Challenges in Formalizing Geometric Theorems:
A Case Study of Pick’s theorem

John Harrison

AUTOMATHEO 2010, Edinburgh

15th July 2010, 10:00–11:00

Pick’s theorem

Take a simple polygon with vertices at integer lattice points, i.e. where both x and y coordinates are integers.

Let I be the number of integer lattice points in its interior and B be the number on its boundary.

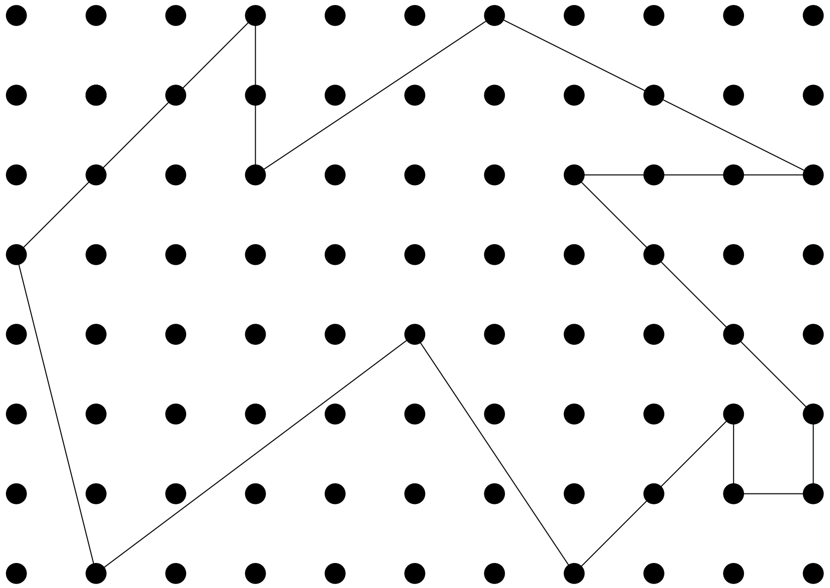
Then the area of the polygon is given by

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

First proved by G. Pick in 1899 in his paper ‘Geometrisches zur Zahlenlehre’

Example

This polyon has 30 interior points and 22 boundary points.



So its area is $30 + 22/2 - 1 = 40$.

Challenge 1 for automated theory exploration

Could Pick’s theorem be discovered automatically? Could generalizations be discovered? For example, to non-simple polygons, or to higher dimensions (Ehrhart polynomials etc.)

Pick’s theorem seems difficult to formalize

The HOL Light proof of Pick’s theorem is quite long, comparable to some much ‘deeper’ results like the PNT:

Dirichlet’s theorem	2082 lines
Pick’s theorem	**3709** lines
Prime Number Theorem	4314 lines

What accounts for this apparent difficulty? (Except possibly for my incompetence.)

Pick theorem difficulties

Some sources of the difficulty:

- Requires formalization of informal geometric concepts like ‘inside’.
- Leads to lemmas with more generality, whose proofs become correspondingly harder.
- Requires simplifying methods to exploit symmetries or choose convenient coordinates.
- Many geometrically intuitive facts need to be explicitly deduced algebraically.

Some of these difficulties also make it interesting from the point of view of automated theory exploration.

Insides and outsides

We want to define ‘inside’ and ‘outside’ of a polygon in a natural way.

We aim for a definition that works in much more general situations, with the polygon as a very special case.

- The *inside* of a set S is the set of points $x \notin S$ such that the connected component of $\mathbb{R}^N - S$ containing x is *bounded*.

- The *outside* of a set S is the set of points $x \notin S$ such that the connected component of $\mathbb{R}^N - S$ containing x is *unbounded*.

So we always have $\mathbb{R}^N = S \cup \mathsf{inside}(S) \cup \mathsf{outside}(S)$ as a disjoint union. Note that $\mathsf{interior}(S)$ is the topological interior of S, which is completely different.

HOL definitions of inside and outside

```
|- 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)}
```

These concepts satisfy many straightforward properties such as:

```
|- ∀s. inside s INTER s = {}

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

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

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

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

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

|- ∀s. closed s ⇒ open(inside s)
```

Challenge 2 for automated theory exploration

There are doubtless several alternative ways of defining ‘inside’ and ‘outside’. For example, we could include the set S in inside(S), or we could use path components instead of connected components, or do something more radically different. Could automated theory exploration discover which definitions satisfy the cleanest or most convenient general properties?

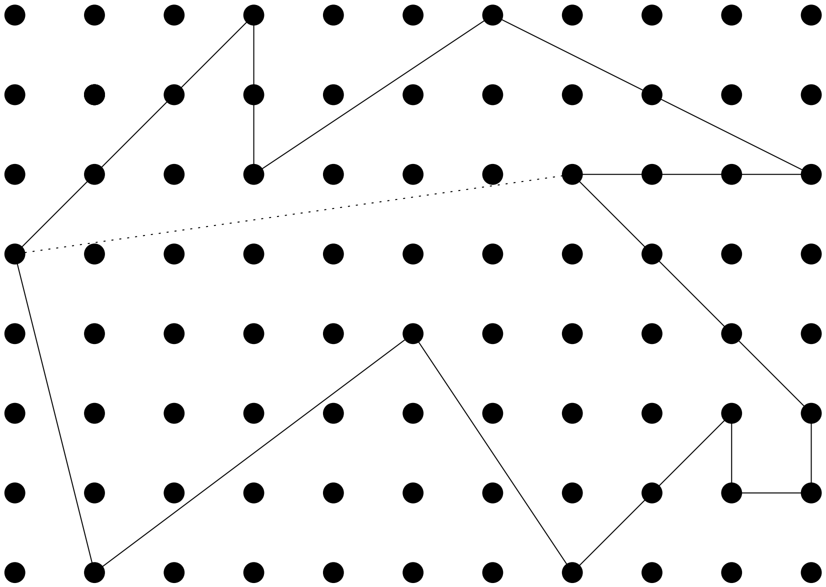
The Jordan Curve Theorem

We can ‘skolemize’ the Jordan Curve Theorem to express it using inside and outside:

```
|- ∀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
```

Chopping a polygon in half

In the proof of Pick’s theorem, a key idea is to split the (inside of) a polygon into two pieces by a line segment connecting two vertices and lying in the inside (except for the endpoints).

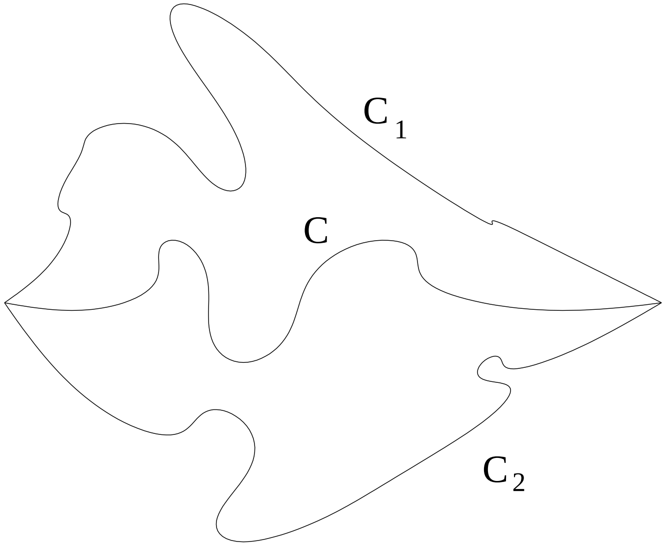


This can drive an ‘inductive’ approach the the proof.

The ‘Jordan triple curve theorem’

We would like to generalize this from the special case of a polygon to any simple closed curve in $\mathbb{R}^2$.

In this kind of situation, we want to prove that the insides of $C_1 + C$ and $C_2 + C$ split the inside of $C_1 + C_2$ in the obvious way.



The ‘Jordan triple curve theorem’ in HOL

```
|- ∀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 Jordan triple curve theorem isn’t trivial

It’s well-known that the Jordan Curve Theorem is surprisingly hard to prove.

However, one might hope that given the JCT, the J3CT and all such elementary properties would be straightforward corollaries.

However, this seems not to be the case. After quite a search we eventually found a 14-line proof based on JCT in Whyburn’s “Topological Analysis”. However, its formalization took 788 lines . . .

Challenge 3 for automated theory exploration

There are many analogous theorems that one might need. Is there a systematic way of discovering proofs for results like J3CT about chopping up or combining the insides of curves?

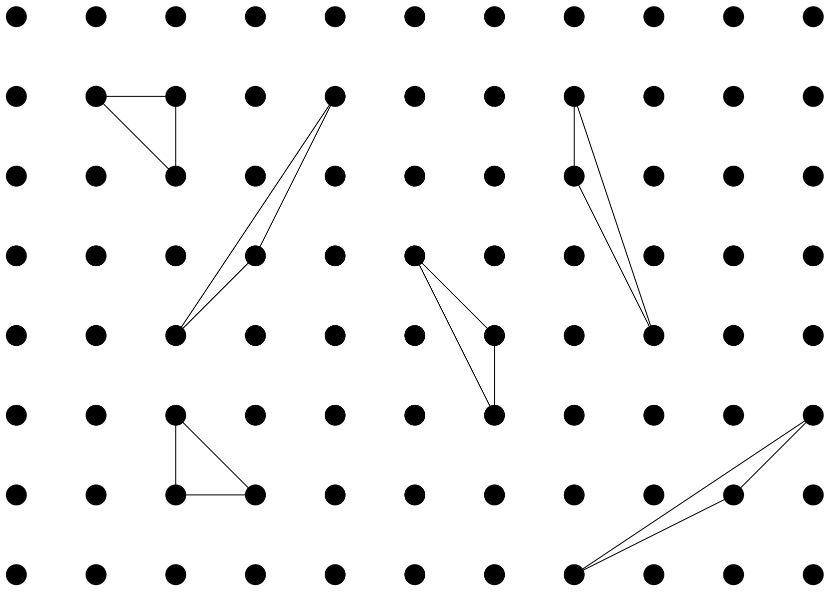
Outline of Pick proof

The proof of Pick's theorem proceeds by establishing it for successively more general classes of polygon:

- For an *elementary* triangle
- For an arbitrary triangle with vertices at lattice points
- For an arbitrary simple polygon with vertices at lattice points

Pick’s theorem for an elementary triangle

An elementary triangle is one with vertices at lattice points but containing no other lattice points inside or on its boundary.



Pick’s theorem then simply says that each such triangle has area $0 + 3/2 - 1 = 1/2$.

Proof for an elementary triangle

In general, if the image of the integer lattice points under a linear map is exactly that same set of lattive points, then the map has determinant ± 1:

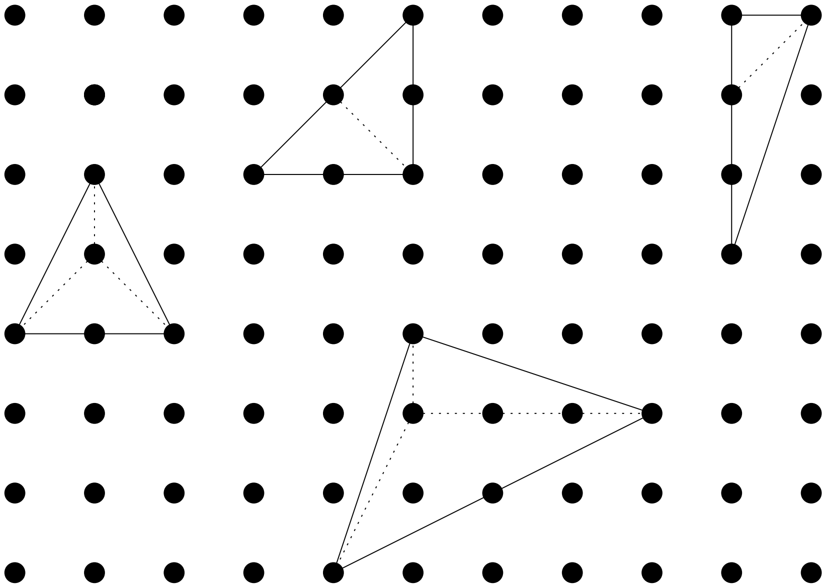
```
|- ∀f:real^N->real^N.
        linear f ∧ IMAGE f integral_vector = integral_vector
        ⇒ abs(det(matrix f)) = &1
```

Use one vertex as the origin and consider the other two points as vectors A and B. Consider the linear transformation of the plane $f : (x, y) \mapsto Ax + By$. This does map lattice points to lattice points, and the area of the triangle is half the determinant. **But note the degenerate cases!**

```
|- ∀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
```

Proof for an arbitrary triangle

Any non-elementary triangle with vertices at lattice points can be split into 2 or 3 others, e.g.



This can then be used to drive an inductive proof for any lattice triangle.

Avoiding case analysis

At a stroke we can avoid considering degenerate triangles while treating subdivision into 2 as a special case of subdivision into 3.

```
|- ∀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
```

We can use this to support the inductive proof, and prove the equivalence to the usual Pick formula as well.

Challenge 4 for automated theory exploration

Would it be possible to arrive at such reformulations automatically?

Pick for an arbitrary polygon

We can continue in the same inductive style for an arbitrary simple lattice polygon, provided we establish that any polygon can be cut into two by a line joining two vertices.

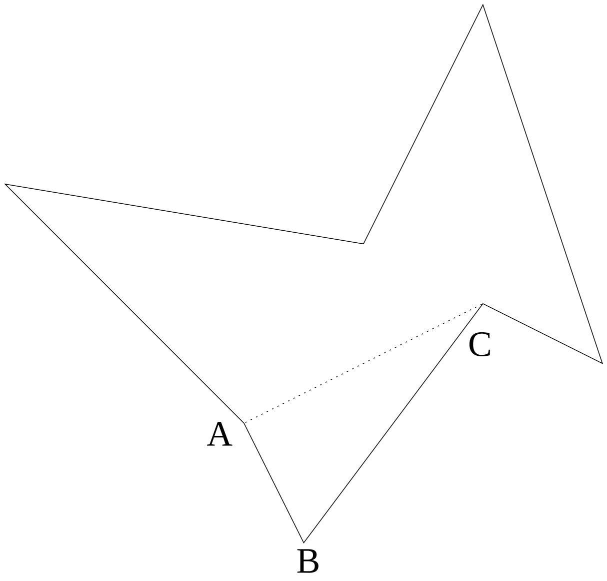
This can also be used as the key lemma to drive a proof that the inside of a simple polygon can be triangulated (without introducing any new vertices).

We haven’t actually formally proved the triangulation as a separate result, but doing so would now be easy.

Informal proof (1/2)

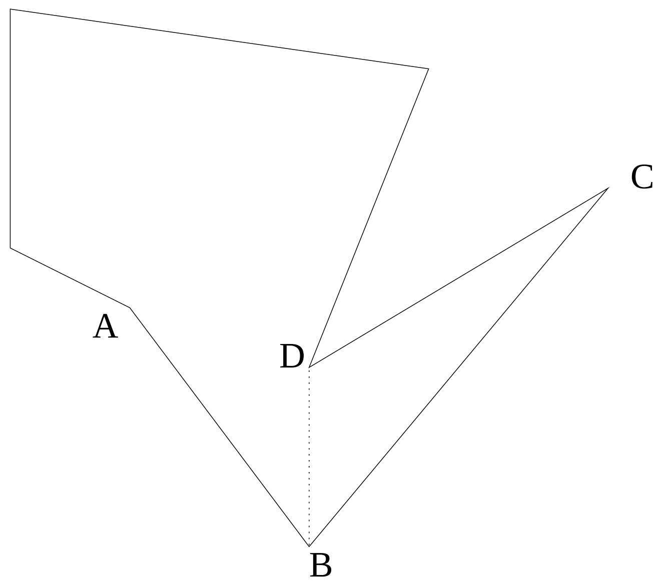
The first step is to pick the coordinate axis so that no two vertices have the same y coordinate. Let B be the lowest vertex on the polygon, and let A and C be adjacent to B.

If AC is an interior diagonal, we draw the diagonal AC, forming a triangle ABC and a polygon without the vertex B.



Informal proof (2/2)

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.



WLOG

The first step, picking a suitable coordinate system 'without loss of generality', is something we've already worked hard to support in HOL Light. (See '*Without Loss of Generality*', TPHOLs 2009.)

This suite of tactics largely automates the task of ensuring without loss of generality that a particular vector x has a suitable property P.

- Proves that there is a suitable translation and orthogonal transformation f so that $P(f(x))$.
- Systematically uses the surjectivity of f to transform quantifiers $\forall x.\ \phi[x]$ to $\forall x.\ \phi[f(x)]$ etc.
- Uses a suite of theorems about how functions and predicates preserve or are unchanged by functions like f to eliminate it.

Enhanced WLOG machinery

To support the present proof, we need to show that given a list of points q there is an orthogonal transformation g such that the preimages under g of the points in q have distinct y coordinates:

```
|- ∀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
```

This can be combined with a standard general WLOG lemma:

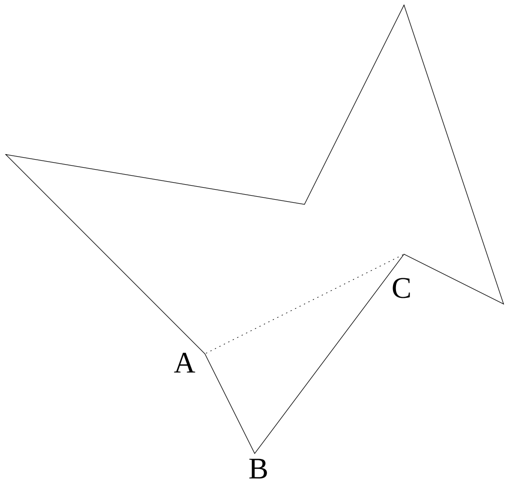
```
|- (∀x. ∃f y. transform f ∧ nice y ∧ f y = x)
   ⇒ ∀P. (∀f x. transform f ⇒ (P(f x) <=> P x)) ∧
          (∀x. nice x ⇒ P x)
          ⇒ ∀x. P x
```

The WLOG machinery needs to be enhanced with quantifier mappings for more higher-order variables such as polygonal paths.

Proving points are inside

We need to show that the line we choose (excluding its endpoints) does lie entirely in the inside of the polygon.

In the second case in the proof, it takes only a little thought. In the first case, it seems utterly evident:



But to deduce it rigorously from our inside definition is still quite a bit of work!

The parity lemma

We prove a parity lemma showing how inside/outsde status ‘flips’ if we cross a polygon (or line segment portion of any path) exactly once:

```
|- ∀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))))
```

The proof is not too hard from the Jordan Curve Theorem, since the segment a, b is a limit point both for the inside and outside. But the details still need work (400 lines of proof).

Using the parity lemma

Here's the approach to showing that (A, C) lies entirely inside the polygon:

- Prove by basic considerations that a point lower than all vertices is in the *outside*.
- Prove by the parity lemma that a point Z just above B in the interior of triangle ABC is in the *inside*.
- Prove that all points of (A, C) can be joined to Z by a segment lying entirely in the interior of triangle ABC.

The result follows.

Challenge 5 for automated theory exploration

Much of this reasoning is justifying the geometrically self-evident. Is there some way of having the machine construct a proof automatically in such cases?

The final result

```
|- ∀p:(real^2)list.
        (∀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
```

Conclusions

The proof of Pick’s theorem is, on the face of it, surprisingly hard to formalize.

Much of the difficulty lies in the formalization of geonetric reasoning that is very simple and obvious.

In itself, it is a good source of possible challenges for automated theory exploration.

Such methods might be applicable in the case of other geometric theorems too.