Under consideration for publication in Math. Struct. in Comp. Science

A formal proof of Pick's theorem

John Harrison

Intel Corporation, JF1-13

2111 NE 25th Avenue, Hillsboro OR 97124, USA

johnh@ichips.intel.com

Received 23 November 2015

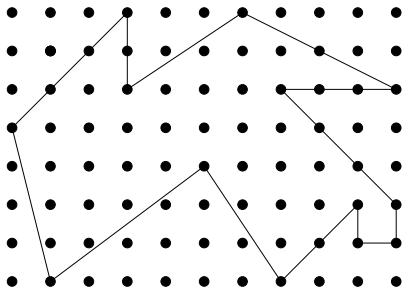
Pick's theorem relates the area of a simple polygon with vertices at integer lattice points to the number of lattice points in its inside and boundary. We describe a formal proof of this theorem using the HOL Light theorem prover. As sometimes happens for highly geometrical proofs, the formalization turned out to be more work than initially expected. The difficulties arise mostly from formalizing the triangulation process for an arbitrary polygon.

1. Introduction

We start with some definitions, whose formal counterparts will be given later. Throughout this paper we call a point $(x, y) \in \mathbb{R}^2$ of the plane an *integer lattice point*, or simply *lattice point*, iff both x and y are integers. We define a *lattice polygon* to be a polygon all of whose vertices are at integer lattice points. Pick's theorem (Pick, 1899) states that given a simple lattice polygon, if I is the number of integer lattice points in its inside and B the number of integer lattice points on its boundary, then the area of the polygon is given by

$$A = I + B/2 - 1$$

For example, the following diagram shows a lattice polygon, with the lattice points in its vicinity marked with black spots. There are 30 lattice points in the inside of the polygon and 22 on its boundary, so according to Pick's theorem the area of the polygon is $30 + 22/2 - 1 = 40$.



To state Pick's theorem more precisely we need to be a bit more careful about what we consider

a polygon and what constitutes its 'inside' and 'boundary'. We will pay more attention to these details later. For now, we just recall that a *curve* or *path* in the plane can formally be considered simply as a continuous function $\gamma : [0,1] \to \mathbb{R}^2$ out of the unit interval, and a *polygonal path* is the special case of a piecewise linear continuous function. A path is said to be *closed* if its starting and finishing points are the same ($\gamma(0) = \gamma(1)$), and *simple* if it does not otherwise self-intersect (i.e. $\gamma(s) = \gamma(t)$ implies either $s = t$ or $\{s,t\} = \{0,1\}$). We can identify a *simple polygon* with the image $\gamma[0,1]$ of the unit interval under a simple closed polygonal path γ.

This paper describes a formalization of Pick's theorem using the HOL Light theorem prover (Harrison, 1996). The proof turned out to be somewhat harder work than expected. Using the crude metric of 'lines of proof script', it seems to be almost as challenging as the Prime Number Theorem and considerably harder than Dirichlet's theorem on primes in arithmetic progression.

Theorem	Lines of proof
Dirichlet's theorem (Harrison, 2009a)	2082 lines
Pick's theorem (the present paper)	**3709** lines
Prime Number Theorem (Harrison, 2009b)	4314 lines

Charming and surprising as Pick's theorem is, one would normally consider it a very straightforward result in comparison with those two jewels of 19th century mathematics. What accounts for the difficulty of formalizing Pick? Of course, it could be just a reflection of our own ineptitude. However we prefer to believe that there is an intrinsic difficulty in formalizing many geometric proofs in stating and proving in a formal way various properties that seem intuitively obvious 'by eye'. In fact, the main difficulties with formalizing Pick's theorem are not at all specific to that theorem, but are connected with the process of 'triangulating' a general polygon, as happens in the informal proof we were using as a model.

2. Preliminary definitions and lemmas

We first describe the formalization of the various concepts involved in stating Pick's theorem. The material connected with paths is fairly standard, but the precise definition of 'inside' is perhaps more interesting.

Lattice points

We define the lattice points as 'integral vectors', i.e. vectors all of whose components are integers. The definition is made for the more general Euclidean space $\mathbb{R}^N$ using the technical setup described in (Harrison, 2005), but we will only use the special case $\mathbb{R}^2$ in what follows.

```
|- integral_vector x ⇔
        ∀i. 1 <= i ∧ i <= dimindex(:N) ⇒ integer(x$i)
```

Paths and polygons

As noted, we consider a path to be simply a continuous function out of the unit interval. Actually, the HOL Light formalization uses the unit interval in the type $\mathbb{R}^1$, which is technically different, though isomorphic to, the unit interval on $\mathbb{R}$. The elements `vec 0` and `vec 1` are the endpoints of this interval.

```
|- path g ⇔ g continuous_on interval[vec 0,vec 1]
```

For most purposes, we want to forget the details of the parametrization using the arbitrary interval $[0, 1]$, so we define various natural abbreviations:

```
|- pathstart g = g(vec 0)

|- pathfinish g = g(vec 1)

|- closed_path g ⇔ pathstart g = pathfinish g

|- path_image g = IMAGE g (interval[vec 0,vec 1])
```

One important special case of a path is the straight-line path from point a to point b. The functions `drop` : $\mathbb{R}^1 \to \mathbb{R}$ and `lift` : $\mathbb{R} \to \mathbb{R}^1$ are the inverse bijections between the type $\mathbb{R}^1$ and $\mathbb{R}$.

```
|- linepath(a,b) = λx. (&1 - drop x) % a + drop x % b
```

It's often convenient to stick together two new paths to make a new path. To retain the canonical parametrization we allocate the two intervals $[0, 1/2]$ and $[1/2, 1]$ to scaled versions of the the two components; formally:

```
|- g1 ++ g2 = (λx. if drop x <= &1 / &2
                   then g1(&2 % x)
                   else g2(&2 % x - vec 1))
```

We define an *arc* to be a path that does not self-intersect at all:

```
|- arc g ⇔
       path g ∧
       ∀x y. x IN interval[vec 0,vec 1] ∧
             y IN interval[vec 0,vec 1] ∧
             g x = g y
             ⇒ x = y
```

and a *simple path* to be one that may intersect only at endpoints

```
|- simple_path g ⇔
       path g ∧
       ∀x y. x IN interval[vec 0,vec 1] ∧
             y IN interval[vec 0,vec 1] ∧
             g x = g y
             ⇒ x = y ∨ x = vec 0 ∧ y = vec 1 ∨ x = vec 1 ∧ y = vec 0
```

Note that every arc is therefore also a simple path, and various other straightforward relationships hold:

```
|- arc g ⇒ simple_path g

|- simple_path g ⇒ path g

|- arc g ⇔ simple_path g ∧ ¬(pathfinish g = pathstart g)
```

We consider a polygonal path to be one defined by a list of vertices $[v_0; v_1; \cdots; v_n]$. Intuitively, this is the path that proceeds through those vertices in order via straight-line segments. It is defined by list recursion as a succession of `linepath`s.

```
|- polygonal_path [] = linepath(vec 0,vec 0) ∧
   polygonal_path [a] = linepath(a,a) ∧
   polygonal_path [a;b] = linepath(a,b) ∧
   polygonal_path (CONS a (CONS b (CONS c l))) =
       linepath(a,b) ++ polygonal_path(CONS b (CONS c l))
```

We will seldom be concerned with the cases of empty or singleton lists, but they are defined for the sake of regularity. In the case of the empty list, the choice of the corresponding linepath as the constant function whose image is $\{0\}$ is made only to avoid technical restrictions on the theorem about the image of a polygonal path under a linear mapping, and has no deep significance.

Inside and outside

Pick's theorem talks about the 'inside' or 'interior' of a polygon. First of all, let us note that we already have an established topological theory in HOL Light, which includes the interior, closure and frontier (boundary) of a set.

```
|- interior s = {x | ∃t. open t ∧ x IN t ∧ t SUBSET s}

|- closure s = s UNION {x | x limit_point_of s}

|- frontier s = (closure s) DIFF (interior s)
```

Despite the disparate definitions, the interior and closure obey well-known dualities:

```
|- closure s = (:real^N) DIFF (interior ((:real^N) DIFF s))

|- closure((:real^N) DIFF s) = (:real^N) DIFF interior(s)
```

However, when one talks about the 'inside' or 'interior' of a polygon in Pick's theorem, one means not the interior of the polygonal path itself (that would be the empty set since it has zero thickness) but rather of the region it encloses. So the fundamental problem is to define that concept. A clearly relevant result is the Jordan Curve Theorem, which asserts that a simple closed curve divides the plane into an 'inside' and 'outside'. At least, that is how one thinks of it intuitively, but the traditional formal statements simply observe that the complement of such a curve has two connected components, each of which has the curve as its frontier, and exactly one of which is bounded. For a discussion of the formal proof of this statement see (Hales, 2007b).

```
|- ∀c:real^1->real^2.
       simple_path c ∧ pathfinish c = pathstart c
       ⇒ ∃ins out.
                ¬(ins = {}) ∧ open ins ∧ connected ins ∧
                ¬(out = {}) ∧ open out ∧ connected out ∧
                bounded ins ∧ ¬bounded out ∧
                ins INTER out = {} ∧
                ins UNION out = (:real^2) DIFF path_image c ∧
                frontier ins = path_image c ∧
                frontier out = path_image c
```

Intuitively, the components '`ins`' and '`out`' whose existence is asserted are thought of as the 'inside' and 'outside' of the curve. We might elect to introduce these notions simply by Skolemizing these existential quantifiers. However, we prefer to define the concepts in a more general way that might be applicable in other situations. In particular these apply to arbitrary sets in $\mathbb{R}^N$, not just paths in $\mathbb{R}^2$ (even if we might not be able to deduce analogously strong properties). We simply say that a point x is inside (resp. outside) a set S if it is not in S and the connected component of $\mathbb{R}^N - S$ containing x is bounded (resp. unbounded).

```
|- inside s = {x | ¬(x IN s) ∧
                   bounded(connected_component ((:real^N) DIFF s) x)}

|- outside s = {x | ¬(x IN s) ∧
                    ¬bounded(connected_component ((:real^N) DIFF s) x)}
```

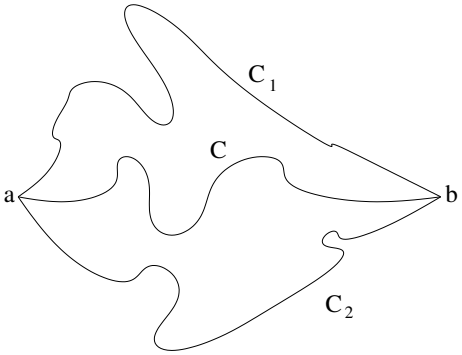
It is fairly straightforward to prove a collection of natural and straightforward properties of these concepts, e.g.

```
|- ∀s. inside s INTER s = {}

|- ∀s. inside s UNION outside s = (:real^N) DIFF s

|- ∀s. inside s = (:real^N) DIFF (s UNION outside s)

|- ∀s t. s SUBSET t ⇒ outside t SUBSET outside s

|- ∀s. bounded s ⇒ bounded((:real^N) DIFF outside s)

|- ∀c c1 c2. c INTER outside(c1 UNION c2) = {}
             ⇒ outside(c1 UNION c2) SUBSET outside(c1 UNION c)

|- ∀s. convex s ⇒ outside s = (:real^N) DIFF s

|- ∀s. bounded s ∧ convex s ⇒ inside(frontier s) = interior s
```

Although in general it might be rather challenging to prove some other significant properties in full generality, we can at least fairly straightforwardly use it to "Skolemize" the Jordan Curve Theorem as we discussed at the outset.

```
|- ∀c:real^1->real^2.
        simple_path c ∧ pathfinish c = pathstart c
        ⇒ ¬(inside(path_image c) = {}) ∧
            open(inside(path_image c)) ∧
            connected(inside(path_image c)) ∧
            ¬(outside(path_image c) = {}) ∧
            open(outside(path_image c)) ∧
            connected(outside(path_image c)) ∧
            bounded(inside(path_image c)) ∧
            ¬bounded(outside(path_image c)) ∧
            inside(path_image c) INTER outside(path_image c) = {} ∧
            inside(path_image c) UNION outside(path_image c) =
            (:real^2) DIFF path_image c ∧
            frontier(inside(path_image c)) = path_image c ∧
            frontier(outside(path_image c)) = path_image c
```

As will be explained later, one of the key ideas in the Pick proof is to divide the inside of a polygon into two pieces by a straight cut across its inside. Once again, we made an attempt to generalize this property from polygonal arcs to arbitrary ones. Suppose that an arc c cuts across the inside of a simple closed curve, meeting it at points a and b, and we separate the original closed curve into two arcs c_1 and c_2 with the endpoints a and b. Our objective is to prove that the inside of the original curve is essentially cut into two pieces, one the inside of the curve defined by c_1 and c, the other the inside of the curve defined by c_2 and c.



We had originally hoped that given the Jordan Curve Theorem, all such natural extensions and variants (we sometimes call this one the 'Jordan Triple Curve Theorem') would be fairly trivial corollaries. But in fact, this is the sort of theorem that is not considered so often in the literature, and we had some trouble finding a suitable proof to formalize. We eventually found a relatively straightforward 14-line proof in (Whyburn, 1964) (1.4, page 31). However, it was still quite hard work to formalize; in particular the process of cutting paths into sub-arcs and glueing them together in different ways proved awkward. This is the final result as stated and proved in HOL Light; note that we merely assume that c has non-empty intersection with the inside of the curve defined by c_1 and c_2 (though this generalization is trivial).

```
|- ∀c1 c2 c a b:real^2.
        ¬(a = b) ∧
        simple_path c1 ∧ pathstart c1 = a ∧ pathfinish c1 = b ∧
        simple_path c2 ∧ pathstart c2 = a ∧ pathfinish c2 = b ∧
        simple_path c ∧ pathstart c = a ∧ pathfinish c = b ∧
        path_image c1 INTER path_image c2 = {a,b} ∧
        path_image c1 INTER path_image c = {a,b} ∧
        path_image c2 INTER path_image c = {a,b} ∧
        ¬(path_image c INTER inside(path_image c1 UNION path_image c2) = {})
        ⇒ inside(path_image c1 UNION path_image c) INTER
            inside(path_image c2 UNION path_image c) = {} ∧
            inside(path_image c1 UNION path_image c) UNION
            inside(path_image c2 UNION path_image c) UNION
            (path_image c DIFF {a,b}) =
            inside(path_image c1 UNION path_image c2)
```

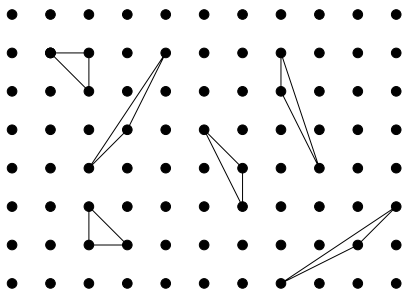
The eventual formal proof of this result was fully 788 lines long. We do not consider this as part of the proof of Pick's theorem specifically, and have installed it in the general multivariate theories. For this reason, it was not included in the line-count of 3709 for Pick's theorem. However, this categorization is arguable, and including these further 788 lines in the count makes the Pick proof even longer than that of the Prime Number Theorem!

3. The proof of Pick's theorem

Having established the necessary background to state and prove Pick's theorem, we can now get to work. Our proof of the theorem is fairly standard, though it does not specifically follow any single informal source. It proceeds by establishing the theorem for increasingly general classes of polygon: first a so-called 'elementary triangle', then an arbitrary lattice triangle and finally a general simple polygon. As will be seen, the first two steps are entirely routine and straightforward to formalize, while the last one presents significant difficulties.

Elementary triangle

An *elementary triangle* is one with vertices at lattice points but containing no other lattice points, either inside or on its boundary. The following picture shows some examples of elementary triangles.



Pick's theorem for such a triangle simply asserts that each such triangle has area $1/2$. There are various relatively straightforward proofs of this result. The one we have formalized is based on linear transformations of the integer lattice.

Given two vectors A and B, we can consider them as defining the linear transformation of the plane $f : (x, y) \mapsto Ax + By$, where juxtaposition indicates scalar-vector multiplication. It is not hard to show that if the image under f of the set of integer lattice points is exactly this same set of integer lattice points, then the determinant of the matrix of f is ± 1:

```
|- ∀f:real^N->real^N.
       linear f ∧ IMAGE f integral_vector = integral_vector
       ⇒ abs(det(matrix f)) = &1
```

Given an elementary triangle OAB, where we take O as the origin, one can show that the integer multiples of the two other vertices generate precisely the integer lattice points. The determinant in the previous theorem is precisely twice the area of the triangle formed by the three vertices, which therefore has area $1/2$.

In formalizing this result, we simply use the convex hull of the set of vertices $\{a, b, c\}$ instead of the more elaborate concept of being 'inside' the triangle abc (though we can connect the two definitions easily enough later). Our characterization of elementary triangle is that the set of integer lattice points inside this convex hull is exactly the vertex set $\{a, b, c\}$. This actually takes in the possibility that the triangle is completely degenerate, i.e. that the three vertices are collinear. So the formal theorem has to take this into account.

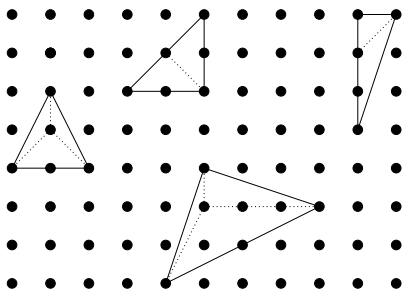
```
|- ∀a b c:real^2.
       {x | x IN convex hull {a,b,c} ∧ integral_vector x} = {a,b,c}
       ⇒ measure(convex hull {a,b,c}) =
              if collinear {a,b,c} then &0 else &1 / &2
```

Arbitrary triangle

Next, we proceed inductively to establish the result for an arbitrary triangle with all its vertices at integer lattice points. If a triangle ABC is not elementary, then it must have a lattice point D either:

— On one if its sides, say AB, in which case we can subdivide it into triangle ADC and BCD.
— In its interior, in which case we can divide it into three triangles ADB, ADC and BDC.

The following diagram illustrates these possibilities for a few examples.



Although this is straightforward enough, we can make it even simpler, by reformulating things slightly so that the first case becomes a special case of the second, and we avoid handling degenerate cases of the theorem itself separately. The key is the following general theorem about 'additivity' of a real-valued function defined on subsets of the plane.

```
|- ∀f:(real^2->bool)->real a b c d.
        (∀s t. compact s ∧ compact t
               ⇒ f(s UNION t) = f(s) + f(t) - f(s INTER t)) ∧
        ¬(a = b) ∧ ¬(a = c) ∧ ¬(b = c) ∧
        ¬affine_dependent {a,b,c} ∧ d IN convex hull {a,b,c}
        ⇒ f(convex hull {a,b,c}) =
            (f(convex hull {a,b,d}) +
             f(convex hull {a,c,d}) +
             f(convex hull {b,c,d})) -
            (f(convex hull {a,d}) +
             f(convex hull {b,d}) +
             f(convex hull {c,d})) +
            f(convex hull {d})
```

In the inductive proof of Pick's theorem for an arbitrary lattice triangle, we can apply this general result twice, once with f giving the number of lattice points in a set and once with it giving the measure (area) of a set. This results in an almost immediate proof of the following reformulation of Pick's theorem:

```
|- ∀a b c:real^2.
    integral_vector a ∧ integral_vector b ∧ integral_vector c
    ⇒ measure(convex hull {a,b,c}) =
        &(CARD {x | x IN convex hull {a,b,c} ∧ integral_vector x}) -
        (&(CARD {x | x IN convex hull {b,c} ∧ integral_vector x}) +
         &(CARD {x | x IN convex hull {a,c} ∧ integral_vector x}) +
         &(CARD {x | x IN convex hull {a,b} ∧ integral_vector x})) / &2 +
        &1 / &2
```

It is straightforward to show it equivalent to the usual formulation with a proviso of nondegeneracy.

```
|- ∀a b c:real^2.
        integral_vector a ∧ integral_vector b ∧ integral_vector c
        ⇒ measure(convex hull {a,b,c}) =
                if collinear {a,b,c} then &0
                else &(CARD {x | x IN interior(convex hull {a,b,c}) ∧
                                 integral_vector x}) +
                     &(CARD {x | x IN frontier(convex hull {a,b,c}) ∧
                                 integral_vector x}) / &2 - &1
```

Arbitrary polygon

Again, we proceed inductively, showing that any polygon can be subdivided into two by a line joining two vertices and otherwise lying entirely in the inside. (This can also be used to drive an inductive proof that any polygon can be triangulated, and that is where we looked for a proof, though we don't explicitly deduce this general result.) The informal proof, essentially excerpted from (Hales, 2007a), seems relatively straightforward:

Pick the coordinate axis so that no two vertices have the same y coordinate. Let B be the lowest vertex on the polygon, and let A and C be adjacent to B. If AC is an interior diagonal, we draw the diagonal AC, forming a triangle ABC and a polygon without the vertex B. Otherwise, let D be a vertex of the polygon at maximal distance from the line AC in the direction of B. Cut the polygon into two along the edge BD.

The first challenge is to formalize the 'pick the coordinate axis' step. An earlier paper (Harrison, 2009c) described an extensive framework for such 'without loss of generality' reasoning, but to support the present proof, this had to be generalized from 'first order' concepts like points and lines to 'higher order' concepts like polygonal paths and lists of points. The following 'WLOG

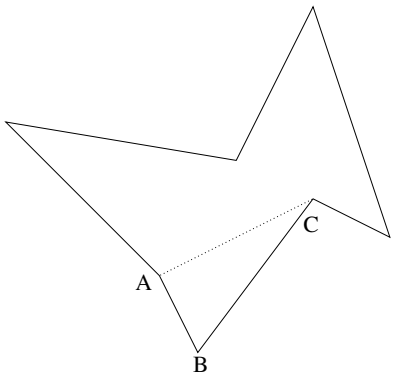
lemma' gives the schematic justification for picking a suitable `transform` f that enables us to assume that some particular object x has a `nice` property.

```
|- (∀x. ∃f y. transform f ∧ nice y ∧ f y = x)
   ⇒ ∀P. (∀f x. transform f ⇒ (P(f x) ⇔ P x)) ∧
         (∀x. nice x ⇒ P x)
         ⇒ ∀x. P x
```

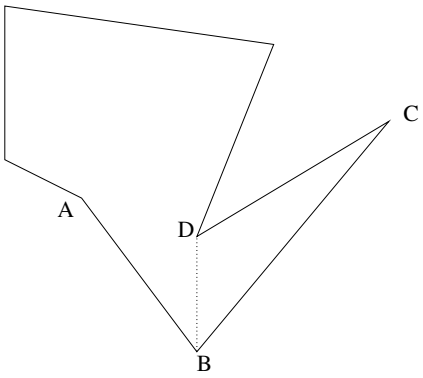
In the applications considered in (Harrison, 2009c), the objects were always individual vectors, and the transformations were aimed at making one of these vectors have some particularly convenient property. In a typical example, one chooses a spatial translation to turn one point into the origin; in another one chooses an orthogonal transformation (roughly a rotation or rotoinversion) to align one point with a chosen rectangular coordinate axis. In our case, our core objects are *lists* of vectors and we desire an orthogonal transformation to ensure that all elements of that list have distinct y coordinates. In order to apply it, we just need to prove that such an orthogonal transformation exists. This is fairly straightforward because there are infinitely many possible angles of rotation in the plane and only finitely many pairs of vertices in the list.

```
|- ∀p:(real^2)list.
       ∃f q. (∃g. orthogonal_transformation g ∧ f = MAP g) ∧
             (∀x y. MEM x q ∧ MEM y q ∧ ¬(x = y) ⇒ ¬(x$2 = y$2)) ∧
             f q = p
```

Having achieved the desired coordinate transformation, the remainder of the informal proof can be carried through. However, even this turned out to be quite difficult. Recall that the informal proof has two cases. Once we have established the triangle ABC with B as the lowest (and *strictly* lowest, since all y coordinates are different) vertex, we either split the polygon with a line AC



or choose a point D in the interior of ABC at maximal distance from the line AC and split the polynomial with the line BD

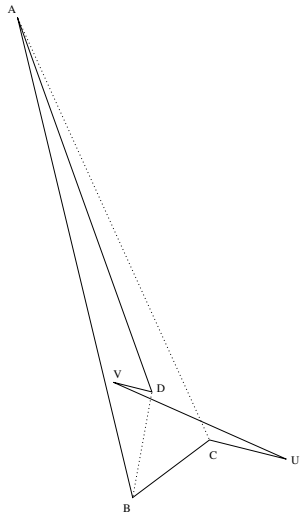


In either case, to apply the Jordan Triple Curve Theorem, we need to establish that this new cut has nonempty intersection with the inside of the original polygon and does not intersect the polygon itself except at the endpoints. Simply establishing the first requirement is not trivial, despite its intuitive obviousness. Essentially, to get started we need to establish some point that is in both the inside of the polygon and the interior of the triangle ABC. Once that is achieved, it's fairly straightforward in either case to argue that we can join this point to at least some point on the new line segment without touching the polygon itself.

We construct a point known to be inside the polygon and the interior of the triangle ABC as follows. First, we pick points A' on AB and C' on AC (neither of them being B) such that of the vertices in the original polygon, only B has a smaller y-coordinate than either A' or C'. After establishing the the triangle $A'BC'$ is not degenerate (i.e. those three points are affinely independent) we can easily show that there is a point in its interior. Intuitively, we believe it is also in the inside of the polygon because there are no other points inside $A'BC'$ to "interfere". To show this formally we use a kind of parity lemma. Roughly this asserts that given any simple closed curve including a line segment from a to b, if another segment from c to d crosses the line segment from a to b exactly once and does not otherwise intersect the curve, then c and d have opposite 'inside' and 'outside' status.

```
|- ∀a b c d p x:real^2.
        simple_path(p ++ linepath(a,b)) ∧
        pathstart p = b ∧ pathfinish p = a ∧
        segment(a,b) INTER segment(c,d) = {x} ∧
        segment[c,d] INTER path_image p = {}
        ⇒ (c IN inside(path_image(p ++ linepath(a,b))) ⇔
             d IN outside(path_image(p ++ linepath(a,b))))
```

It is easy to establish that points below our vertex B are outside the polygon, and then to use this parity lemma to deduce that our constructed point lies inside. This still leaves the task of proving that our new cut does not intersect the original polygon except at its endpoints. In the first case above, this is relatively easy, but the second is a bit more involved. The idea of picking the new vertex D as far from the line AC as possible (as opposed to, for example, the lowest vertex in the interior of ABC) is to avoid cases like the following where another vertex V manages to 'sneak inside' so that our new cut BD does after all intersect the polygon.



The geometric reasoning underlying the fact that our choice of vertex D indeed ensures non-intersection is in principle straightforward, but we found it quite hard work reasoning about obvious facts like 'this point and that point are on opposite sides of the line'. Nevertheless, we do ultimately obtain our goal of proving that a polygon can always be chopped into two pieces with a cut as required. Note that the condition that the length of the vertex list is at least 5 corresponds to the fact that the polygon has at least 4 vertices (since the first and last elements of the vertex list are the same).

```
|- ∀p:(real^2)list.
        simple_path(polygonal_path p) ∧
        pathfinish(polygonal_path p) = pathstart(polygonal_path p) ∧
        5 <= LENGTH p
        ⇒ ∃a b. ¬(a = b) ∧ MEM a p ∧ MEM b p ∧
                  segment(a,b) SUBSET inside(path_image(polygonal_path p))
```

As a result, we can rather easily drive an inductive proof based on number of vertices to obtain the final Pick theorem:

```
|- (∀x. MEM x p ⇒ integral_vector x) ∧
   simple_path(polygonal_path p) ∧
   pathfinish(polygonal_path p) = pathstart(polygonal_path p)
   ⇒ measure(inside(path_image(polygonal_path p))) =
       &(CARD {x | x IN inside(path_image(polygonal_path p)) ∧ integral_vector x}) +
       &(CARD {x | x IN path_image(polygonal_path p) ∧ integral_vector x}) / &2 -
       &1
```

4. Conclusions and related work

A presentation by Narboux et al.† discusses a Coq proof of part of the theorem, but leaves a number of gaps and as far as we know has not been completed. So, as far as we are aware, the work we describe here is the first successful formalization of Pick's theorem. Nevertheless, it

† "Vers une preuve formelle du théorème de Pick en Coq", `http://galapagos.gforge.inria.fr/December2008/mns-galapagos-dec-2008.pdf`.

was somewhat chastening for us to discover how difficult some apparently elementary pieces of intuitive geometric reasoning were to formalize. The main culprit seems to be the concept of 'inside', which is a very natural notion but difficult to reason about. It would be a very appealing project to try to automate more of it, perhaps by a more searching analysis of our everyday intuition.

Our general approach to Pick's theorem was, naturally enough, chosen based on our belief that it would be relatively easy to formalize. In fact, the proof up to the case of an arbitrary lattice triangle worked very nicely, and was streamlined by the use of the inclusion-exclusion result for additive functions. As one of the reviewers pointed out, this is reminiscent of the inclusion-exclusion expansion over faces in the Euler-Poincaré formula and some of its common proofs (Barvinok, 2002; Webster, 1995). We believe that the same sort of reasoning would easily generalize to the case of a convex polygon. Indeed, one proof we looked at in the literature and were considering adopting for our formalization (Murty and Thain, 2007) actually proves the theorem only for a convex lattice polygon.

All in all, the main difficulties are connected with the treatment of arbitrary polygons and their 'inside' and 'outside'. In this domain, many intuitively obvious facts require a considerable amount of work to prove rigorously, and it is often difficult to find a thorough informal treatment to use as a model without digging into fairly old literature such as (Lennes, 1911). We may have made things more difficult for ourselves by generalizing where possible from polygons to arbitrary paths, but our hope is that in compensation the background lemmas are more likely to be useful elsewhere. Here again we needed to explore older literature to find proofs of results like the 'Jordan Triple Curve theorem'. As noted, we took our proof from (Whyburn, 1964). The same result is discussed in (Newman, 1939), but it is proved by a modified form of a particular proof of the Jordan Curve Theorem, not deduced directly from the result itself. A restricted form for polygons, which would have sufficed for our present purposes, is proved by (Thomassen, 1992).

One naturally wonders whether a radically different proof of Pick's theorem might avoid all the difficulties associated with triangulation of polygons, perhaps using results such as the Gauss-Green theorem relating line and plane integrals. Some intriguingly different proofs are given by (Kurogi and Yasukura, 2005; Diaz and Robins, 1995; Blatter, 1997), all of which may generate interesting formalization challenges. It would also be natural to consider formalizing extensions and generalizations of Pick's theorem, and other related results. One possibility is to extend it to non-simple polygons using Euler characteristics (Dubeau and Labbé, 2007). Another is to consider higher-dimensional generalizations using Ehrhart polynomials (Ehrhart, 1967). One more interesting future direction would be to use Pick's theorem as a case study for automated theory exploration, since the concepts it involves are combinatorial and finite, lending themselves to automated conjecture formation and testing.

Acknowledgements

The author is grateful to Tom Hales for help with the Jordan Triple Curve Theorem proof, to the referees for their corrections and insightful suggestions, and to the audiences in several talks on the subject for additional ideas.

References

Barvinok, A. (2002). *A Course in Convexity*, volume 54 of *Graduate Texts in Mathematics*. American Mathematical Society.

Blatter, C. (1997). Another proof of Pick's area theorem. *Mathematics Magazine*, 70:200.

Diaz, R. and Robins, S. (1995). Pick's formula via the Weierstrass $\wp$-function. *The American Mathematical Monthly*, 102:431–437.

Dubeau, F. and Labbé, S. (2007). Euler's characteristics and Pick's theorem. *International Journal of Contemporary Mathematical Sciences*, 2:909–928.

Ehrhart, E. (1967). Sur un problème de géométrie diophantienne linéaire II. *Journal für die reine und angewandte Mathematik*, 227:25–49.

Hales, T. C. (2007a). Easy pieces in geometry. Available at `http://www.math.pitt.edu/~thales/papers/`.

Hales, T. C. (2007b). The Jordan curve theorem, formally and informally. *The American Mathematical Monthly*, 114:882–894.

Harrison, J. (1996). HOL Light: A tutorial introduction. In Srivas, M. and Camilleri, A., editors, *Proceedings of the First International Conference on Formal Methods in Computer-Aided Design (FMCAD'96)*, volume 1166 of *Lecture Notes in Computer Science*, pages 265–269. Springer-Verlag.

Harrison, J. (2005). A HOL theory of Euclidean space. In Hurd, J. and Melham, T., editors, *Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005*, volume 3603 of *Lecture Notes in Computer Science*, pages 114–129, Oxford, UK. Springer-Verlag.

Harrison, J. (2009a). A formalized proof of Dirichlet's theorem on primes in arithmetic progression. *Journal of Formalized Reasoning*, 2(1):63–83.

Harrison, J. (2009b). Formalizing an analytic proof of the Prime Number Theorem (dedicated to Mike Gordon on the occasion of his 60th birthday). *Journal of Automated Reasoning*, 43:243–261.

Harrison, J. (2009c). Without loss of generality. In Berghofer, S., Nipkow, T., Urban, C., and Wenzel, M., editors, *Proceedings of the 22nd International Conference on Theorem Proving in Higher Order Logics, TPHOLs 2009*, volume 5674 of *Lecture Notes in Computer Science*, pages 43–59, Munich, Germany. Springer-Verlag.

Kurogi, T. and Yasukura, O. (2005). From Homma's theorem to Pick's theorem. *Osaka Journal of Mathematics*, 42:723–735.

Lennes, N. J. (1911). Theorems on the simple finite polygon and polyhedron. *American Journal of Mathematics*, 33:37–62.

Murty, M. R. and Thain, N. (2007). Pick's theorem via Minkowksi's theorem. *The American Mathematical Monthly*, 114:732–736.

Newman, M. H. A. (1939). *Elements of the Topology of Plane Sets of Points*. Cambridge University Press.

Pick, G. (1899). Geometrisches zur Zahlenlehre. *Sitzungsberichte des deutschen naturwissenschaftlich-medicinischen Vereines für Böhmen "Lotos" in Prag, Series 2*, 19:311–319.

Thomassen, C. (1992). The Jordan-Schoenflies theorem and the classification of surfaces. *The American Mathematical Monthly*, 99:116–130.

Webster, R. (1995). *Convexity*. Oxford University Press.

Whyburn, G. T. (1964). *Topological Analysis*, volume 23 of *Princeton Mathematical Series*. Princeton University Press, revised edition.