Propositional Proofs via Combinatorial Geometry and the Search for Symmetry

Alessandra Carbone[1] and Stephen Semmes[2] *

[1] Mathématiques/Informatique
Université de Paris XII
e-mail: ale@univ-paris12.fr
[2] Department of Mathematics
Rice University
e-mail: semmes@math.rice.edu

Abstract. This paper is motivated by questions of complexity and combinatorics of proofs in the sequent calculus. We shall pay particular attention to the role that *symmetry* plays in these questions.
We want to have combinatorial models for proofs, models that will reflect complexity phenomena as well as accommodate other kinds of mathematical structures or relationships. For *interpolation* in proofs such a model was given in [3], and indeed this model could adapt itself to all sorts of structures (concerning graphs, polygons, surfaces...). However this was a model for *cut-free* proofs, and we would like to deal with the combinatorial structure of proofs with cuts. Proofs with cuts can have *cycles*, and the structure of these cycles is considered in [6], where they are analyzed through associated *finitely presented groups*. In some cases these groups are strongly *distorted*, and this can be seen as a reflection of the processes underlying the proof.
The model discussed here is related to the concept of the *universal covering surface* in topology. We shall deal only with phenomena connected to propositional proofs, without the cycling and substitutions and recursion to be found with quantifiers. We shall be working in the range of exactly one level of exponential expansion, and we shall investigate the role of symmetry.

* The second author was supported by the U.S. National Science Foundation. Both authors are grateful to IHES for its hospitality.

1 Proofs, algorithms, and implicit descriptions

What is a proof? What is it that a proof really *does*?

One might try to think of a proof as being like an algorithm, in which terms evolve and a construction is in progress. This is an appealing idea, and indeed one can often see natural "inputs" for a proof. To make these ideas more concrete let us consider the concept of *feasible numbers*. As in [13], one works in arithmetic but allows an extra unary predicate $F(\cdot)$ for which the intended meaning of $F(n)$ is that "n is feasible". All natural numbers are feasible, but one is interested in the sizes of proofs of feasibility (without allowing induction over formulas containing F). A proof of the feasibility of a number implicitly describes a construction of the number.

More precisely, one allows special rules to say that 0 is a feasible number, and that feasibility is preserved by taking sums, products, or applying the successor function. One may wish to consider other special rules, but in any case a proof of feasibility ought to contain (at least implicitly) a recipe for constructing the number.

More generally, one can apply the idea of feasibility to other objects in mathematics, like the elements of a finitely generated group. A proof of feasibility again describes a construction, according to whatever primitive rules or building blocks one wants to allow. This is discussed in [7]. Thus one can use lengths of proofs of feasibility as a new way to measure information content. It provides alternatives to the word metric for finitely generated groups, for instance. It lends itself well to situations in which there are natural dynamical processes present. In the context of the rational numbers, for instance, one has the action of $SL(2, \mathbf{Z})$ on $\mathbf{Q}$ by projective transformations, and this can be used to make short proofs of feasibility [7].

Feasibility provides a language to make explicit the idea of an algorithmic construction. Through it one can see clearly how proofs can code recursive substitutions, what is the meaning of cycling within a proof [4], and the exponential and multiexponential effects in proofs in various mathematical structures [7]. We use it here to motivate combinatorial choices, and to present a simplified model for the complexity of proofs based only on combinatorial and geometric notions, without the logical interpretation.

Substitution, recursion, and cycling are phenomena which occur with quantifier rules, but even with only propositional rules of logic one can already see interesting effects. Short proofs of feasibility of large numbers can be built using cuts and contractions, and may present exponential speed-up over direct constructions [4, 7]. In the basic example one proves $F(t) \rightarrow F(t^2)$ for any term t, and then combines a series of these proofs using cuts to win an exponential in the exponent. This type of argument is very robust, and can be extended to other mathematical structures, as in [7].

Proofs, like algorithms, may be seen as *descriptions* of constructions rather than actual constructions. To decode them into actual constructions may require a procedure like cut elimination, which induces large expansion. (See [9] for an introduction to the combinatorics and complexity of cut elimination.) However, the idea of proofs as algorithms suffers a deficiency. A proof of the feasibility of a particular number deals only with that number. Thus proofs may deal with specific objects while algorithms are in essence more general.

These considerations might lead one to conclude that a proof can represent an *idea* for an algorithm rather than an actual algorithm. What about infinite families of proofs which "look alike"? Can one capture the notion of "algorithm" in a more complete way through families of proofs? The families of short proofs of the feasibility of large numbers in [4] have this flavor, reflecting a single basic idea even if no one proof alone expresses the general rule.

Proofs often come in natural infinite families, as in the examples for feasible numbers, or the propositional codings of the Pigeon-Hole Principle. One would like to have ways to compare proofs, to see patterns among them, and to be able to see infinite families of related proofs as finite pieces of infinite limiting objects. Limiting objects which one might expect to enjoy more symmetry than the individual finite approximations, and to reflect the compactness of some underlying algorithm. Imagine infinite limiting objects which then have some kind of simple tiling, so that they can be "folded" like paper down to a small figure. This kind of compression is relevant for complexity, for obtaining smaller representations.

There is another aspect of symmetry for proofs, and it concerns cut elimination. If a short proof with cuts becomes much longer after cut elimination, then one might expect the cut-free version to have a lot of symmetry, i.e., many similar "patterns", like many copies of the same subproof. In particular, if all propositional tautologies admit proofs of polynomial size (as a function of the size of the tautology), then one would want to conclude that this kind of symmetry is frequently present, since propositional tautologies can often be expected to have only exponentially large cut-free proofs. One should keep in mind here the well-known fact that there is a propositional proof system in which *all* tautologies have polynomial-sized proofs if and only if NP = co-NP [10].

One would also like to say that if a large object is somehow described by a short proof – like a short proof of the feasibility of a large number – then that should imply the presence of some kind of symmetry of the large object, like many similar patterns in the term describing the large number. This theme is discussed in [7], but it is hard to make it precise in a general way. It should be related to the matter of cut elimination, because it is often the cut-free proofs that describe the underlying objects in a complete way, gram by gram, so to speak. Thus one might expect that symmetries in cut-free proofs correspond directly in some approximate way to symmetries in the underlying objects, and that short proofs with cuts lead to symmetries of both the cut-free proofs and the underlying mathematical objects.

These questions of symmetry appear to be fundamental. They are compatible with the idea sometimes discussed these days that questions like P = NP or NP = co-NP have negative answers, but that the sequence of examples which express their failure are hard to write down, themselves not following a simple pattern. To "break" these questions one may have to find intelligent ways to break underlying symmetries. In the context of algorithms, proofs, and combinatorics the notion of *symmetry* is quite flexible, and that makes for part of the difficulty.

To make progress on these ideas we would like to have combinatorial models which bring out some of the aspects related to complexity and symmetry in a more manageable way. For this purpose we consider the concept of *optical graphs*.

2 Proofs and optical graphs

By an *optical graph* we mean an oriented graph with the properties that each vertex has at most three edges attached, and that there are at most two edges oriented away from any given vertex, and at most two oriented towards a given vertex.

A basic point is that *logical flow graphs* associated to formal proofs are optical graphs. The logical flow graph of a proof traces the flow of occurrences of formulas in a proof, and was introduced by Buss [2]. A related graph associated to proofs was introduced earlier by Girard [11]. For our purposes it is better to use a variant of the notion of Buss, in which we restrict ourselves to *atomic* occurrences of formulas, as in [5].

Logical flow graphs carry a natural orientation, as discussed in [5]. This orientation reflects the underlying flow of information; roughly speaking, the positive orientation goes from *hypotheses* to *conclusions*. It is easy to check that logical flow graphs are indeed optical graphs. The main point is that branch points come from the use of the contraction rule in formal proofs.

Conversely, we can sometimes convert optical graphs into proofs, proofs of feasibility, and this conversion captures some of the phenomena of complexity in which we are interested. One of the main points here is that the complexity of proofs is related to the use of contractions in an essential way, and this plays a large role in the combinatorial model as well, through the branch points. We shall return to this point soon.

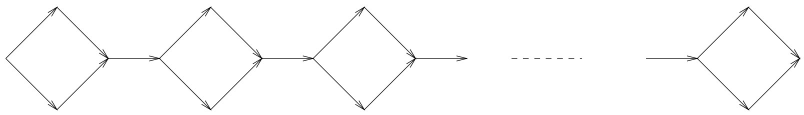
Optical graphs provide a simple way to model some of the activity which takes place within a proof. A node in the logical flow graph of a proof might be considered as some kind of "atomic fact", and then it is natural to try to trace this fact through the proof, to ask how it is used or where it came from. To see how it is used one might follow *positively* oriented paths from the given point, while *negatively* oriented paths tell where the fact came from. *Branching* phenomena are fundamental to proofs; one establishes a piece of information, and afterwards one is permitted to use it twice. This is a consequence of the contraction rule. One might establish a piece of information, use it twice to establish a second piece of information, use that twice to establish a third, and so forth.

Such a chain of reasoning uses *cuts* in an essential way. It corresponds to the ability to make lemmas, and can lead to *exponential* effects; for a chain of reasoning of length n, the same piece of information might be used concretely 2^n times, even if the process itself is coded in a compact way and can be described in merely n steps.

This kind of exponential effect has a simple counterpart in the context of optical graphs. To see this we define the notion of the *visibility* of an optical graph G starting at a vertex v. This is an oriented graph $\mathcal{V}_+(v, G)$ which we define in the following way. For the set of vertices in $\mathcal{V}_+(v, G)$ we take the set of positively oriented paths in G which begin at v. Given two such paths s and t, we have an edge in $\mathcal{V}_+(v, G)$ from s to t if t is obtained by adding an edge at the end of s.

This is an analogue of the idea of covering surfaces from topology (see [1, 12]), and it arises in a natural way in the context of proofs as well. A vertex in the logical flow graph of a proof represents a single atomic fact, and the visibility at that vertex represents the totality of ways that that fact is used in the proof. Alternatively one can reverse orientations so that the visibility represents the totality of information which supports the given fact.

For optical graphs one can see exponential behavior in a simple way. Consider the following example:



If this is our graph G and we choose v to be the leftmost endpoint, then the visibility of G at v is a binary tree with root corresponding to v. If we have n loops in G, then our tree has n levels, and exponentially many vertices and edges.

This provides a pretty good model for what can happen in propositional proofs. For straight propositional logic one has examples as in [14], which involve phenomena of the type which appear in the example above. A further analysis is given in [6], in a way which brings out the natural *oscillations* in these examples through the notion of "almost cyclic structures". In a proof without cuts an oriented path in the logical flow graph can go "up" in the proof and then "down" again, but afterwards it has no opportunities to turn up again. To do so would represent a kind of "oscillation" in the graph, and to understand exactly which kind of oscillations matter for complexity is very much the matter at hand. The analysis developed in [6] also treats the examples of short proofs of the feasibility of large numbers, which manifest the same phenomena more directly and simply. In fact the simplest examples of short proofs of feasibility can be coded completely by optical graphs with some auxiliary data.

These considerations suggest that we *define* a new notion of feasibility more directly in terms of optical graphs (and in analogy with the notion for proofs introduced in Section 1). For this notion each vertex would have a number or other term attached to it, or rather we would think of each vertex being associated to an atomic formula $F(t)$. Each vertex would have to be obtained from the "preceding" ones (as defined through the orientation) using an allowable rule. Let us call a vertex p in an optical graph *focussing* if it has two edges oriented towards it and *defocussing* if it has two edges oriented away from it. The binary rules which say that sums and products of feasible numbers are feasible would correspond to focussing branch points, while the unary rule to the effect that the successor of a feasible number is feasible would not require a branch point. Defocussing branch points would reflect the fact that once the feasibility of a particular term is established one may use that information twice. A vertex with no edges going into it would represent either an assumption or an "axiom" such as the feasibility of 0. A vertex with no edge coming from it could be interpreted as a "conclusion", e.g., the feasibility of some given number. With this correspondence the proof of the feasibility of 2^{2^n} in $O(n)$ steps obtained by stringing together proofs of $F(t) \to F(t^2)$ (as mentioned in Section 1) can be described completely by an optical graph.[3]

By following these rules we can have a way to pass *from* optical graphs *to* proofs. Thus optical graphs provide a way of modelling proofs in two different directions: every proof gives rise to an optical graph, through the logical flow graph, and conversely in certain circumstances optical graphs can practically be reconverted into proofs. One should keep in mind here that the general idea of *feasibility* is quite flexible in its ability to speak about mathematical structures, as discussed in [7].

3 Visibility and cut elimination

One might wonder why we approach the study of symmetry in proofs through the notion of visibility, instead of looking directly at combinatorial models for cut elimination (see [9] for background information). Since cut-free proofs have a fairly simple combinatorial structure (see [3]), to have a clear combinatorial model for cut elimination would hopefully mean to have a good understanding of how short proofs might be built. Unfortunately local rules of cut elimination have very "unusual" combinatorial behavior, and global properties of graphs of proofs are in general *not* preserved under these local transformations. (This aspect is analyzed thoroughly in [5].)

Just as for the visibility, *contractions* play a very important role for cut elimination. Indeed, to eliminate cuts over formulas which have been contracted, one needs to *duplicate* subproofs of the original proof. This fact is independent of the method we use for cut elimination, but instead intrinsic to the combinatorics of proofs. Duplication can occur repeatedly during the process of elimination of cuts and might lead to large expansion in the size of the proof [9]. In this respect, the idea of visibility makes a pretty good model for exponential expansion as in cut elimination. Similar duplications occur there.

For the logical connections between formulas in a proof the combinatorial model based on visibility is more rough as a picture of cut elimination. Suppose

[3] Although there are two exponentials in 2^{2^n}, only one of them really counts. The other comes for free since we allow a multiplication rule for feasibility. That is, we can get a proof of the feasibility of 2^n in $O(n)$ lines just by multiplying a bunch of 2's together, without using contractions or cuts. By using them we win an additional exponential.

that we are given a proof for which G is the logical flow graph. Let v be a vertex in G, representing an occurrence of an atomic formula in the proof. If that atomic formula lies in the endsequent of the proof, then it makes sense to try to compare the visibility $\mathcal{V}(v, G)$ of v (positively or negatively oriented, as appropriate) with what happens in the proof after cut elimination. That is, if Π is the original proof and Π' is obtained from Π by cut elimination, then Π and Π' have the same endsequent, and so we can view v as being a vertex in each of the logical flow graphs G and G'. One might think that $\mathcal{V}(v, G)$ makes a reasonable model for what v can "see" inside G', but this is not really what happens. The visibility is much more homogeneous. Whatever opportunities an (oriented) path has in G persist in the visibility. This is *not* preserved in general by any procedure of cut elimination, because of the way that paths can be split [5]. Imagine a vertex w of G which is visible to v and lies inside the proof Π. It may have several natural "relatives" in G'. Each of these may see much less in G' than what w itself sees in G, even if the total of what the relatives of w see in G' is comparable to what w sees in G. Now w will also have a different set of "relatives" in $\mathcal{V}(v, G)$, but each of the relatives of w in $\mathcal{V}(v, G)$ will see inside $\mathcal{V}(v, G)$ about the same as what w sees inside G. (The point here is that a connected component of G might split under cut elimination into several disconnected subgraphs. As a consequence, different "relatives" of w in G' might lie in different connected components of G' and therefore "see" only what lies in a specific component.) In other words, the visibility always has a certain kind of homogeneity that cut elimination does not typically preserve.

From the above one concludes that the visibility gives a more natural combinatorial model for recognizing and relating symmetric patterns in proofs. The "partial" view that one has after applying cut elimination corresponds to having only "parts" of patterns lying inside a cut-free proof. The problem of recognizing patterns from their "partial" representation describes well the difficulty and limitations in the analysis of cut elimination.

To conclude let us remark that in the discussion above, if we simply start with a vertex v in G which does not correspond to an atomic formula in the endsequent, then we experience similar problems from the start.

4 Implicit descriptions and explicit constructions

Let us start over now and think about optical graphs as a mathematical notion in its own right.

Our thinking about optical graphs will parallel our initial discussion of proofs. We can think of an optical graph as being related to an algorithm, for instance. The optical graph, together perhaps with auxiliary data, might code an implicit construction of some object, as discussed in Section 2.

How does one convert the implicit description into an explicit construction from primitive pieces? For optical graphs this is achieved by passing to the *visibility*[4]. It is easy to see why this is true. Imagine building an actual physical

[4] Strictly speaking one should use the "negatively oriented" version of the visibility

house, where one cannot simply say "I have constructed a room, and therefore I have two rooms". This explains *how* to build a house, but for the physical building one has to produce everything. If one needs 100 bricks, one has to produce 100 actual physical bricks. Similarly, the word metric in a finitely generated group measures the amount of material needed to build the physical group element, while metrics based on feasibility (as in [7]) make measurements based on the length of the "explanations" required. For free groups the metrics based on feasibility are already quite different from the word metric, in the way that they treat cyclic subgroups. Optical graphs provide another way to measure the size of the explanation, while passing to the visibility corresponds to passing to the construction of the word metric.

This leads one to question how the visibility behaves, what kinds of expansions can occur, etc. These are analyzed in some detail in [8]. One of the first points is that the visibility never has more than exponential growth. Exponential growth can occur, as in the example described in Section 2. A more subtle point is that configurations like those in the example are actually *necessary* for exponential expansion. Roughly speaking, if N is the number of vertices in the optical graph G, if the visibility is finite, and if n is the length of the longest configuration "like" the example in Section 2 to be found in G, then the size of the visibility is at least 2^n and at most $(2N)^n$. See [8] for a precise statement. The lower bound is about right for the example in Section 2, but there are more complicated examples which are closer to the upper bound. In the latter there are many configurations like those in Section 2 interlacing each other. These examples can also be represented through short proofs of feasibility of large numbers.

In summary, [8] provides a fairly precise analysis of the way that the visibility can behave in terms of size, and we refer to [8] for more information. This analysis has the pleasant features that it is purely geometric and fits fairly well with what one might expect from the perspective of proofs.

5 The universal feasibility construction

We have discussed before how optical graphs can sometimes be used to code completely a proof of feasibility. With a little bit of flexibility one can in fact say that this is *always* true. That is, an optical graph can *always* be viewed as representing the "feasibility" of its own visibility graph.

If we want this to match perfectly with what we did before we should sort out the technical problem of finding a way to make a formal theory about oriented rooted trees, in such a way that an optical graph can be converted into a formal proof of its feasibility. We shall not pursue this here. Instead we take a broader view, towards the idea of implicit descriptions of mathematical objects in general, and their connection with questions of complexity and algorithms. Formal proofs and optical graphs both provide models for a general question which has a life of its own.

---

here to be consistent with our other definitions, but this is not a serious point.

This perspective is simple but conceptually quite striking. It says that whatever phenomena we see on the geometric side should be "accepted" as part of the general story, rather than as an artifact of the model. Some particular phenomenon may not be relevant for the particular case of feasibility of numbers, or of elements of finitely generated groups, but it is always relevant for the general idea of feasibility, because the visibility is always an example for feasibility.

In the beginning of this paper we mentioned various questions about proofs and their relations to algorithms and implicit descriptions, and then we went on to optical graphs as combinatorial models for proofs. We should return now to some of the questions of Section 1 but for optical graphs instead of proofs.

6 Symmetry and oscillations

We mentioned before the idea that if one has a short proof describing a "large" object then that should imply the presence of a substantial amount of symmetry. For proofs in general it is not at all clear how to formulate this principle in a precise way.

For our model the question becomes more concrete. Given an optical graph G and a vertex v in G, let $\mathcal{V} = \mathcal{V}_+(v, G)$ be the visibility of G starting from v. Assume for simplicity that $\mathcal{V}$ is finite, so that its size is at most exponential in the size of G, as discussed in [8]. It may be much smaller than exponential, but if it is of exponential size then our principle would say that $\mathcal{V}$ should have a lot of symmetry.

To put the matter into perspective let us think first about universal covering surfaces from topology, for which the visibility is a kind of analogue. (See [1, 12] for information about covering spaces and the universal covering.) In this case one starts with a topological space X and a basepoint b, and one looks at all paths in X which start at b. (Assume that X is pathwise connected to avoid pathologies.) Two such paths are considered *equivalent* if they have the same endpoint and if there is a continuous deformation from one to the other that keeps the endpoints fixed. The set $\widetilde{X}$ of all such equivalence classes is called the *universal covering space* of X, and it carries a natural topology. If one takes X to be a circle, for instance, this construction gives back the real line.

For the universal covering space there is a natural group action, an action by the *fundamental group* of X, which is the group of all loops based at b modulo the equivalence relation of continuous deformation again. The quotient of the universal covering $\widetilde{X}$ by this group action exactly gives X back in a simple and natural way. This is not deep or mysterious, just a part of the definitions.

Thus for universal covering spaces there is a kind of perfect version of our principle, the growth in size of the universal covering as compared to the original space is exactly matched by the size of the covering group. If there is a natural volume measure for X, for instance, then there will be a natural volume measure for $\widetilde{X}$ too, and the ratio of the total volume of $\widetilde{X}$ to the total volume of X will be exactly the order of the covering group.

For the visibility of an optical graph there is some symmetry but not quite as much. This stems in part from the fact that we restrict ourselves to *oriented* paths to define the visibility. This can lead to a substantial difference between going "forward" and "backward", in such a way that the range of future choices diminishes with each step.

If an oriented path beginning at the initial vertex v returns to v, then all the future choices that were possible before are possible again. This situation is reminiscent of universal coverings, but it is not quite the same, since for universal coverings one can simply reverse the path, while for the visibility that possibility is prevented by the orientations.

Even if oriented paths from v in our graph G do not return to v, as in the example in Section 2, one can still see a lot of symmetry for the visibility in that case. In general one can often expect to see a fair amount of symmetry in the following manner. A vertex in the visibility $\mathcal{V}_+(v, G)$ is an oriented path α in G which begins at v. If w is the endpoint of α, then the visibility $\mathcal{V}_+(w, G)$ of G beginning at w lies naturally within $\mathcal{V}_+(v, G)$. Indeed, a vertex in $\mathcal{V}_+(w, G)$ corresponds to an oriented path that begins at w, and that can be added to the end of α to give a vertex in $\mathcal{V}_+(v, G)$. This construction also respects edges and orientations. Now if the visibility $\mathcal{V}_+(v, G)$ is large compared to the size of G, then there will be vertices w in G which correspond to many different α's in $\mathcal{V}_+(v, G)$. For each of these there will be a copy of $\mathcal{V}_+(w, G)$ in $\mathcal{V}_+(v, G)$, and all of these copies will look exactly the same.

In this way the visibility will have a lot of symmetry when it is large, even if the symmetry is not as simple as for covering spaces, and not as simple as having a group action. See [8] for more precise information about the structure and symmetry of visibility graphs.

7 Intermediate symmetries

As in Section 1 we would like to have ways to *compare* optical graphs and their visibilities, to see patterns, to find symmetry, etc. To this end we can use the language of *mappings* between them. This means mappings from vertices to vertices and from edges to edges with the obvious compatibility conditions. We shall restrict our attention to mappings which respect orientations.

A very simple way to compare graphs is to look for *embeddings* between them, injective mappings. The family of examples in Section 2 (parameterized by the number n of successive loops) can be embedded one into the other in simple ways. Embeddings like these are compatible with interpretations of feasibility along the lines presented in Section 2.

A basic point about mappings between optical graphs is that they induce mappings between visibilities. That is, if G and H are optical graphs, $f : G \to H$ is a mapping (which, as above, we ask to respect orientations), and if $v \in G$ and $w \in H$ are vertices which satisfy $f(v) = w$, then f lifts to a mapping F between the visibilities $\mathcal{V}_+(v, G)$ and $\mathcal{V}_+(w, H)$ in a natural way. That is, an oriented path in G starting from v is mapped by f to an oriented path in H starting from w, and so we get a mapping from the vertices of $\mathcal{V}_+(v, G)$ to the vertices of $\mathcal{V}_+(w, H)$. Similarly f induces a mapping from edges in $\mathcal{V}_+(v, G)$ to edges in $\mathcal{V}_+(w, H)$, and the resulting $F : \mathcal{V}_+(v, G) \to \mathcal{V}_+(w, H)$ preserves orientations.

In addition to *embeddings* between optical graphs one can look at mappings which might look more like "projections". For this the theory of covering surfaces provides a good model; before we were interested in universal coverings as an analogy with visibilities, now we look to covering spaces in general for inspiration. Again [1, 12] provide good references for background information, and for the definitions and basic observations that we are now going to emulate. Note that the orientations of paths and graphs in our context changes the story somewhat from the classical one.

Mappings more like projections are very natural for looking for intermediate symmetries within proofs, for understanding better what it means to say that a proof can be shortened. Being able to make a proof shorter is like being able to project an optical graph into a smaller one, thereby making manifest some of its internal symmetry. One would like to find criteria for shortening proofs, or to exhibit "counterexamples" of families of statements which admit *only* large optical graphs as proofs.

To make these ideas more precise we need to introduce some definitions. Let G and H be optical graphs, and let $f : G \to H$ be a mapping which preserves orientations. We say that f is *locally one-to-one* if for each vertex $u \in G$ and each pair of distinct edges a, b attached to u we have that a and b are sent to distinct edges in H by f. Note that locally one-to-one mappings need not be globally injective, but may instead behave like some kind of projection.

This concept is very useful for comparing optical graphs and seeing patterns. Consider the situation where Z is some enormous optical graph, G is a much smaller one, and $\phi : Z \to G$ preserves orientations and is locally one-to-one. This is exactly a way of saying that Z has many repetitions in its behavior, which is constrained by the mapping into G. Imagine Z being "projected" into G, or wrapped around it.

Let us say that f is *$+$-complete* if it satisfies the following condition. Let u be any vertex in G, and suppose that e is an edge in H attached to $f(u)$ which is oriented *away* from u. Then we ask that there be an edge a in G attached to u and oriented away from u which is mapped to e by f.

One can think of $+$-completeness as a *local surjectivity* condition, but only for the edges oriented *away* from the given vertex. This definition differs from the customary theory for covering surfaces and reflects the role of orientation in our story; we are putting more importance on where something goes than where it came from.

We shall think of mappings which are both locally one-to-one and $+$-complete as being analogous to covering maps in topology. (Actually they are like "complete" covering maps, our finite situation enjoying an extra compactness which is not automatic in the topological setting.)

A fundamental example of such mappings comes from the visibility. That is, let G be an optical graph, let v be a vertex in G, and let $\pi : \mathcal{V}_+(v, G) \to G$ be the canonical projection, defined as follows. Each vertex s in $\mathcal{V}_+(v,G)$ represents an oriented path in G which begins at v, and $\pi(s)$ is simply the endpoint of this path. One can define π similarly for edges and check that it preserves orientations. (This mapping is discussed further in [8].) It is easy to see that π is locally one-to-one and $+$-complete.

This fits with the idea that the visibility is analogous to the universal covering surface, but the point now is that we can talk about "covering spaces" in general. Let us describe a bit of the analogous theory in our situation (which is much simpler than the usual one).

Suppose that G and H are both optical graphs, that $v \in G$ and $w \in H$ are vertices, and that $f : G \to H$ is a mapping which preserves orientations. Assume also that f is locally one-to-one and $+$-complete, and that it satisfies $f(v) = w$. As above f induces a mapping $F : \mathcal{V}_+(v,G) \to \mathcal{V}_+(w,H)$ which preserves orientations. Under our present assumptions this mapping F is actually an *isomorphism*. Let us sketch the argument. Suppose that s and t are vertices in $\mathcal{V}_+(v,G)$ which are mapped to the same vertex in $\mathcal{V}_+(w,H)$. This means that s and t represent distinct oriented paths in G which start at v and which are mapped to the same path in H by f. Since s and t are *distinct* paths they may agree for some time after v but they must diverge somewhere. The parts that agree will of course have the same image in H. At the moment where they diverge one must also have a divergence in H; this follows from the local injectivity of f. This implies that the images of s and t are not the same path in H, and that $F : \mathcal{V}_+(v,G) \to \mathcal{V}_+(w,H)$ is injective on vertices. It is easy to see that F is locally injective in the sense described above, and this implies that F must be injective on *edges* as well.

For surjectivity we use $+$-completeness. We assume that we are given a vertex in $\mathcal{V}_+(w,H)$, and we want to show that it is the image of a vertex in $\mathcal{V}_+(v,G)$ under F. This is the same as saying that any oriented path in H which begins at w should be the image under f of an oriented path in G which begins at v. This is easy to prove, by induction on the length of the given path in H, for instance. That is, $+$-completeness is precisely the condition which ensures that if you can lift a path in H to one in G up to a certain point, then you can go one step further in G if there is another step to take in H. This gives surjectivity of F on vertices, and it is easy to show surjectivity on edges as well.

To summarize, F is injective when f is locally injective, F is surjective when f is $+$-complete, and the two assumptions on f imply that F is an isomorphism.

This is all very nice. It provides a way to say that the structure of G lies between the structure of H and the structure of the common visibility. Think of the visibility as being "larger", while H might well be "smaller" than G, with f somehow coiling G up inside H. The visibility disentangles all the paths in H while G only disentangles some of them. Mathematically, some of these ideas are represented by the formula

$$\pi_H = f \circ \pi_G \circ F^{-1}. \tag{1}$$

Here $\pi_H : \mathcal{V}_+(w,H) \to H$ and $\pi_G : \mathcal{V}_+(v,G) \to G$ denote the canonical projections as defined above. We can think of $\pi_H : \mathcal{V}_+(w,H) \to H$ as representing the way that the visibility of H is coiled up inside H. It represents much of the structure of H, the way that paths are coiled up in H. The formula says that we can compute it as follows. We first use F to identify the visibilities $\mathcal{V}_+(w, H)$ and $\mathcal{V}_+(v, G)$. Then we proceed in two steps, first applying π_G, which represents the coiling of paths in G, and then we apply f, which may wrap G up inside H.

In short we are saying that the way that paths are coiled up in H can be "factored" into two contributions, reflecting the coiling of paths in G and the coiling of G in H. Thus G indeed represents a kind of "intermediate" structure.

To help us to interpret this it is useful to go back to the case of covering spaces. In topology we can think of a surface X as providing a geometric representation of a certain group, the fundamental group. In the universal covering $\widetilde{X}$ of X all the paths are "unwrapped".

If $f : Y \to X$ is a (complete) covering of X by another surface Y, then f and Y represent a partial unwrapping of the curves in X. One can show that the canonical projection from the universal covering $\widetilde{X}$ to X can be factored into a mapping from $\widetilde{X}$ to Y followed by $f : Y \to X$. Indeed, $\widetilde{X}$ is actually isomorphic to the universal covering of Y, and the mapping from $\widetilde{X}$ to Y just mentioned is equivalent to the canonical projection.

For covering surfaces the fundamental group of the intermediate space Y can be seen as a subgroup of the fundamental group of X. Moreover, Y is essentially characterized in this manner, and all subgroups of the fundamental group arise in this way.

Thus we might think of a surface as representing a group in a geometric way, and then (complete) coverings correspond exactly to subgroups of this group. We can forget about groups and work entirely within this geometric language of covering spaces.

In our situation we do not have such a good connection with groups. We have "broken" some of the symmetry as discussed in Section 6, but we still have a reasonable version of the geometric language. With this geometric language we can still make comparisons and measurements of patterns even if it is not quite so "regular" as in the case of covering spaces where the story admits simple descriptions through groups and group actions on topological spaces. We still have patterns, even if they are not so simple or regular.

In short we trade groups for combinatorial geometry. We broke the symmetry and lost the groups, but we gained extra flexibility in our combinatorial geometry.

References

1. L. Ahlfors and L. Sario. *Riemann Surfaces.* Princeton University Press, 1960.
2. S. R. Buss. The undecidability of k-provability. *Annals of Pure and Applied Logic*, 53:75–102, 1991.
3. A. Carbone. Some combinatorics behind proofs. Submitted for publication, 1995.
4. A. Carbone. Cycling in proofs, feasibility and no speed-up for nonstandard arithmetic. IHES preprint M/96/55, 1996.
5. A. Carbone. Interpolants, cut elimination and flow graphs for the propositional calculus. *Annals of Pure and Applied Logic*, 1996. To appear.

6. A. Carbone. Proofs and groups with distorted length function. Unpublished manuscript, 1996.
7. A. Carbone and S. Semmes. Looking from the inside and the outside. IHES preprint M/96/44, 1996.
8. A. Carbone and S. Semmes. Algorithms and geometry: Hidden symmetry and implicit description. Book manuscript (in preparation), 1997.
9. A. Carbone and S. Semmes. Making proofs without modus ponens: An introduction to the combinatorics and complexity of cut elimination. *Bulletin of the American Mathematical Society*, 1997. to appear.
10. S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44:36–50, 1979.
11. J.-Y. Girard. Linear Logic. *Theoretical Computer Science*, 50:1–102, 1987.
12. W. Massey. *A Basic Course in Algebraic Topology*, volume 127 of *Graduate Texts in Mathematics*. Springer-Verlag, 1960.
13. R. Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, 36:494–508, 1971.
14. R. Statman. Bounds for proof-search and speed-up in predicate calculus. *Ann. Math. Logic*, 15:225–287, 1978.