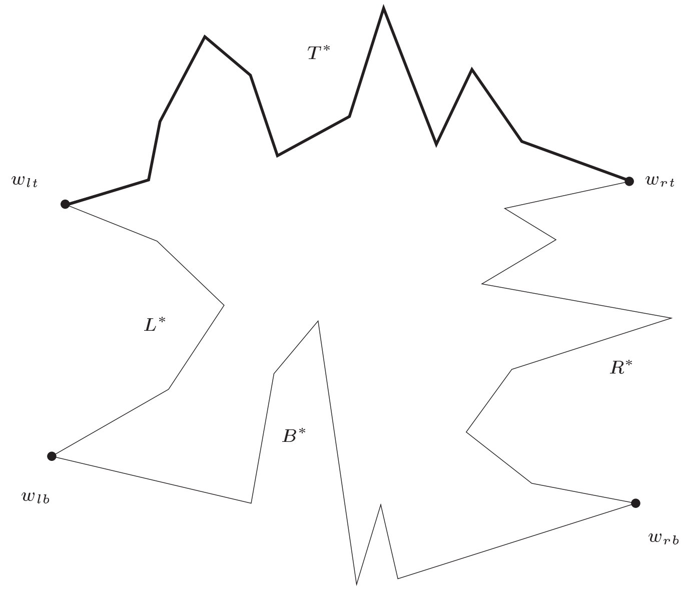

Fig. 2. The Minkowski sum of two monotone polygons, and (one possible) partition of its boundary into x-monotone portions T^*, B^*, and y-monotone portions L^*, R^*, delimited by the points w_{tr}, w_{tl}, w_{bl}, w_{br}. The portion T^* is highlighted.

4. The Boundary of the Sum of Two Monotone Polygons

In what follows we assume that P and Q are monotone in the x- and y-directions, respectively. As noted above, this involves no loss of generality.

Theorem 4.1. *Let P and Q be two simple polygons monotone in the x- and y-directions, respectively. Then the boundary of $S = P \oplus Q$ is the concatenation of two x-monotone and two y-monotone connected polygonal chains, which are pairwise openly disjoint.*

See Fig. 2.

In the proof we use the interpretation, already mentioned above, of $S = P \oplus Q$ as the space of all "forbidden" translations of $\widetilde{P} = (-P)$ at which it intersects Q, which we regard as stationary. The boundary of S is the set of all translations where $\widetilde{P}$ touches Q, but does not intersect its interior.

By Theorem 2.1, each point $v \in \partial S$ can be classified into one (or more) of the four following types:

Top, if $\widetilde{P}$ can be moved from v to infinity in the positive y-direction without penetrating into Q.

Bottom, if $\widetilde{P}$ can be moved from v to infinity in the negative y-direction without penetrating into Q.

Left, if $\widetilde{P}$ can be moved from v to infinity in the negative x-direction without penetrating into Q.

Right, if $\widetilde{P}$ can be moved from v to infinity in the positive x-direction without penetrating into Q.

We can therefore write ∂S as the union of four subsets T, B, L, R, where T (resp., B, L, R) consists of all top (resp., bottom, left, right) points on ∂S. By definition, all of these sets are closed. These sets are not necessarily disjoint, but the only points of ∂S that belong to $T \cap B$ are the leftmost and rightmost points of S. Similarly, only the topmost and bottommost points of S can belong to $L \cap R$. Any other pair of sets can have a more substantial intersection.

Proof of Theorem 4.1. Let w_t, w_b, w_l, w_r denote respectively the highest, lowest, leftmost, and rightmost points of ∂S. (We assume general position which makes these points unique). These four points partition ∂S into four connected portions, which we denote as the northeastern portion *NE* (lying clockwise from w_t to w_r), the southeastern portion *SE* (lying clockwise from w_r to w_b), the southwestern portion *SW* (lying clockwise from w_b to w_l), and the northwestern portion *NW* (lying clockwise from w_l to w_t). Note that the points w_t, w_b, w_l, w_r need not be distinct, although we always have $w_t \neq w_b$ and $w_l \neq w_r$. See Fig. 3.

It is easily seen that

$$NE \subseteq T \cup R, \qquad SE \subseteq B \cup R, \qquad SW \subseteq B \cup L, \qquad NW \subseteq T \cup L.$$

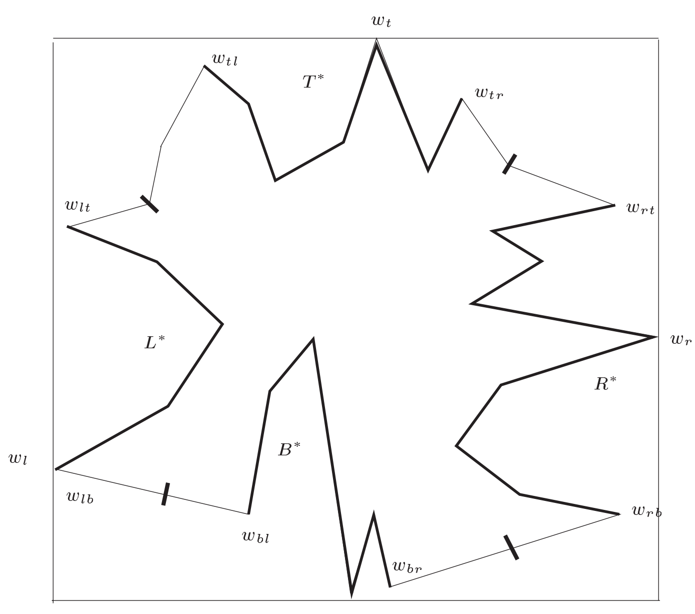

Fig. 3. The partition of ∂S into the portions *NE*, *SE*, *SW*, and *NW*.

More precisely, except possibly for their endpoints, these chains satisfy

$$NE \cap (B \cup L) = \emptyset, \qquad SE \cap (T \cup L) = \emptyset, \qquad SW \cap (T \cup R) = \emptyset, \qquad NW \cap (B \cup R) = \emptyset.$$

Lemma 4.2. *Let $u, v \in NE$ such that v lies clockwise to u. It is impossible that $u \in R \setminus T$ and $v \in T \setminus R$. Symmetric statements hold for SE, SW, and NW.*

Proof. Suppose to the contrary that such a pair of points u, v exists. Clearly, we have $u \in R$ and $v \in T$. Hence the rightward-directed ray ρ_u emanating from u and the upward-directed ray ρ_v emanating from v are both openly disjoint from S. It is easily seen that these rays must cross each other. Indeed, it is impossible to draw the simple clockwise-directed connected polygonal chain *NE*, so that it starts at w_t, ends at w_r, lies below w_t and to the left of w_r, and passes first through u and then through v, so that the rays ρ_u and ρ_v are openly disjoint from *NE* and from each other. This is because such a drawing would yield a plane embedding of $K_{3,3}$, as is illustrated in Fig. 4.

Hence the two rays intersect, at some point z, as is illustrated in Fig. 5. Let $\widetilde{P}_u = \widetilde{P} + u$, $\widetilde{P}_v = \widetilde{P} + v$, denote the placements of $\widetilde{P}$ with its reference point placed at u, v, respectively. Since $u, v \in \partial S$, $\widetilde{P}_u$ and $\widetilde{P}_v$ touch Q, but do not penetrate into it. Move $\widetilde{P}_u$ to the right until its reference point reaches z, and then move it down until the reference point reaches v. That is, the reference point traces the chain $J := \overline{uz} \cup \overline{zv}$, and the area swept by $\widetilde{P}$ during this motion is $P' := \widetilde{P} \oplus J$. By construction, P' and Q are openly disjoint. See Fig. 6. By construction, J is both (weakly) x- and y-monotone. Thus P' is also x-monotone, since it is the Minkowski sum of two x-monotone polygons (see, e.g., [10]). Since P' and Q are openly disjoint, it follows from Theorem 2.1 (and the subsequent Remark (3)) that we can move P' to infinity along one of the four coordinate directions, without penetrating into Q. However, P' contains both $\widetilde{P}_u$ and $\widetilde{P}_v$, and thus

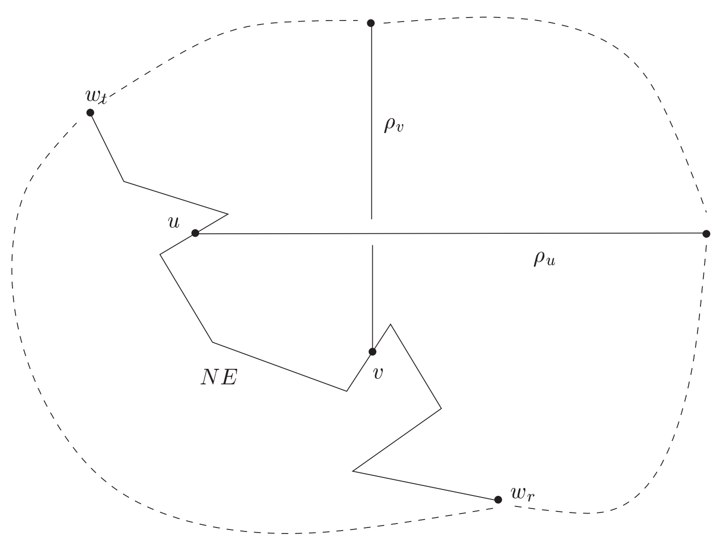

Fig. 4. If ρ_u and ρ_v do not intersect, we obtain an impossible plane embedding of $K_{3,3}$.

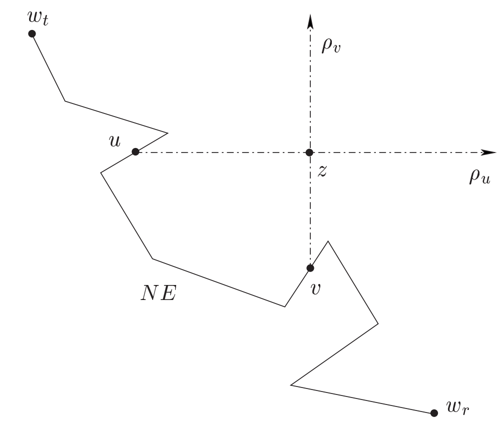

Fig. 5. The configuration in Lemma 4.2.

both these polygons can be moved to infinity *in the same direction*, or, in other words, both u and v belong to the same subset of ∂S, which contradicts the facts that $u, v \notin B \cup L$, $u \notin T$ and $v \notin R$. The corresponding statements for *SE*, *SW*, and *NW* are proved in a fully symmetric manner. □

Lemma 4.2 implies that we can partition *NE* into two openly disjoint connected subchains, T_{NE} and R_{NE}, with a common endpoint w, such that T_{NE} connects w_t to w and is contained in T, and R_{NE} connects w to w_r and is contained in R. Symmetrically, we obtain similar partitions $SE := R_{SE} \cup B_{SE}$, $SW := B_{SW} \cup L_{SW}$, and $NW := L_{NW} \cup T_{NW}$. The point w need not be unique. For example, if *NE* is monotone in both the x- and y-

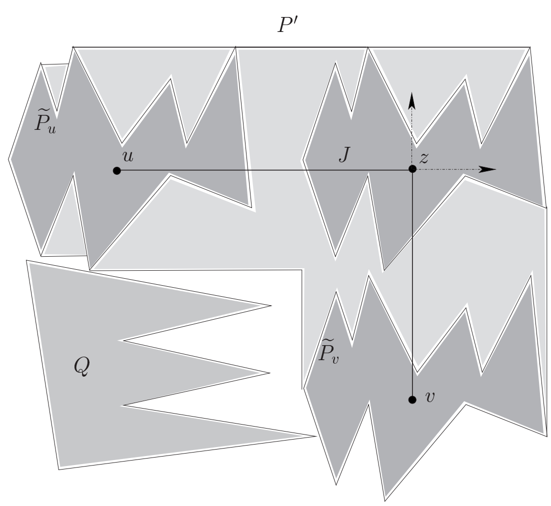

Fig. 6. The swept polygon P', obtained as the area swept by $\widetilde{P}$ as it translates from u to v via z, is disjoint from Q.

directions, *any* point along it can serve as the delimiter w. See Fig. 3 where the delimiters w are highlighted; for *NE*, any point between w_{tr} and w_{rt} can serve as a delimiter, and similarly for the other three chains. We refer to the loci of the delimiters w as the *buffer zones* of *NE*, *SE*, *SW*, and *NW*. Set $T^* := T_{NW} \cup T_{NE} \subseteq T$, $R^* := R_{NE} \cup R_{SE} \subseteq R$, $B^* := B_{SE} \cup B_{SW} \subseteq B$, and $L^* := L_{SW} \cup L_{NW} \subseteq L$. Each of these four sets is connected, and they constitute the desired partition of ∂S, as asserted in Theorem 4.1. □

5. Pockets in the Minkowski Sum of Monotone Polygons

We next bound the number of *pockets* in S. A *top pocket* is a maximal connected portion γ of T^* which is the concatenation of two connected portions α and β, such that (when proceeding in clockwise direction) α is monotone decreasing in y and monotone increasing in x, and β is monotone increasing in both x and y. Consequently, the common endpoint of α and β is locally y-minimal in T^*, and the two other endpoints of α, β are locally y-maximal. *Bottom pockets* in B^*, *left pockets* in L^*, and *right pockets* in R^* are defined in an analogous manner. See Fig. 7. Note that the pockets are pairwise openly disjoint, and that their union is ∂S minus the four buffer zones of *NE*, *SE*, *SW*, and *NW*.

Theorem 5.1. *Let P be an x-monotone simple polygon with m edges, and let Q be a y-monotone simple polygon with n edges. Then the number of pockets in $P \oplus Q$ is $O(m + n)$.*

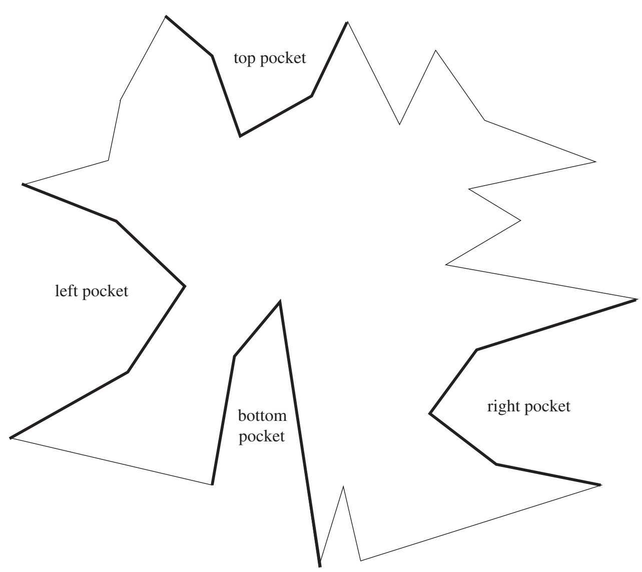

Fig. 7. Pockets of ∂S.

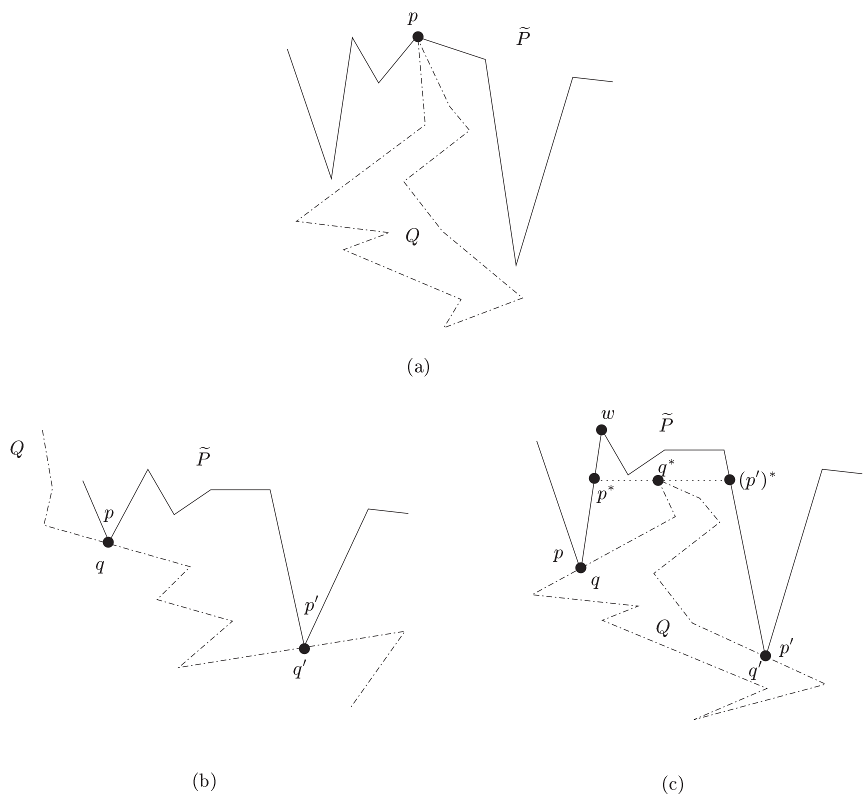

Fig. 8. Three kinds of placements that correspond to the lowest point of a top pocket. (a) A single contact. (b) An impossible double contact. (c) A double contact, and the (pocket) vertex w of $\widetilde{P}$ being charged.

Proof. Let $\gamma = \alpha \| \beta$ be a top pocket with lowest vertex v, incident to both α and β. In the interpretation of placing $\widetilde{P}$ around Q, v is a translation of $\widetilde{P}$ into a free placement at which one of the following situations arises:

Single contact: A vertex p of $\widetilde{P}$ that is locally y-maximal on (the bottom portion of) $\partial \widetilde{P}$ touches the unique y-maximal vertex of Q. See Fig. 8(a).

(To see that this is the only possible situation, note first that the contact must be a vertex–vertex contact, or else it would not represent a local minimum of a top pocket. A similar reasoning shows that one of the vertices must be convex and the other reflex, and that if the reflex vertex is of $\widetilde{P}$ (resp., of Q) then it must be locally y-maximal (resp., y-minimal) on the boundary of the respective polygon. Since Q, being a y-monotone polygon, does not have a reflex vertex that is locally y-minimal on ∂Q, the latter case is impossible, and the asserted characterization follows.)

Double contact: Two points p, p' of the bottom portion of $\partial \widetilde{P}$ touch two corresponding points q, q' of ∂Q that are locally top boundary points. Moreover, if we move $\widetilde{P}$ slightly to the left then it penetrates into Q in the vicinity of one of these contacts, and if we move $\widetilde{P}$ slightly to the right then it penetrates into Q in the vicinity of the other contact. (It is easily checked that these penetrations cannot both occur in the vicinity of the same contact, because this would contradict the monotonicity of either P or Q.) See Fig. 8(b),(c).

Clearly, we may assume that one of p, q is a *vertex* of its polygon, and the same holds for p', q'. However, we make no assumption on which ones are the vertices, and the arguments carry through in all cases. (The only exceptions are Figs. 8 and 9, where we have chosen, for convenience, to make p, p' vertices.)

A top pocket whose lowest point is generated by a single contact can be uniquely charged to the corresponding vertex p of $\widetilde{P}$, for a total of $O(m)$ such pockets. The same bound holds for bottom pockets of this kind, and the number of left and right pockets of this kind is $O(n)$. (The constants of proportionality in these bounds are smaller than 1.)

Consider next a top pocket whose lowest point v is generated by a double contact of two points p, p' on the bottom boundary of $\widetilde{P}$ with two corresponding points q, q' of ∂Q that are locally top boundary points. Assume, without loss of generality, that p lies to the left of p'. Since v is the lowest point of a pocket, we cannot move $\widetilde{P}$ to the left or to the right without immediately penetrating into Q. As noted above, one of the following two cases must arise:

Case A. As we move $\widetilde{P}$ slightly to the left, p penetrates into Q, and as we move it slightly to the right, p' penetrates into Q. See Fig. 8(b).

Case B. As we move $\widetilde{P}$ slightly to the left, p' penetrates into Q, and as we move it slightly to the right, p penetrates into Q. See Fig. 8(c).

We first claim that case A is impossible. Indeed, consider the portion of the bottom boundary of $\widetilde{P}$ at the placement v as the graph of a continuous function $y = f(x)$. Let $\bar{q}$ (resp., $\bar{q}'$) be a point in the interior of Q which coincides with p (resp., with p') as we move $\widetilde{P}$ slightly to the left (resp., to the right). Connect $\bar{q}$ and $\bar{q}'$ by a y-monotone polygonal chain within Q (which is always possible, since Q is y-monotone), which we regard as the graph of a continuous function $x = g(y)$. By construction, $g(y)$ has a point directly to the right of p' (namely, $\bar{q}'$), and a point directly to the left of p (namely, $\bar{q}$), which contradicts Lemma 2.2, thus showing that case A is impossible.

In case B the local penetrations of $\widetilde{P}$ into Q in the vicinities of the two pairs of coincident points $p = q$ and $p' = q'$ (at the placement v of $\widetilde{P}$) imply that the following property holds: Separate $\widetilde{P}$ and Q locally near $p = q$ by a line ℓ, and locally near $p' = q'$ by a line ℓ'. Then ℓ has a positive slope, $\widetilde{P}$ lies locally to its left and above it, and Q lies locally to its right and below it. Symmetrically, ℓ' has a negative slope, $\widetilde{P}$ lies locally to its right and above it, and Q lies locally to its left and below it. This implies, arguing as in Section 2, that the top vertex q^* of Q lies above p and p' and below the portion δ of the graph $y = f(x)$ that connects p and p' (as defined in case A). Let p^* (resp., $(p')^*$) denote the closest point on δ that lies directly to the left (resp., to the right) of q^*; the preceding analysis implies that both points exist. It follows that the global maximum of $y = f(x)$ between p^* and $(p')^*$ must occur at an interior point w, which is the highest point of a bottom pocket of $\widetilde{P}$ (alternatively, the lowest point of a top pocket of P). See Fig. 8(c).

We charge the top pocket of v to w, and claim that *this charging is unique*. Indeed, suppose to the contrary that another top pocket is also charged to w. Let v_1 denote its

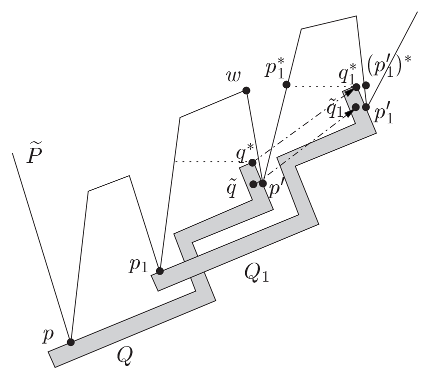

Fig. 9. Illustrating the proof that two top pockets of S of case B cannot both charge the same pocket of P.

lowest point, and let the corresponding double contact be determined by points p_1, p_1' of $\widetilde{P}$ with corresponding points q_1, q_1' of Q.

It is more convenient for this stage of the analysis to regard $\widetilde{P}$ as the stationary and Q as the translating polygon. We thus have two placements of Q, which for simplicity we denote as Q and Q_1. The stationary bottom boundary of $\widetilde{P}$ contains the five points p, p_1, w, p', p_1', so that p and p_1 lie to the left of w, p' and p_1' lie to the right of w, and w lies above the four other points. The polygon Q touches $\widetilde{P}$ at the two points p, p', which coincide with the respective points $q, q' \in Q$, and the polygon Q_1 touches $\widetilde{P}$ at the two points p_1, p_1', which coincide with the respective points $q_1, q_1' \in Q_1$. Let t be the translation vector that satisfies $Q_1 = Q + t$. Since the situation is symmetric in Q and Q_1 we may assume, without loss of generality, that t has a positive x-component, and consider two cases, depending on the sign of the y-component of t. (The general position assumption allows us to assume that t is not horizontal.)

Suppose first that t has a positive y-component. Refer to Fig. 9. Since case B applies to both Q and Q_1, there exists a point $\tilde{q} \in Q$ directly to the left and arbitrarily close to p'. Hence, $\tilde{q}_1 = \tilde{q} + t \in Q_1$ lies to the right and above p'. This implies that the highest point q_1^* of Q_1 lies above and to the right of p'. Indeed, it clearly lies above p'. Suppose to the contrary that it lies to the left of p'. By the preceding arguments, q_1^* lies below the graph of $y = f(x)$ and has a point on that graph directly to its right. By continuity, the closest such point must lie to the left of p'. This, combined with the fact that $\tilde{q}_1 = \tilde{q} + t \in Q_1$ lies to the right and above p', is easily seen to contradict Lemma 2.2, thus showing that q_1^* lies to the right of p'.

It follows that the two closest points p_1^*, $(p_1')^*$ on $y = f(x)$ that lie directly to the left and to the right of q_1^* both lie to the right of p'. However, then the pocket associated with Q_1 should have charged a vertex of $\widetilde{P}$ that lies between p_1^* and $(p_1')^*$, and thus lies to the right of p', so it could not have charged w.

The case where t has a negative y-component is argued in a fully symmetric fashion. Using the point p_1 instead of p', we show that the highest point q^* of Q is "trapped" in a pocket of $\widetilde{P}$ that lies fully to the right of p_1, and thus the pocket associated with the placement Q cannot charge w. This shows that w can be charged at most once, as asserted.

To sum up, we have shown that the top pockets of S that are generated by double contacts can be uniquely charged to top pockets of P. Symmetrically, bottom pockets of S of this kind can be uniquely charged to bottom pockets of P. Yet another symmetric argument, in which the roles of P and Q are interchanged, shows that left (resp., right) double-contact pockets of S can be uniquely charged to left (resp., right) pockets of Q. Adding the bound on the number of single-contact pockets, we conclude that the total number of pockets of S is $O(m+n)$. □

6. Minkowski Sum of Nonmonotone Simple Polygons

The preceding machinery allows us to derive the main result of this paper:

Theorem 6.1. *Let P be a simple polygon with m edges, which can be decomposed into k simple subpolygons, all monotone in the x-direction, and let Q be a simple polygon with n edges, which can be decomposed into ℓ simple subpolygons, all monotone in the y-direction. Then the complexity of $P \oplus Q$ is $O(k\ell mn\alpha(\min\{m,n\}))$. The same holds if the x- and y-directions are replaced by two arbitrary directions.*

Proof. Let $P_1, \ldots, P_k$ be the k subpolygons in the decomposition of P, and let $Q_1, \ldots, Q_\ell$ be the ℓ subpolygons in the decomposition of Q. Let m_i denote the number of edges of P_i, for $i = 1, \ldots, k$, and let n_i denote the number of edges of Q_i, for $i = 1, \ldots, \ell$. We have $\sum_{i=1}^{k} m_i = O(m)$ and $\sum_{i=1}^{\ell} n_i = O(n)$.

Put $S_{ij} := P_i \oplus Q_j$, for $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$. Clearly, $S = P \oplus Q = \bigcup_{i,j} S_{ij}$. By Theorem 3.1, the complexity of S_{ij} is $O(m_i n_j \alpha(\min\{m_i, n_j\}))$. Hence, the total number of edges of all the sums S_{ij} is

$$O\left(\left(\sum_{i=1}^{k}\sum_{j=1}^{\ell} m_i n_j\right)\alpha(\min\{m,n\})\right) = O(mn\alpha(\min\{m,n\})).$$

For each i and j, let T_{ij}^*, B_{ij}^*, L_{ij}^*, and R_{ij}^* denote the four connected portions of ∂S_{ij}, as provided by Theorem 4.1. Let X denote the collection of all the chains T_{ij}^* and B_{ij}^*, and let Y denote the collection of all the chains L_{ij}^* and R_{ij}^*. X is a set of $2k\ell$ x-monotone polygonal chains. The number of intersections of any pair of such chains is proportional to the number of their edges, which is easily seen to imply that the complexity of the arrangement $\mathcal{A}(X)$ is $O(k\ell mn\alpha(\min\{m,n\}))$. Similarly, the complexity of $\mathcal{A}(Y)$ is also $O(k\ell mn\alpha(\min\{m,n\}))$.

The complement of S is the union of some faces of the arrangement $\mathcal{A}(X \cup Y)$. Let H denote the collection of these faces. H contains one (the unique) unbounded face, and the rest are bounded faces ("holes" of S). By the Combination Lemma for planar arrangements (see [15]), the overall complexity of all the faces of H (that is, the complexity of S) is proportional to the complexity of $\mathcal{A}(X)$ plus the complexity of $\mathcal{A}(Y)$ plus $|H|$. Hence, Theorem 6.1 is an immediate consequence of the following lemma.

Lemma 6.2. *The number of holes of $P \oplus Q$ is $O(k\ell mn\alpha(\min\{m,n\}))$.*

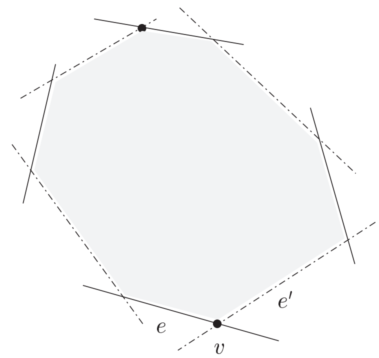

Fig. 10. A convex hole in H.

Proof. Let f be a bounded hole in H. If ∂f contains a vertex of either $\mathcal{A}(X)$ or $\mathcal{A}(Y)$, we charge f to that vertex, and thus conclude that the number of such holes is $O(k\ell mn\alpha(\min\{m, n\}))$. Otherwise, f is a convex polygon, whose boundary consists of a sequence of edges, alternating between edges of $\mathcal{A}(X)$ and edges of $\mathcal{A}(Y)$. Clearly, f has an even number of edges.

Let v be the lowest vertex of f, and suppose that it is incident to an edge e of $\mathcal{A}(X)$ and to an edge e' of $\mathcal{A}(Y)$. Suppose, without loss of generality, that e bounds f to the left of v and that e' bounds f to the right of v; see Fig. 10. In this case e is (a portion of) an edge of some T^*_{ij} and e' is (a portion of) an edge of some $L^*_{i'j'}$. Clearly, f is a portion of a face f_0 of the arrangement $\mathcal{A}(T^*_{ij} \cup L^*_{i'j'})$, and v is a local y-minimum of f_0. (Note that the case $i = i'$, $j = j'$ is impossible, because T^*_{ij} cannot meet L^*_{ij} in such a way.)

A simple application of Morse theory to f_0 shows that if f_0 is not y-monotone, then the number of local y-minima of f_0 is proportional to the number of points of ∂f_0 which are local y-extrema of the complement of f_0 (i.e., reflex locally y-extremal vertices of ∂f_0). See, e.g., Lemma 2.4 of [7] for a similar argument. Any such point u is a local y-extremal vertex of either T^*_{ij} or $L^*_{i'j'}$. The latter chain has only two such vertices, and the number of such vertices on the former chain T^*_{ij} is 1 plus the number of top pockets of S_{ij}. Hence, this number is $O(m_i)$, by Theorem 5.1. We repeat this argument to all the faces of $\mathcal{A}(T^*_{ij} \cup L^*_{i'j'})$ which are not y-monotone, to all other combinations of sub-boundaries of S_{ij} and $S_{i'j'}$, and to all combinations of i, j, i', j', to conclude that the overall number of holes f that satisfy all the above conditions is

$$O\left(\sum_{i,j,i',j'} (m_i + m_{i'})\right) = O(mk\ell^2) = O(k\ell mn).$$

Suppose then that f_0 is y-monotone. Then f_0 has a unique y-minimal point (namely, v). If ∂f_0 contains a vertex of either S_{ij} or of $S_{i'j'}$ then we charge v (uniquely) to such a vertex. Summing over all such faces f_0 and over all i, j, i', j', we conclude that the number of holes f that fall into this subcase is

$$O\left(\sum_{i,j,i',j'} (m_i n_j + m_{i'} n_{j'})\alpha(\min\{m, n\})\right) = O(k\ell mn\alpha(\min\{m, n\})).$$

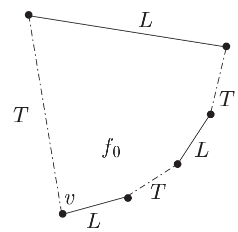

Fig. 11. A convex hole f_0 of $T^*_{ij} \cup L^*_{i'j'}$ with more than four edges, and the pair (e_1, e_2) of edges to which f_0 is charged.

We may thus assume that f_0 is convex and bounded, and that the edges of its boundary alternate between (portions of) edges of T^*_{ij} and (portions of) edges of $L^*_{i'j'}$. No edge of the top boundary of f_0 can belong to T^*_{ij}, and thus the top boundary of f_0 consists of a *single edge* of $L^*_{i'j'}$. Similarly, the left boundary of f_0 consists of a single edge of T^*_{ij}. We distinguish between two cases:

Case A: f_0 *has more than four edges*. See Fig. 11. Note that the unique left edge e_1 of f_0 spans the entire y-range of the hole. We can therefore connect e_1 to the highest right edge e_2 of f_0 (excluding the top edge of f_0) by a horizontal segment g_0, as shown in Fig. 11. Note that e_2 is also (a portion of) an edge of T^*_{ij}. Let $\bar{e}_1$ and $\bar{e}_2$ denote the edges of T^*_{ij} that contain, respectively, e_1 and e_2. We draw in the plane a graph G whose vertices are the edges of T^*_{ij}, which are drawn as they are. The edges of G are all the pairs $(\bar{e}_1, \bar{e}_2)$ obtained from the faces f_0 that fall into this subcase, and are drawn in the plane as the above horizontal connecting segments g_0. Clearly, G is planar.

We claim that G is simple. Indeed, suppose to the contrary that there exist two faces f_0, f_1 that cause the same pair of edges $\bar{e}_1, \bar{e}_2$ of T^*_{ij} to be connected by two respective horizontal segments g_0, g_1. Suppose, without loss of generailty, that g_0 lies higher than g_1, and refer to Fig. 12. Note that f_0 has at least one additional right edge e_3 that is contained in an edge $\bar{e}_3$ of T^*_{ij}. The x-monotonicity of T^*_{ij} implies that the entire edge

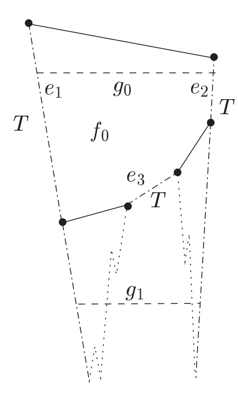

Fig. 12. Illustrating the proof that G is simple.

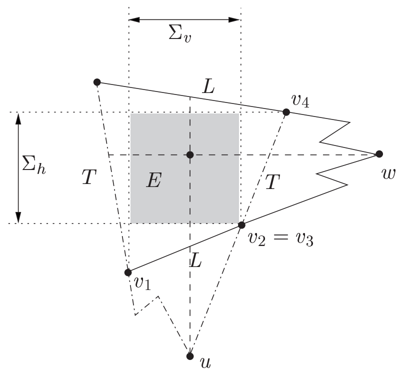

Fig. 13. A convex quadrangular hole f_0 of $T^*_{ij} \cup L^*_{i'j'}$, and the pair of pockets to which f_0 is charged.

$\bar{e}_3$ must be contained in the upper wedge defined by the lines containing $\bar{e}_1$ and $\bar{e}_2$. Therefore, $\bar{e}_3$ appears along T^*_{ij} between $\bar{e}_1$ and $\bar{e}_2$. Moreover, $\bar{e}_3$ lies above g_1, and the right endpoint of $\bar{e}_1$ lies below g_1. It follows that the portion of T^*_{ij} between $\bar{e}_1$ and $\bar{e}_3$ must cross g_1, which is impossible. This contradiction shows that G is simple. Hence the number of edges of G, and thus the number of faces f_0 of the type under consideration, is proportional to the number of edges of T^*_{ij}. Summing this bound over all i, j, i', j', we conclude that the overall number of holes f that give rise to faces f_0 of case A is $O(k\ell mn\alpha(\min\{m, n\}))$.

Case B: *f_0 is a quadrilateral*. See Fig. 13. Let v_1, v_2, v_3, v_4 denote the vertices of f_0 in counterclockwise order, starting from the bottom vertex $v = v_1$. Let Σ_{v} denote the vertical strip between v_1 and v_2, and let Σ_{h} denote the horizontal strip between v_2 and v_3. It follows that the rectangle $E = \Sigma_{\mathrm{v}} \cap \Sigma_{\mathrm{h}}$ is fully contained in f_0. See Fig. 13. Since the left edge of f_0 has negative slope, and the right edge (which belongs to T^*_{ij}) has positive slope, it follows that the portion of T^*_{ij} between these edges must contain at least one pocket, and that its lowest vertex u lies in Σ_{v}. Similarly, the portion of $L^*_{i'j'}$ between the bottom and the top edges of f_0 must contain at least one pocket, and its rightmost vertex w lies in Σ_{h}. Hence, the vertical line through u and the horizontal line through w meet inside E, and thus inside f_0. We can therefore charge f_0 to the pair of pockets of T^*_{ij} and $L^*_{i'j'}$ that are associated with u and w, respectively, and conclude that any such pair of pockets is charged at most once. Hence, by Theorem 5.1, the number of faces f_0 of case B is $O(m_i n_{j'})$. Summing this bound over all i, j, i', j', we conclude that the overall number of holes f that give rise to faces f_0 of case B is $O(k\ell mn)$.

Thus the number of holes of $P \oplus Q$ is $O(k\ell mn\alpha(\min\{m, n\}))$. This completes the proof of Lemma 6.2, and thus also the proof of Theorem 6.1. □

Remarks. (1) As already remarked, the bound in Theorem 6.1 is slightly suboptimal when $k = \Theta(m)$ and $\ell = \Theta(n)$, which is the case of arbitrary simple polygons P and Q. In this case the worst-case tight bound is $O(m^2n^2) = O(k\ell mn)$ [3]. It would be interesting to fine-tune our analysis to make our bound equal to this bound in the general case.

(2) The subpolygons $P_1, \ldots, P_k$ in the decomposition of P need not be pairwise openly disjoint, and the theorem continues to hold provided that the overall number of their edges is still $O(m)$. A similar extension applies to the decomposition of Q.

References

1. P.K. Agarwal, E. Flato and D. Halperin, Polygon decomposition for efficient construction of Minkowski sums, *Comput. Geom. Theory Appl.* **21** (2002), 39–61.
2. R.V. Benson, *Euclidean Geometry and Convexity*, McGraw-Hill, New York, 1966.
3. M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf, *Computational Geometry, Algorithms and Applications*, 2nd edition, Springer-Verlag, Berlin, 2000.
4. E. Flato, Robust and efficient construction of planar Minkowski sums, M.Sc. Dissertation, School of Computer Science, Tel Aviv University, 2000.
5. E. Flato and D. Halperin, Robust and efficient construction of planar Minkowski sums, in *Abstracts* 16*th European Workshop on Computational Geometry*, Eilat, 2000, pages 85–88.
6. L. Guibas, L. Ramshaw and J. Stolfi, A kinetic framework for computational geometry, *Proc.* 24*th Annu. IEEE Symp. on Foundations of Computer Science*, 1983, pages 100–111.
7. D. Halperin and M. Sharir, Almost tight upper bounds for the single cell and zone problems in three dimensions, *Discrete Comput. Geom.* **14** (1995), 385–410.
8. S. Har-Peled, Multicolor combination lemmas, *Comput. Geom. Theory Appl.* **12** (1999), 155–176.
9. S. Har-Peled, T.M. Chan, B. Aronov, D. Halperin and J. Snoeyink, The complexity of a single face of a Minkowski sum, *Proc.* 7*th Canadian Conf. on Computational Geometry*, 1995, pages 91–96.
10. A. Hernańdez-Barrera, Computing the Minkowski sum of monotone polygons, *IEICE Trans. Inform. & Syst.*, **E80-D**(2) (1997), 218–222.
11. K. Kedem, R. Livne, J. Pach and M. Sharir, On the union of Jordan regions and collision-free translational motion amidst polygonal obstacles, *Discrete Comput. Geom.* **1** (1986), 59–71.
12. T. Lozano-Perez and M.A. Wesley, An algorithm for planning collision-free paths among polyhedral obstacles, *Comm. ACM* **22** (1979), 560–570.
13. G. Pustylnik and M. Sharir, The Minkowski sum of a simple polygon and a segment, *Inform. Process. Lett.* **85** (2003), 179–184.
14. M. Sharir, Efficient algorithms for planning purely translational collision-free motion in two and three dimensions, *Proc. IEEE Symp. on Robotics and Automation*, 1987, pages 1326–1331.
15. M. Sharir and P.K. Agarwal, *Davenport–Schinzel Sequences and Their Geometric Applications*, Cambridge University Press, Cambridge, 1995.
16. G.T. Toussaint, Solving geometric problems with the rotating calipers, *Proc. IEEE MELECON '*83, Athens, 1983, pages A.10.02/1–4.
17. G.T. Toussaint and H.A. ElGindy, Separation of two monotone polygons in linear time, *Robotica* **2** (1984), 215–220.

Received May 31, 2004, *and in revised form August* 1, 2005. *Online publication December* 22, 2005.