ELSEVIER

Annals of Pure and Applied Logic 100 (1999) 1–67

ANNALS OF PURE AND APPLIED LOGIC

www.elsevier.com/locate/apal

Duplication of directed graphs and exponential blow up of proofs

A. Carbone *

Mathématiques/Informatique, Université de Paris XII, 61 Avenue du Général de Gaulle, 94010 Créteil, France

Received 5 January 1998; revised 15 September 1998; accepted 26 January 1999

Communicated by J.Y. Girard

Abstract

We develop a combinatorial model to study the evolution of graphs underlying proofs during the process of cut elimination. Proofs are two-dimensional objects and differences in the behavior of their cut elimination can often be accounted for by differences in their two-dimensional structure. Our purpose is to determine geometrical conditions on the graphs of proofs to explain the expansion of the size of proofs after cut elimination. We will be concerned with exponential expansion and we give upper and lower bounds which depend on the geometry of the graphs. The lower bound is computed passing through the notion of universal covering for directed graphs.

In this paper we present ground material for the study of cut elimination and structure of proofs in purely combinatorial terms. We develop a theory of duplication for *directed graphs* and derive results on graphs of proofs as corollaries. © 1999 Elsevier Science B.V. All rights reserved.

MSC: 03F07; 03F20; 03F03; 03F05; 05C20; 05C38; 68R10; 68R05

Keywords: Directed graphs; Optical graphs; Logical flow graphs; Cut elimination; Cycles; Structure of proofs; Proof complexity; Duplication; Visibility graphs

1. Introduction

The reasons to study the relations between the geometry of proofs and the expansion of a proof after cut elimination are several and they come from different directions. The order we will present them does not correspond to their importance. We need just to start somewhere.

* Fax: +33-01-45171649.

E-mail address: ale@univ-paris12.fr, carbone@ihes.fr (A. Carbone)

PII: S0168-0072(99)00009-3

Let us look first at complexity theory. It is known that the existence of a proof system for propositional logic where classical tautologies can be proved by polynomial size proofs in the size of the tautology would imply $\mathrm{NP} = \mathrm{CO}-\mathrm{NP}$ [11]. There has been a lot of work done to exhibit families of tautologies with only "large" proofs (i.e. proofs which are of exponential size in the size of the tautology) for the various layers of the hierarchy of propositional proof systems [20]. Positive answers have been obtained for the layers corresponding to proofs which present very simple combinatorial structure (e.g. resolution proofs, cutting plane proofs, cut-free proofs). The first system in the hierarchy that presents proofs with complicated graph structure is the *sequent calculus LK*, which is characterized by the presence of *cut* and *contraction* rules. Here it seems a particularly difficult task to find hard tautologies. A deeper understanding of the combinatorics of proofs might be of help for the search of such examples. See [3, 9] for a discussion and some related results, and [20] for a detailed account of the most recent work done in the area of proof complexity.

There is a well-known link between proofs with and without cuts. It is the Gentzen Cut Elimination Theorem [12] which says that any proof in the sequent calculus *with* cuts can always be transformed in some effective way into a proof *without* cuts. This theorem furnishes an algorithm to formally convert proofs formalized in *LK* into proofs formalized in a system (i.e. *LK* without the cut rule) which lies in the lowest levels of the hierarchy and for which we already have examples of tautologies which are "hard" to prove (a finite version of the Pigeon Hole Principle is one of them [15]).

The known examples for the cut-free sequent calculus are nevertheless very few, and there are no *uniform* proofs of exponential lower bounds on the size of their deductions. One would like to analyze the entire set of hard tautologies and a possible approach might be to look at the amount of *symmetry* lying in the tautologies and in their proofs. It seems plausible that hard tautologies for the sequent calculus *with* cuts would present very little symmetry. Roughly speaking, a tautology of such a type should represent in a *concise* way a state of "chaos" in the proof. Its proofs should look like a complete search in truth tables.

In general one might like to believe that

> *The symmetry present in a tautological statement reflects the symmetry in its proofs, as well as the symmetry present in a mathematical object reflects the symmetry in its construction.*

It is very difficult to argue anything precise here. One can say that there is a lot of evidence supporting this principle in mathematics. (This is discussed at length in [7].) On the other hand, this principle would suggest that the existence of short proofs for some propositional proof system would reflect a kind of *universal symmetry* underlying all propositional tautologies and this might seem unlikely. Indeed $\mathrm{NP} = \mathrm{CO}-\mathrm{NP}$ is considered unlikely as well.

Let us go back to cut elimination. In most cases, the procedure of cut elimination unfolds a proof into a new one having much simpler structure but much larger size. The cut-free proof will turn out to be formed by many building blocks which are copies of the same subproof. These identical components might be formed by several copies of some other building blocks, and so on. In other words, *symmetry* is found at repeated levels. The upper bound in Theorem 31 reflects well this idea and shows how patterns lying in cut-free proofs might be recoverable from the graph of the original proof *with* cuts. In Sections 5 and 13 we analyze how patterns in proofs evolve through cut elimination and which are the combinatorial structures of proofs that might induce an exponential blow-up after cut elimination. We give examples of proofs with cuts where these patterns appear explicitly and examples where they appear in more subtle ways.

How the symmetry of a statement is related to the internal symmetry of its proofs? Given a proof, how can we reduce it to a smaller proof by exploiting the symmetricity of its subparts? These are fundamental questions in automated deduction and answers to them (even partial answers) might open up new ways to the creation of alternative proof systems which are based on different underlying combinatorics. Fixing a proof system, we fix also the combinatorial structure of the proofs we construct and as a consequence we are able to generate specific structural patterns. We would like to think that the power of a deduction system depends on how complex its structural patterns are. But where this overlook can lead to? Suppose for a moment to have some informal language where proofs can be constructed and to know for this setting what it means for a proof to be short. (Afterall we do this all the time since we hardly think of a mathematical proof in formal terms.) It is plausible to think that short proofs of hard theorems (in our informal language) exhibit specific structural patterns and that these patterns might vary from theorem to theorem. If we consider now *formal* systems of deduction, we might see that they forbid the creation of some of these patterns, and if so we will never be able in a feasible time to show inside them some of their theorems. At the moment we know a relatively small number of different proof systems (this holds for both propositional and predicate logic) and it is by no means the case that these proof systems should be the only interesting ones.

Another aspect which deserves to be investigated concerns the constraints of purely *logical* nature that the rules of inference of a calculus impose on the graphs of proofs they generate. Along this line, Theorem 51 says that not all topological structures of graphs can be realized by the logical system *LK* and Theorem 54 shows how graphs of proofs in the sequent calculus cannot contain a connected component which is a simple cycle (i.e. a sequence of edges starting and ending in the same vertex). (More general results of this nature are given in Section 8.) It is plausible that *no* logical systems can produce proofs whose graphs contain simple cycles and one could ask what are the combinatorial structures allowed by *logic*. A discussion on this and other related points is developed at the end of Section 11.

The relations between proofs and other existing models of computation (more or less expressive, as for instance circuits and automata) are not at all understood. About whether or not proofs are comparable as combinatorial objects to circuits, we will show (see Theorem 50) that any circuit can be simulated by a proof with atomic cuts and no quantifiers in such a way that the graph underlying the circuit and the graph underlying the proof are the same. The procedure of cut elimination corresponds combinatorially to the transformation of circuits into boolean expressions.

The plan of the paper goes as follows. Section 2 will contain a quick review of known concepts. We present the rules of the sequent calculus *LK*, we introduce Gentzen Cut Elimination Theorem and we define the notion of logical flow graph. Section 3 illustrates the ideas developed in the paper through a concrete example. Section 4 introduces the notion of *optical graph* and the combinatorial operation of *duplication* on optical graphs. These are two fundamental concepts for this paper. The first notion defines a class of graphs that includes graphs of proofs and the latter is used to describe the 'topological' changes of the graph of a proof during the process of cut elimination. In Section 5 we present several examples of evolution of optical graphs through repeated duplication. They are examples of linear as well as polynomial (of any degree) and exponential growth. Sections 6 and 7 discuss basic properties of duplication. In Section 8 we study the effect of duplication on cycles in an optical graphs. We show that not always cycles in a graph can be disrupted by duplication and we give conditions under which this happens. In Section 9 we focus our attention to a specific strategy of duplication and discuss its properties. We compute upper and lower bounds of the size of the expansion and will show how these bounds depend on the geometry of the starting graph. This strategy is 'natural' and we give concrete examples of proofs where it applies. The notions of *visibility* graph and *focal pair* of branching points are introduced here and will turn out to be crucial notions for the computation of the lower bound. In Section 9.3 we give an exponential upper bound for the visibility graph improving a result obtained in [10]. In Section 11 we relate optical graphs to formal proofs and we prove that given an acyclic graph there is always a proof with the same underlying structure (Theorem 49). Theorem 50, which we discussed above, is a consequence of this latter result. In Sections 12 and 13 we revisit cut elimination through our combinatorial model, discuss the effect of duplication on chains of focal pairs, and analyze the creation of patterns in the graphs of proofs. The paper ends with the definition of a new notion of graphs of proofs (a refinement of the notion of logical flow graph) and two conjectures which relate in a precise way exponential expansion to geometrical properties of proofs with cuts.

The results on this paper are general and obtained for *directed graphs*. In particular the notion of visibility graph formalizes the idea of "universal covering" applied to directed graphs. The results on graphs of proofs are obtained as corollaries of more general statements.

2. The sequent calculus

In this section we quickly recall known concepts. We present the rules of the *sequent calculus LK* [14, 24], we introduce Gentzen Cut Elimination Theorem [12, 14, 24] and define what is a logical flow graph of a proof [1].

For *LK* the *axioms* are sequents of the form

$$A, \Gamma \to \Delta, A$$

where A is any formula and Γ, Δ are any collections of formulas (where we allow multiple occurrences of the same formula). A *proof* of a particular sequent consists of derivations from axioms through rules of inference, namely the *logical rules* and *structural rules*. To describe these rules we write Γ, Γ_1, Γ_2, Δ_1, Δ_2, for collections of formulas, and we write $\Gamma_{1,2}$ as a shorthand for the combination of Γ_1 and Γ_2 (counting multiplicities).

The *logical rules* are used to introduce connectives, and they are given as follows:

$$\neg : \textit{left} \quad \frac{\Gamma \to \Delta, A}{\neg A, \Gamma \to \Delta} \qquad \neg : \textit{right} \quad \frac{A, \Gamma \to \Delta}{\Gamma \to \Delta, \neg A}$$

$$\wedge : \textit{right} \quad \frac{\Gamma_1 \to \Delta_1, A \quad \Gamma_2 \to \Delta_2, B}{\Gamma_{1,2} \to \Delta_{1,2}, A \wedge B}$$

$$\wedge : \textit{left} \quad \frac{A, B, \Gamma \to \Delta}{A \wedge B, \Gamma \to \Delta}$$

$$\vee : \textit{left} \quad \frac{A, \Gamma_1 \to \Delta_1 \quad B, \Gamma_2 \to \Delta_2}{A \vee B, \Gamma_{1,2} \to \Delta_{1,2}}$$

$$\vee : \textit{right} \quad \frac{\Gamma \to \Delta, A, B}{\Gamma \to \Delta, A \vee B}$$

$$\supset : \textit{left} \quad \frac{\Gamma_1 \to \Delta_1, A \quad B, \Gamma_2 \to \Delta_2}{A \supset B, \Gamma_{1,2} \to \Delta_{1,2}}$$

$$\supset : \textit{right} \quad \frac{A, \Gamma \to \Delta, B}{\Gamma \to \Delta, A \supset B}$$

$$\exists : \textit{left} \quad \frac{A(b), \Gamma \to \Delta}{(\exists x)A(x), \Gamma \to \Delta} \qquad \exists : \textit{right} \quad \frac{\Gamma \to \Delta, A(t)}{\Gamma \to \Delta, (\exists x)A(x)}$$

$$\forall : \textit{left} \quad \frac{A(t), \Gamma \to \Delta}{(\forall x)A(x), \Gamma \to \Delta} \qquad \forall : \textit{right} \quad \frac{\Gamma \to \Delta, A(b)}{\Gamma \to \Delta, (\forall x)A(x)}$$

The *structural rules* do not involve connectives and are the following:

$$\textit{Cut} \quad \frac{\Gamma_1 \to \Delta_1, A \quad A, \Gamma_2 \to \Delta_2}{\Gamma_{1,2} \to \Delta_{1,2}}$$

$$\textit{Contraction} \quad \frac{\Gamma \to \Delta, A, A}{\Gamma \to \Delta, A} \qquad \frac{A, A, \Gamma \to \Delta}{A, \Gamma \to \Delta}$$

The rules for the quantifiers can be used provided some requirements. In $\exists$: *right* and $\forall$: *left*, any term t is allowed which does not include a variable which lies already within the scope of a quantifier in the given formula A. In $\exists$: *left* and $\forall$: *right* the "eigenvariable" b should not occur free in Γ, Δ.

This is the system *LK* for classical predicate logic. For propositional logic it is the same except that one drops the quantifier rules. In the usual formulation of *LK* one has two more structural rules, namely the *permutation* rule and the *weakening* rule. The first permutes the position of formulas in a sequent. Since we defined a sequent through collections instead of sequences of formulas we do not really need this rule.

The *weakening* rule adds formulas to a sequent during the derivation. By defining axioms as above we assume that weak formulas are all introduced from the beginning of the proof.

A proof in *LK* is a *binary tree* of sequents, where each occurrence of a sequent in the proof can be used at most once as premise of a rule. The root of the tree is labelled by the theorem, its leaves are labelled by axioms and its internal nodes are sequents derived from one or two sequents (which are labels for the antecedents of the node in the tree) through the rules of *LK*.

The system *LK* has very nice combinatorial properties. The formulas never simplify in the course of a proof; they can disappear through the cut rule, and their repetitions can be reduced through the contraction rule, but they cannot be "decomposed" directly. One can do this in effect through the cut rule however.

Before concluding this introductory part let us give some more notation and a few definitions which we will use in the course of the paper. The *size* $|A|$ of a formula A is its number of symbols. The size $|S|$ of a sequent S is the sum of the sizes of its formulas. The size $|\Pi|$ of a proof Π is the sum of the sizes of its sequents.

A *weak connective* is either an $\wedge$ occurring negatively in a formula or an $\vee$ occurring positively. A *strong connective* is either an $\wedge$ occurring positively or an $\vee$ occurring negatively.

In the following we will frequently use the notion of *occurrence* of a formula in a proof as compared to the formula itself which may occur many times.

2.1. Cut elimination

In 1934 Gentzen ([12]; see also [14, 24]) proved the following

Theorem 1 (Gentzen). *Any proof in LK can be effectively transformed into a proof which never uses the cut rule. This works for both propositional and predicate logic.*

This is a striking result, particularly in view of its combinatorial consequences for formal proofs. In a proof without cuts there is no way to simplify formulas. In fact every formula which appears in a proof without cuts also occurs as a subformula of a formula in the end-sequent. This fact is known as *subformula property*. (One should be careful about the notion of a *subformula* in the presence of quantifiers, which can have the effect of changing the terms within.) One of the effects of cut-elimination is the simplification of the dynamical processes which can occur within proofs. Predicate logic gives a way to code substitutions which may not be expressed explicitly, and an effect of cut-elimination is to make all the substitutions explicit. A related point is that proofs with cuts can have *oriented cycles*, while proofs without cuts cannot. See [3, 4] for more information and examples. Note that [4] provides an alternative proof system in which there are no (oriented) cycles and in which it is easier to make explicit the construction of terms which lies below a proof.

The 'price' of cut-elimination is that the cut-free proof may have to be much larger than proofs without cuts. There are propositional tautologies for which cut-free proofs must be exponentially larger than proofs with cuts, and in predicate logic the expansion can be non-elementary. See [16, 17, 21–23, 25].

The *contraction rule* plays a key role in this expansion. While it may seem harmless enough, it can be very powerful in connection with the cut rule. Imagine that you have some piece of information which is represented by a formula A which you can prove, but that in your reasoning you actually need to use this piece of information twice. By using a contraction rule (on the left hand side of the sequent arrow) and a cut you can code your argument in such a way that you only need to verify A once. On the other hand, a cut-free proof represents 'direct' reasoning, where lemmas are not allowed, and in practice this forces one to duplicate the proof of A. (See [9] for more details.)

Thus the cut rule provides a mechanism by which the contraction rule can have the effect of a "duplication" rule. It is this point that it was aimed for in the geometric models introduced in [10] and analyzed thoroughly in this paper.

Let us look more closely at how the procedure introduced by Gentzen works. There are distinguished cases and different recipes that we need to follow depending on the structure of the cut formula and whether it came from a contraction. The idea is to push the cuts up towards the axioms and eliminate them afterwards.

Let us begin by considering the case of a cut applied over a formula which comes directly from an axiom, either as a distinguished occurrence or as a weak occurrence. Consider first the situation where the cut formula comes from a distinguished occurrence in an axiom, as in the following:

$$\frac{\Gamma_1, A \to A, \Delta_1 \quad A, \Gamma_2 \overset{\Pi_*}{\to} \Delta_2}{\Gamma_1, A, \Gamma_2 \to \Delta_1, \Delta_2} \tag{1}$$

In this case we can remove the axiom from the proof and simply add the weak occurrences in Γ_1 and Δ_1 to the subproof Π_* without trouble, thereby obtaining a new proof of the sequent $\Gamma_1, A, \Gamma_2 \to \Delta_1, \Delta_2$ in which the last cut has been eliminated.

Suppose instead that we have a cut over a formula which comes from a weak occurrence in an axiom, as in the following situation

$$\frac{\Gamma_1, A \to A, \Delta_1, C \quad C, \Gamma_2 \overset{\Pi_0}{\to} \Delta_2}{\Gamma_1, A, \Gamma_2 \to A, \Delta_1, \Delta_2} \tag{2}$$

To eliminate the cut one can simply eliminate the subproof Π_0, take out the (weak) occurrence of C in the axiom, and add Γ_2 and Δ_2 to the axiom as weak occurrences. In other words, the sequent

$$\Gamma_1, \Gamma_2, A \to A, \Delta_1, \Delta_2 \tag{3}$$

is itself an axiom already. By doing this one removes a possibly large part of the proof.

If the two cut-formulas have been introduced by logical rules the procedure will substitute the cut with two new ones of smaller logical complexity. The idea is illustrated in the following proof:

$$\dfrac{\dfrac{\begin{matrix}\Pi_1'\\ \Gamma_1 \to \Delta_1, A, B\end{matrix}}{\Gamma_1 \to \Delta_1, A \vee B} \qquad \dfrac{\begin{matrix}\Pi_2' \\ A, \Gamma_2 \to \Delta_2\end{matrix} \quad \begin{matrix}\Pi_3' \\ B, \Gamma_3 \to \Delta_3\end{matrix}}{A \vee B, \Gamma_{2,3} \to \Delta_{2,3}}}{\begin{matrix}\Gamma_{1,2,3} \to \Delta_{1,2,3} \\ \vdots \lambda\end{matrix}} \tag{4}$$

where the cut-elimination procedure will reduce the complexity of $A \vee B$ as follows:

$$\dfrac{\dfrac{\begin{matrix}\Pi_1'\\ \Gamma_1 \to \Delta_1, A, B\end{matrix} \quad \begin{matrix}\Pi_2' \\ A, \Gamma_2 \to \Delta_2\end{matrix}}{\Gamma_{1,2} \to \Delta_{1,2}, B} \qquad \begin{matrix}\Pi_3' \\ B, \Gamma_3 \to \Delta_3\end{matrix}}{\begin{matrix}\Gamma_{1,2,3} \to \Delta_{1,2,3} \\ \vdots \lambda\end{matrix}} \tag{5}$$

The other connectives are treated similarly. In case the cut formula is quantified the procedure proceeds in a similar manner. The following proof

$$\dfrac{\dfrac{\begin{matrix}\Pi_1\\ \Gamma_1 \to \Delta_1, A(t)\end{matrix}}{\Gamma_1 \to \Delta_1, \exists x. A(x)} \quad \dfrac{\begin{matrix}\Pi_2\\ A(a), \Gamma_2 \to \Delta_2\end{matrix}}{\exists x. A(x), \Gamma_2 \to \Delta_2}}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2} \tag{6}$$

will be transformed into

$$\dfrac{\begin{matrix}\Pi_1\\ \Gamma_1 \to \Delta_1, A(t)\end{matrix} \quad \begin{matrix}\Pi_2'\\ A(t), \Gamma_2 \to \Delta_2\end{matrix}}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2} \tag{7}$$

where the proof Π_2' is obtained by substituting all occurrences of the eigenvariable a in Π_2 with the term t.

We now consider the case of contractions. The following diagram shows the basic problem.

$$\dfrac{\begin{matrix}\Pi_1\\ \Gamma_1 \to \Delta_1, A\end{matrix} \quad \dfrac{\begin{matrix}\Pi_2\\ A^1, A^2, \Gamma_2 \to \Delta_2\end{matrix}}{A, \Gamma_2 \to \Delta_2}}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2} \tag{8}$$

That is, A^1 and A^2 denote two occurrences of the same formula A, and they are contracted into a single occurrence before the cut is applied. (The contraction could just as well be on the left, and this would be treated in the same way.) To push the cut above the contraction one duplicates the subproof Π_1 as indicated below.

$$
\dfrac{\begin{matrix}\Pi_1 \\ \Gamma_1 \to \Delta_1, A\end{matrix} \qquad \dfrac{\begin{matrix}\Pi_1 \\ \Gamma_1 \to \Delta_1, A\end{matrix} \quad \begin{matrix}\Pi_2 \\ A^1, A^2, \Gamma_2 \to \Delta_2\end{matrix}}{A^2, \Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2}}{\begin{matrix}\Gamma_1, \Gamma_1, \Gamma_2 \to \Delta_1, \Delta_1, \Delta_2 \\ \vdots \ \textit{contractions} \\ \Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2\end{matrix}} \tag{9}
$$

The steps of transformation we described are essentially all what one needs. To make them to work though, one needs to change the order of the rules in a proof. In the cases considered above, the cut rule was applied to formulas which were main formulas, i.e. their principal connective was introduced by the immediate preceding rule. This configuration is in general not there and one needs to change the order of the rules in the proof by pushing cuts upwards until a pair of cut formulas which are main formulas has been reached; this can always be done as the following diagrams illustrate:

$$
\dfrac{\begin{matrix}\Pi_1 \\ \Gamma_1 \to \Delta_1, C\end{matrix} \qquad \dfrac{\begin{matrix}\Pi_2 \\ A, C, \Gamma_2 \to \Delta_2\end{matrix} \quad \begin{matrix}\Pi_3 \\ B, \Gamma_3 \to \Delta_3\end{matrix}}{C, A \vee B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}}{A \vee B, \Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3} \tag{10}
$$

which will be transformed into

$$
\dfrac{\dfrac{\begin{matrix}\Pi_1 \\ \Gamma_1 \to \Delta_1, C\end{matrix} \quad \begin{matrix}\Pi_2 \\ C, A, \Gamma_2 \to \Delta_2\end{matrix}}{A, \Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2} \qquad \begin{matrix}\Pi_3 \\ B, \Gamma_3 \to \Delta_3\end{matrix}}{A \vee B, \Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3} \tag{11}
$$

It is important to point out a fundamental feature of cut elimination: there is no canonical way to do it. In the passage from (4) to (5) we could have chosen to cut first B and then A instead of the other way around. In the passage from (8) to (9), if both appearences of the cut formula A in (8) were obtained from contractions, then we would have a choice as to which subproof duplicate first (either Π_1 or Π_2). In principle we can have procedures of cut elimination which go on forever. Of course the point of the theorem is that one can always find a way to eliminate cuts in a finite number of steps. One can even make deterministic procedures by imposing conditions on the manner in which the transformations are carried out. See [13].

In essence the principle of the procedure is to push the cuts up higher in the proof while being careful about the notion of “progress”. In fact, we typically increase the number of cuts at each stage of the process as well as the number of contractions. In the case of contractions we have made progress in the sense that we reduced the number of contractions above the cut-formula, even though we may increase the total number of contractions by adding them below the cut. In the case of conjunctions we reduced the complexity of the cut-formula. It is not hard to exhaust the possibilities, but a complete proof requires a tedious verification of cases and we shall not provide it. See [14, 24].

2.2. The logical flow graph

To each proof in the sequent calculus we can associate a *logical flow graph* which is an oriented graph which traces the flow of formula occurrences within the proof. This concept was introduced by Buss [1], although we shall modify his definition slightly by restricting ourselves to *atomic* formulas (as in [3]). A different but related graph was introduced earlier by Girard [13].

The logical flow graph of a proof Π of a sequent S is defined as follows. For the set of *vertices* in the graph we take the set of all the occurrences of *atomic* formulas in Π. We add edges between these occurrences in the following manner. If we have an axiom

$$A, \Gamma \to \Delta, A$$

then we attach an edge between the vertices which correspond to the "same" atomic formula in A. That is, each atomic subformula of A has an occurrence within each of the occurrences of A in the axiom above, and we connect the corresponding pair of vertices by an edge. We leave undisturbed the vertices which come from Γ and Δ.

When we apply a rule from LK to one or two sequents in the proof Π we have that every atomic occurrence in the upper sequents has a counterpart in the lower sequent, except for those occurrences in the cut formulas in the cut rule. In all but the latter case we simply attach an edge between the occurrences in the upper sequents and their counterparts in the lower sequent. In the case of atomic occurrences within cut formulae we attach edges between the occurrences which occupy the same position within the cut formula.

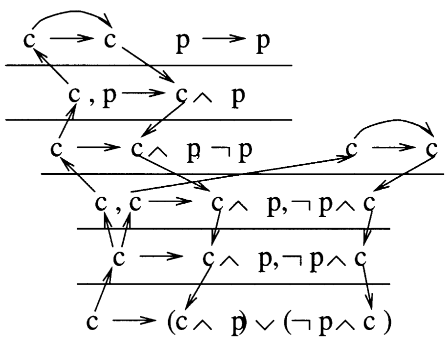
Note that there is a subtlety with the contraction rule. Each atomic occurrence in the contraction formula in the lower sequent will be connected by an edge to an atomic occurrence in *each* of the upper sequents. In all other cases there is exactly one counterpart in the upper sequents of an atomic occurrence in the lower sequent of a rule.

Thus in all cases *except* axioms, cuts, and contractions the logical flow graph simply makes a kind of "prolongation" which does not affect the topology of the graph.

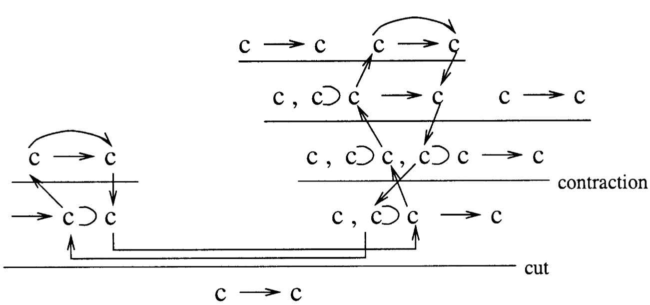
This defines the set of edges for the logical flow graph. We can define an *orientation* in the following manner. We first define the notion of the *sign* of an atomic occurrence within a sequent. One can do this inductively, but it is easier to do the following. If P is an *atomic* subformula of a formula A, then we say that P occurs *positively* in A if it lies within the scope of an *even* number of negations, and we say that it occurs *negatively* if it occurs an odd number of times within the scope of a negation. Suppose now that A is a formula which appears in a given sequent $\Gamma \to \Delta$, i.e., A appears as one of the elements of Γ or Δ. If A appears within Δ then we say that P occurs *positively* in the sequent if it occurs positively in A, and we say that P occurs *negatively* in the sequent if it occurs negatively in A. If A appears within Γ, then we do the opposite, saying that P occurs positively in the sequent if it is negative as a subformula of A, and that P occurs negatively in the sequent if it is positive in A. This takes into account the negation which is implicit in the sequent.

With this notion of *sign* of an atomic formula we can define the *orientation* for the logical flow graph as follows. If an edge in the logical flow graph comes from an axiom as above, then we orient the edge *from* negative occurrences *to* positive occurrences. If the edge comes from a cut formula, then we do the opposite and orient it *from* positive occurrences *to* negative occurrences. Otherwise the edge goes between "upper" and "lower" occurrences of an atomic formula in the application of a rule. If the occurrence is *negative* then we orient the edge *from* the lower occurrence *to* the upper occurrence, and for positive occurrences we go the other way around.

The following example illustrates the orientation of *logical paths* (i.e. sequences of consecutive oriented edges in a logical flow graph) and the way that oriented paths can split apart at contractions



Logical flow graphs of proofs with cuts can contain nontrivial *oriented cycles* (i.e. paths starting with an occurrence of a formula and going back to it; this was remarked by Sam Buss [2]), but this cannot occur for proofs without cuts as proved in [3]. In proofs without cuts oriented paths can go up and over an axiom, but once they start going down they have no chance to turn up again. This is illustrated by the example above, and one can see by comparison how the presence of a cut in the proof below permits the path to turn up again



One can think of the orientation on the logical flow graph as showing the natural flow of information in the proof. Individual oriented paths reflect the way that the "information" in a given atomic occurrence is being used in the proof. This is especially relevant for *substitutions* which can occur within a proof in predicate logic. This point emerges clearly in [4], where one sees how the absence of *oriented* cycles can enable one to track more easily the substitutions involved in the construction of terms in a proof, in such a way that one can obtain elementary estimates on the complexity of terms as a function of the size of the proof.

Notice that only the use of the contraction rule leads to branching in the logical flow graph. In the context considered in the next section there can be additional branch points in the logical flow graph, corresponding to the extra rule of inference $F{:}times$ that govern the feasibility of products of arbitrary terms. This type of branching reflects the fact that two different pieces of information are being combined into one. See [4] for more information.

Before concluding this section let us add some more terminology. A *bridge* is a path starting from a positive occurrence in the end-sequent S of a proof Π and ending in a negative occurrence in S. The starting and ending point of a bridge are called *extremes*. Notice that the proof Π is not assumed to be cut-free and this allows a bridge to pass through cuts in Π.

We say that a logical path is *direct* when it links a formula occurrence in some sequent of a proof Π to an occurrence lying in an axiom without passing through cut-edges or axiom-edges. Because of the presence of contractions in proofs, there might be *several* direct paths linking a given formula occurrence to axioms in Π.

We say that a *weak occurrence* of a formula in a proof Π is a formula whose direct paths all go to weak formulas in axioms of Π.

3. A concrete example

We are concerned with the study of the expansion of a proof under cut elimination and our main aim is to relate the geometry of the logical flow graph of a proof to the rate of growth of the expansion. We present here a basic example which illustrates the main points of our analysis.

We consider an extension of the predicate calculus LK by the rule

$$F : times \quad \frac{\Gamma_1 \to \Delta_1, F(s) \quad \Gamma_2 \to \Delta_2, F(t)}{\Gamma_{1,2} \to \Delta_{1,2}, F(s*t)}$$

where F is a unary predicate and $*$ a binary function symbol. We also add to our language the constant symbol 2. The reader might like to observe that the constructions we will present in this section and others have a natural arithmetical interpretation (the symbol 2 can be interpreted with the number 2 and the operation $*$ by multiplication) even though we will make no use of the axioms of arithmetic. In the following we will be using the usual exponential function to denote a term written down through the symbols $2, *$. We will do it freely but one should keep in mind that what we write is a short notation for a term which is made out of 2's and $*$'s and in most of cases is very large.

Let us look at a proof of $F(2) \to F(2^{2^n})$. There is no use of quantifiers and the formalization takes place on the propositional part of predicate logic. Our basic building block is given by

$$F(2^{2^{j-1}}) \to F(2^{2^j}), \tag{12}$$

which can be proved for each j in only a few steps. (One starts with two copies of the axiom $F(2^{2^j}) \to F(2^{2^j})$ and combines them with the *F-times* rule to get

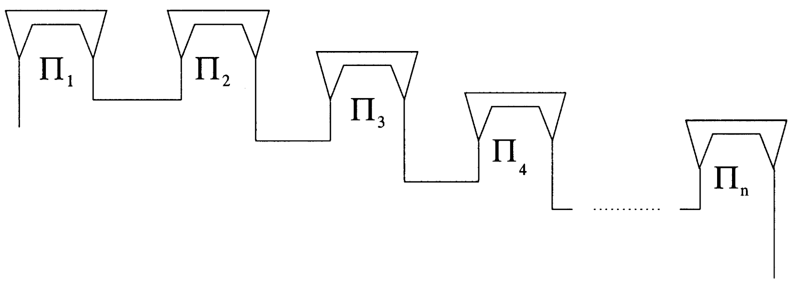
$$F(2^{2^j}), F(2^{2^j}) \to F(2^{2^{j+1}})$$

Then one applies a contraction to the two occurrences of $F(2^{2^j})$ on the left and derives the sequent.) We can then combine a sequence of these proofs together using cuts to get a proof of

$$F(2) \to F(2^{2^n}) \tag{13}$$

in $O(n)$ steps.

The logical flow graph for the proof of (13) looks roughly as follows



where the notation Π_j, $1 \leqslant j \leqslant n$ refers to the proofs of (12). The logical flow graph of each Π_j contains two branches, one for the contraction of two occurrences of $F(2^{2^{j-1}})$ on the left, and another for the use of the *F:times* rule on the right. Along the graph we notice a *chain* of n pairs of branches which are supposed to be eliminated one by one by the intermediate steps of the procedure of cut elimination. It is during their resolution that the expansion of the proof takes place. In particular any cut-free proof of (13) has size $2^{\mathcal{O}(n)}$ and we will see that this number corresponds to the number of paths passing through the chain above which starts at $F(2)$ and ends in $F(2^{2^n})$.

We will see that the chain of rhomboidal patterns occurring in the example turns out to be a *crucial* configuration to ensure an exponential expansion of proofs. We will see this also in the examples of Section 5 as well as in the lower bounds proved in Section 9.2. The idea of long chains of these rhomboidal patterns has a very natural logical interpretation. One can think of having a chain of "facts", each of which is used twice in order to derive the next. By the end of the proof the first fact is used an exponential number of times, at least implicitly.

In Section 14 we discuss some other concrete example of proofs with exponential expansion which does not present the geometry above in the same explicit way. We will see how the same geometry can be recaptured however at a more implicit level.

One of the main points of our combinatorial treatment of proofs is the notion of *visibility* graph which represents the way that the graph looks from the perspective of one of its vertices. Roughly speaking, we will show that the process of cut elimination transforms a proof into a cut-free one where the visibility of its vertices is "explicit". In the example discussed in this section, any cut-free proof of (13) will "compute" $F(2^{2^n})$ from an exponential number of copies of $F(2)$ and the logical flow graph of these proofs will "contain" an explicit tree of multiplications whose root is associated to the formula $F(2^{2^n})$.

4. Optical graphs and duplication

The notion of optical graph has been introduced in [10].

Definition 2 (Carbone–Semmes)**.** An *optical graph* is an oriented graph with the property that each vertex v has at most three edges attached to it. No more than *two* are oriented away from v and no more than *two* are oriented towards v.

A logical flow graph is obviously an optical graph. The word "optical" refers to the idea of "looking" through the graph, following rays of light. Suppose to stand in a vertex v of the graph with a source of light and imagine the rays of light to run along its oriented paths starting at v. In the context of proofs the idea of "rays of light" corresponds to fixing a piece of information and asking where it came from or how it will be used later.

We say that a vertex p *follows* a vertex q in an optical graph G if there is a directed path from q to p in G. It is clear that the points lying in a cycle follow one the other.

A vertex v in an optical graph is called a *branch point* if it has exactly three edges attached to it. We say that v is a *focussing* branch point if there are two edges oriented towards it. We call v a *defocussing* branch point if the two edges are oriented away from it.

A vertex in an optical graph is called *input vertex* if there are no edges in the graph which are oriented towards it. A vertex is called *output vertex* if there are no edges in the graph which are oriented away from it.

A *cycle* in an optical graph is a path which starts and ends in the same vertex.

Definition 3. A *labelled optical graph* is an optical graph with the property that the edges oriented away from a defocussing point and the edges oriented towards a focussing point are labelled 1 and 2, respectively.

A logical flow graph is a labelled optical graph where we think of the pairs of edges at a branch points as labelled by 1 and 2 depending on whether they lie on the *left* or on the *right* of each other (in the proof).

Definition 4. The *duplication* $\mathscr{D}$ is a binary operation applied to a labelled optical graph G and a subgraph G' of G with the property that
1. if a vertex of G' is a focussing point in G then either its immediate predecessor vertices both lie in G' or none of them does, and
2. if a vertex of G' is a defocussing point in G then either its immediate successor vertices both lie in G' or none of them does, and
3. at least one input vertex in G' is a focussing point, or at least one output vertex in G' is a defocussing point.

Let $v_1,\dots,v_n$ be the input vertices of G' which are not inputs in G, and let $w_1,\dots,w_n$ be the output vertices of G' which are not outputs in G. The result of duplication applied to G, G' is a graph $\mathscr{D}(G',G)$ which is defined as G except on the subgraph G' which will be substituted by two copies of it.

Namely all vertices of G which are not in G' lie in $\mathscr{D}(G',G)$ as well as those edges in G which connect any two of these vertices. Moreover, $\mathscr{D}(G',G)$ contains two copies G_1', G_2' of G' which are attached to the "rest" of $\mathscr{D}(G',G)$ as follows:

a. Let v_i be an input vertex of G' which is not focussing in G and let v_i^1, v_i^2 be the two copies of it in G_1', G_2' of $\mathscr{D}(G',G)$ respectively. Add a new vertex u_i in $\mathscr{D}(G',G)$ and two edges from u_i to v_i^1, v_i^2 with labels 1 and 2, respectively. If s_i is the vertex in G with an edge to v_i, add an edge from s_i to u_i in $\mathscr{D}(G',G)$;

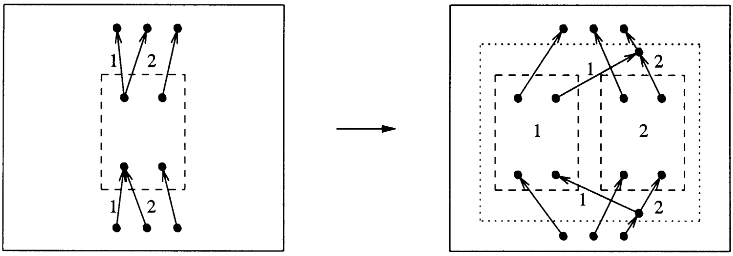
b. Let w_i be an output vertex of G' which is not defocussing in G and let w_i^1, w_i^2 be the two copies of it in G_1', G_2' of $\mathscr{D}(G',G)$, respectively. Add a new vertex t_i in $\mathscr{D}(G',G)$ and two edges from w_i^1, w_i^2 to t_i with labels 1 and 2, respectively. If r_i is the vertex in G with an edge from w_i to it, add an edge from t_i to r_i in $\mathscr{D}(G',G)$;

c. Let v_i be an input vertex of G' which is focussing in G and let v_i^1, v_i^2 be the two copies of it in G_1', G_2' of $\mathscr{D}(G',G)$, respectively. If s_i^1, s_i^2 are the vertices in G with edges to v_i labelled 1 and 2, respectively, add an edge from s_i^1 to v_i^1 in $\mathscr{D}(G',G)$, and add an edge from s_i^2 to v_i^2 in $\mathscr{D}(G',G)$.

d. Let w_i be an output vertex of G' which is defocussing in G and let w_i^1, w_i^2 be the two copies of it in G_1', G_2' of $\mathscr{D}(G',G)$, respectively. If s_i^1, s_i^2 are the vertices in G with an edge from w_i, add an edge from w_i^1 to s_i^1 in $\mathscr{D}(G',G)$, and add an edge from w_i^2 to s_i^2 in $\mathscr{D}(G',G)$.

This concludes the definition of the graph $\mathscr{D}(G',G)$.

A symbolic illustration of the definition is



Here the graph G is represented by the larger box on the left, the graph $\mathscr{D}(G', G)$ by the large box on the right and the subgraphs G', G'_1, G'_2 (which respectively lie in G and $\mathscr{D}(G', G)$) by the dotted ones. The dotted box on the left has one input branch point, one output branch point and two other points (one input and the other output) which are duplicated and linked to the supporting structure of G on the right-hand side. In general there are no limits on how many input and output vertices G' might contain. Moreover, notice that there might be input and output vertices in G' which are inputs and outputs vertices in G. We did not indicate them in our symbolic picture because they will remain input and output vertices in $\mathscr{D}(G', G)$ and will not involve the introduction of new edges. The first copy G'_1 of the subgraph G' in $\mathscr{D}(G', G)$ is linked to the vertices of $\mathscr{D}(G', G)$ which are reached by edges of label 1 in G. There are no edges to those vertices linked in G by edges of label 2. The second copy G'_2 of the subgraph G' in $\mathscr{D}(G', G)$ is linked in a similar manner to the rest of the graph G. The input and output vertices in G' which are neither focussing nor defocussing will be joint together through defocussing and focussing points, respectively.

Remark 5. Notice that requirement 3 in Definition 4 is *not* necessary for the operation $\mathscr{D}$ to make sense. In fact the definition of $\mathscr{D}$ would perfectly work without this requirement. We ask for it because we want to 'see' $\mathscr{D}$ as a way to resolve focussing and defocussing points in G.

Remark 6. From the definition of *input* vertex (the same follows from the definition of output vertex), a point in G' cannot be at the same time input *and* output, except if G' is a graph constituted by a single point.

Definition 7. Let G be a graph and G' be a subgraph of it. We say that a point in G is a *boundary* point if it does not belong to G' but it is linked by an edge to some point of G'.

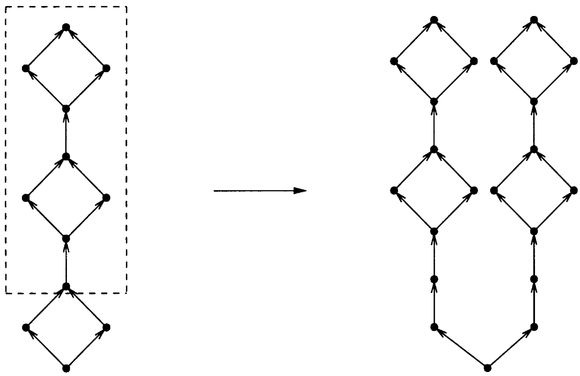
We denote $G'_1 \oplus G'_2$ the subgraph of $\mathscr{D}(G', G)$ which has as boundary points the copies of the boundary points of G' in G. In the symbolic picture following Definition 4 the external dotted box on the right represents $G'_1 \oplus G'_2$.

We shall discuss the basic properties of *duplication* in Sections 6–8, after describing some concrete examples of growth of duplicated graphs first.

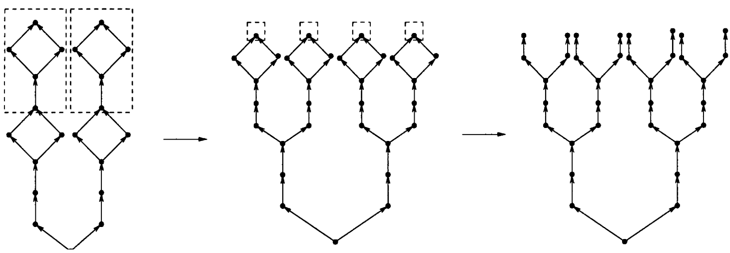
5. Some examples of growth

The repeated application of duplication to a graph might transform it in arbitrary ways. We are interested to study those transformations which will render it "explicit". The focussing and defocussing points in a graph will be redistributed and rearranged. They will be pushed on the boundary of the graph while the central structure of the graph will simplify. This is to say that no focussing point will be following a defocussing one. We present here a number of examples where this point emerges clearly. In particular we will see that the growth rate of the size of the graph varies from graph to graph depending on their *geometric* properties.

Let us start with the graph G on the left

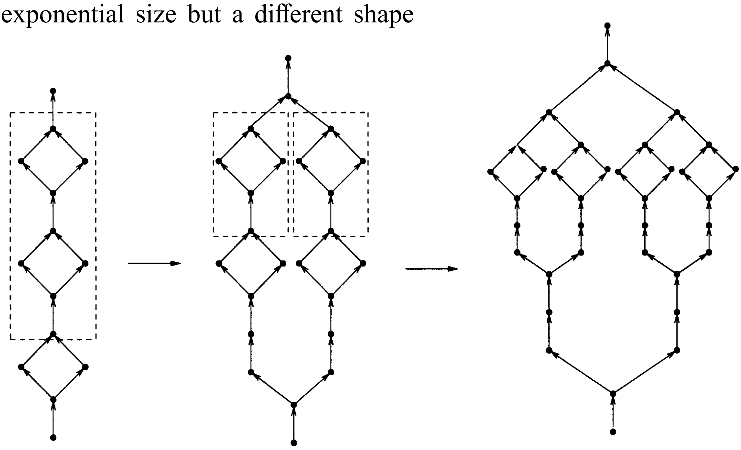


and let G' be the subgraph of G lying in the dotted box of the picture. The graph $\mathscr{D}(G,G')$ is represented on the right. The repeated application of duplication to G represents a kind of branching process, in which the duplications that are implicit in G are made explicit.

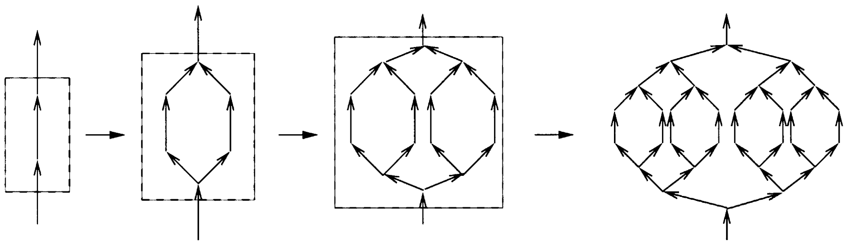


One can see that the result of the repeated application of duplication to G is of finite but *exponential* size compared to the size of G. This example is basic and we shall discuss it again later.

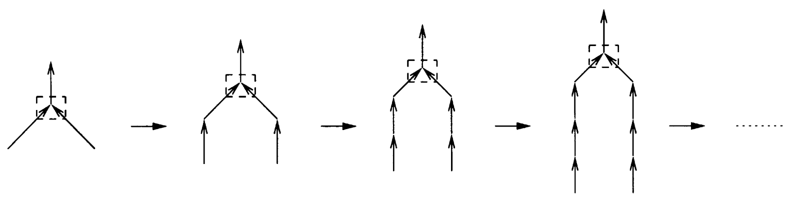
Notice that a slight modification of the starting graph can lead to a graph which has again exponential size but a different shape



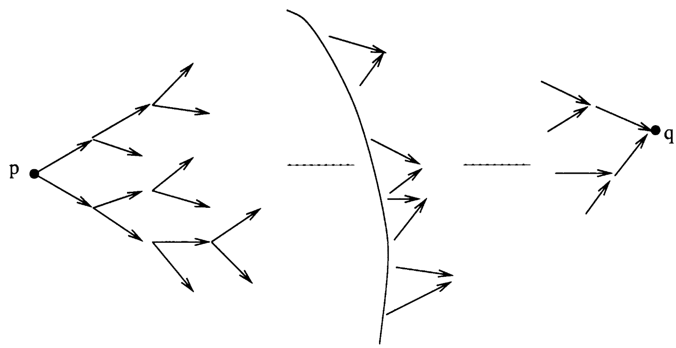
Notice that in the definition of duplication we do not ask for connectedness of G'. This means that there might be subgraphs of G' as pictured in the box below, which would induce an exponential expansion of the graph.



An example of *linear* growth of the graph can be seen as follows. Take the tree with four vertices and three edges illustrated below. By applying repeatedly the duplication to the subgraph constituted by the only branching point of the tree we obtain the following linear transformation:



We see now that any *polynomial* growth can be obtained. Let us consider the graph H given as follows:



All of the branch points on the left side of the picture are *defocussing*, while on the right hand side they are all *focussing*. For simplicity we assume that the transition between the two occurs after some number k of steps away from p. From the interface to the vertex q we assume to have the same number of steps k.

As argued in [10], the total number N of vertices in H is roughly proportional to 2^k, i.e. N is bounded from below and above by constant multiples of 2^k. This is easy to check by summing a geometric series.

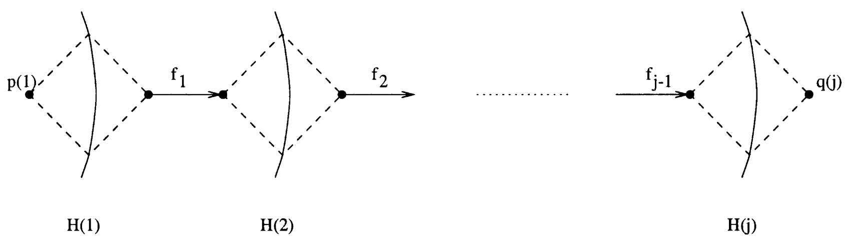
If we repeatedly duplicate subgraphs of H having only input vertices which are $k+i$ edges away from p (for $0 \leqslant i \leqslant k$) (they will be focussing points in H by construction) and no output vertices, we find after a finite number of steps a tree of size $N \log N$. (This bound was pointed out by L. Levin.) However one can obtain a tree of *quadratic* size (compared to the underlying graph) by adding a "tail" to the end of H. More precisely, think to add a sequence of vertices $v_1, \ldots, v_L$ and oriented edges $e_1, \ldots, e_L$ between them in such a way that e_1 goes from q to v_1 and e_i goes from v_{i-1} to v_i, for $1 < i \leqslant L$. Call this graph H^*.

Let us suppose that L is at least as large as N. Then the total number of vertices in H^* is $N + L$, while the number of vertices of the tree obtained by duplicating H^* as before is roughly proportional to $N \cdot L$. This is easy to show. If we take L to be equal N, then the number of vertices in the tree is roughly the square of the number of vertices in H^*. If we take L to be approximately N^{α}, $\alpha \geqslant 1$, then the total number of vertices of the tree is roughly proportional to $N^{\alpha+1}$, which is approximately the same as the number of vertices in H^* raised to the power $(\alpha + 1)/\alpha$.

Since every real number s in the interval $(1, 2]$ can be realized as $(\alpha + 1)/\alpha$ for some $\alpha \geqslant 1$, the preceding construction shows that for any such s we can find families of graphs for which the size of the tree is roughly proportional to the size of the graph raised to the s-th power. One could also get more complicated functions through suitable choices of L, and obtain similar effects by choosing L to be less than N, e.g. a fractional power of N.

In the type of constructions we discussed, the size of the tree is never more than quadratic in the size of the starting graph. To go beyond quadratic growth one can proceed as follows.

Fix a $j \geqslant 2$. We define a new optical graph H_j as follows. We begin by taking j identical but disjoint copies of H, which we denote by $H(i)$, $1 \leqslant i \leqslant j$. Let $p(i)$ and $q(i)$ denote the input and output vertices of $H(i)$. The graph H_i is obtained by taking the union of the $H(i)$'s, $1 \leqslant i \leqslant j$, together with oriented edges f_i that go from $q(i)$ to $p(i+1)$ for $i = 1, 2, \ldots, j-1$. We define H_j for $j = 1$ to be H.



The number of vertices in H_j is equal to $j \cdot N$ but the number of vertices of the tree obtained by duplication as in the above construction, is roughly proportional to N^j, by the same considerations as above. (The constants which are implicit in this estimate are allowed to depend on j but not on N.)

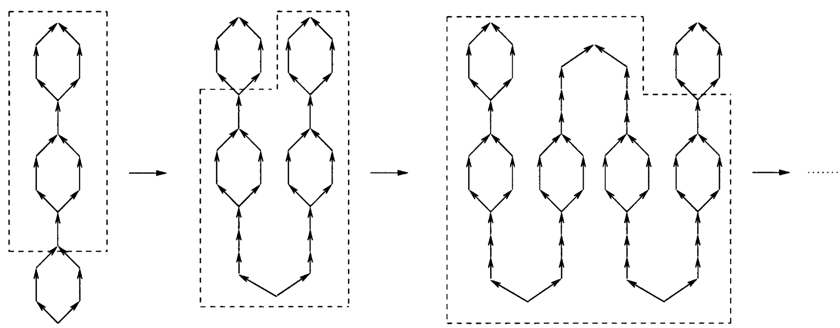
As before one can also add a string of L vertices to $q(j)$ to obtain a graph H_j^* which is analogous to the graph H^* above. This way one defines graphs whose associated tree is approximately prescribed as a function of the size of the original graph. For instance, if we take L to be N^{α}, $\alpha \geqslant 1$, then the total number of vertices of the associated tree (starting at $p(1)$) will be roughly comparable to $N^{j+\alpha}$ while the size of H_j^* will be roughly N^{α}. (As before one can think of fixing j and letting N be arbitrarily large.)

For fixed j and any choice of L the size of the tree associated to H_j^* as above is never more than a constant multiple of the size of H_j^* raised to the $j+1$-rst power. This is because the number of vertices in H_j^* is $jN + L$ while the number of vertices in the tree is $\mathcal{O}(N^j(N+L))$.

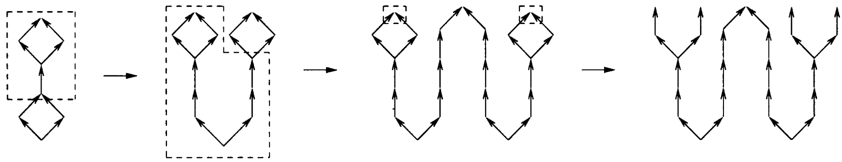
To obtain rates of growth which are larger than polynomial one should allow j to vary as well. For instance, one can take the graph G discussed at the beginning of this section and attach a chain of L vertices to it similarly as for H^* and H_j^*. By choosing L appropriately one can obtain almost any rate of growth in the size of the tree compared to the original graph.

We have seen how the growth of the graphs under repeated duplication might be linear, polynomial (of any degree) and exponential. From the constructions we discussed it starts to emerge that exponential growth depends on the existence of chains of alternating branching points in the graph. This is a crucial point that we will discuss further in detail.

For the moment, notice that the size of the graphs one can reach by duplication is arbitrarily large. In fact there are graphs where we can *infinitely* often duplicate and doubling the size of the graph at each step. Here is an example

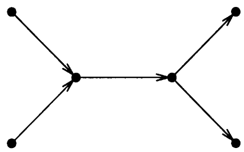


The process we illustrated in the figure can always terminate after having reached a graph of the desired size and having eliminated all alternations of branching points. An example of this process of termination is illustrated in the simpler graph below:



where in the last step one applies duplication to the pair of focussing points as depicted in the boxes on the left. This is not the only way we can proceed to eliminate all alternations. For instance we could have duplicated one of the boxes to obtain only one alternation and afterwards we could have applied duplication to the defocussing point of this alternation to resolve finally the graph. The results would look different.

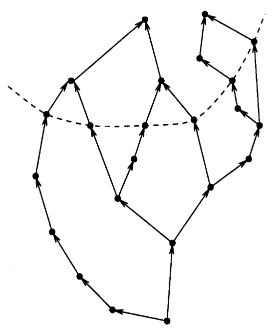
It is clear that given a graph G there are in general many choices of subgraphs G' to which one can apply the operation of duplication. In the future we are interested to transform G into graphs where no focussing point is followed by a defocussing one. The configurations



will be repeatedly eliminated and no new ones will be introduced by the strategies we will consider. Notice that the graph H described above does not contain any such configuration and since we will refer to its shape very often in the future, we say that

Definition 8. An optical graph is an *H-graph* if none of its focussing points is followed by a defocussing one.

An example of H-graph with much less regular shape than H is



where there is one input vertex and two output vertices. (In the terminology of [10] an H-graph was intended to have exactly one input and one output. Here we drop this restriction.)

The interest in the transformation of an optical graph into an H-graph lies into the interpretation of graphs as graphs of proofs. The configuration above corresponds to the presence of cuts in proofs and logical flow graphs of cut-free proofs are H-graphs. We will come back to this point in Sections 11 and 12.

Definition 9. A graph G' is the *resolution* of a graph G if it is obtained through repeated duplication from G and it is an H-graph.

Given a graph G there might be several resolutions of it associated to different strategies of application of $\mathscr{D}$. In Section 9 we will study one of these strategies and we will derive upper and lower bounds for the expansion. From now on we tacitly assume that the subgraphs G' of G satisfy conditions 1–3 of Definition 4 and therefore that the graph $\mathscr{D}(G', G)$ is defined on the choice of G'.

6. The canonical projection

Let G be an optical graph and G' be a subgraph of it. Let us borrow the notation of Definition 4 and also we use the symbol $G'_1 \sqcup G'_2$ to denote the subgraph of $\mathscr{D}(G', G)$ constituted by the two disconnected components G'_1 and G'_2. The notation G/G' will denote the set of vertices and edges lying in the graph G but not in its subgraph G' (for an edge to belong to G/G' we ask both its extremes to belong to G/G').

There is a *canonical projection*

$$\pi : \mathscr{D}(G', G) \to G$$

defined on the vertices as follows. Let us first define π on the vertices lying in G'_1, G'_2 and $\mathscr{D}(G', G)/G'_1 \oplus G'_2$. Then we use the definition on these vertices to establish the behavior of π on the remaining vertices of G.

$v1$. If x is a vertex lying either in G'_1 or in G'_2, then $\pi(x)$ is its copy in G' of G.

$v2$. If x lies in $\mathscr{D}(G', G)/G'_1 \oplus G'_2$ then $\pi(x)$ is its copy in G/G'.

$v3$. The vertices x in $G'_1 \oplus G'_2/G'_1 \sqcup G'_2$ which are output vertices in $G'_1 \oplus G'_2$ are sent to $\pi(y)$ where y is a predecessor of x in $G'_1 \oplus G'_2$.

$v4$. The vertices x in $G'_1 \oplus G'_2/G'_1 \sqcup G'_2$ which are input vertices in $G'_1 \oplus G'_2$ are sent to $\pi(y)$ where y is a successor of x in $G'_1 \oplus G'_2$.

This concludes the definition of the map π from vertices in $\mathscr{D}(G', G)$ to vertices in G. We now extend this map to a canonical projection on *edges* but we need to let it undefined over those edges in $\mathscr{D}(G', G)$ between vertices x lying in $G'_1 \oplus G'_2/G'_1 \sqcup G'_2$ and y lying in $G'_1 \sqcup G'_2$. For all other edges π will be defined as follows.

$e1$. If ε is an edge from a vertex x to a vertex y in $\mathscr{D}(G', G)$ and x, y do not lie in $G'_1 \oplus G'_2/G'_1 \sqcup G'_2$ then there is an edge σ in G between $\pi(x)$ and $\pi(y)$. Let $\pi(\varepsilon)$ be σ. The orientation of the edge is preserved.

$e2$. If ε is an edge between a vertex x in $\mathscr{D}(G', G)/G'_1 \oplus G'_2$ and y in $G'_1 \oplus G'_2$ then there is an edge σ in G between $\pi(x)$ and $\pi(y)$. Let $\pi(\varepsilon)$ be σ. The orientation of the edge is preserved.

This defines a map from edges in $\mathscr{D}(G', G)$ to edges in G with the obvious compatibility conditions between the two graphs. This mapping also preserves the orientations on the two graphs by construction.

Let us record here a few basic facts.

Proposition 10. *The following statements are true*:

a. *If y lies in $\mathscr{D}(G', G)/G'_1 \oplus G'_2$ then $\pi(y)$ and y have the same indegree and the same out-degree.*

b. *If y is an input* (*output*) *vertex in $G'_1 \sqcup G'_2$ then its in-degree* (*out-degree*) *is either* 0 *or* 1.

c. *If* y *is an input* (*output*) *vertex in* $G_1' \oplus G_2'$ *then its in-degree* (*out-degree*) *is either* 0 *or* 1. (*Notice that the in-degree* (*out-degree*) *of* $\pi(y)$ *might be* 2.)
d. *If* y *lies in* $G_1' \oplus G_2' / G_1' \sqcup G_2'$ *then its degree is* 3 *while* $\pi(y)$ *has degree* 2.

We do not give the proof of the statements because it is an immediate consequence of the definitions. The reader might like to check it by inspection of the symbolic picture given after Definition 4.

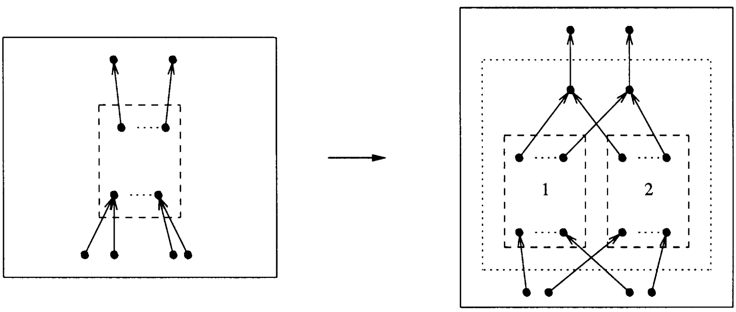
We denote $\#paths(v, G)$ the number of paths between the point v and any output vertex in G.

Proposition 11. *If the input vertices in* G' *are only focussing points and no output vertex in* G' *is defocussing then*

$$\#paths(v, \mathscr{D}(G', G)) = \#paths(\pi(v), G)$$

for any vertex v *in* $\mathscr{D}(G', G)$.

Proof. Consider the following symbolic picture of the duplication of a subgraph G' satisfying the hypothesis of the statement



The claim is easily derived by inspection of the picture. One only needs to take into account that the output vertices in G' might either be linked to G/G' or might be outputs of G. □

An analogous statement holds for output vertices as well

Proposition 12. *If output vertices in* G' *are only defocussing points and no input vertex in* G' *is focussing then*

$$\#paths(v, \mathscr{D}(G', G)) = \#paths(\pi(v), G)$$

for any vertex v.

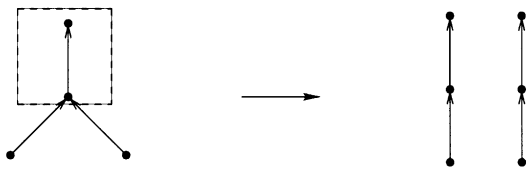
7. Basic properties of duplication

We present here a few basic facts on $\mathscr{D}(G', G)$ and on the behavior of its paths with respect to paths in G. Let us state first

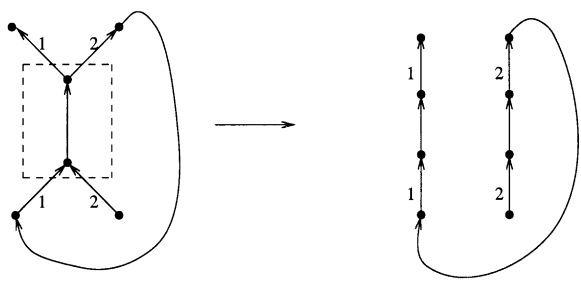
Proposition 13. *Let G be a labelled optical graph and G' a subgraph of it. Then $\mathscr{D}(G,G')$ is a labelled optical graph.*

Proof. This is proved by chasing the definition. □

If the graph G is connected then we do *not* necessarily have that the graph $\mathscr{D}(G',G)$ is also connected. Take for instance the situation illustrated here

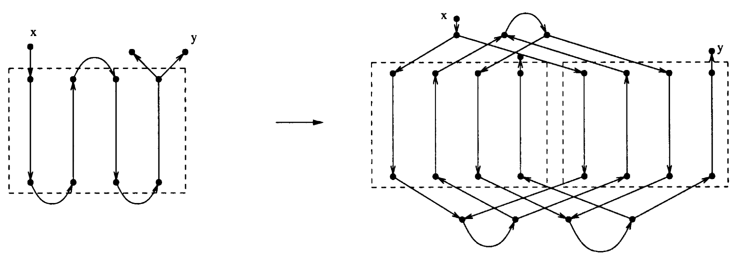


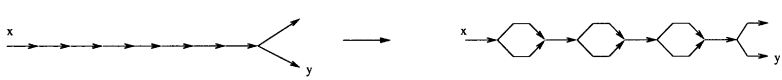
where the elimination of the branching point implies the disconnection of the resulting graph. At times this phenomenon has some global effect. In fact, it might eliminate cycles and as a consequence of this, infinite paths in G become finite in $\mathscr{D}(G',G)$. Let us consider the following optical graph:



The application of duplication gives here an acyclic graph. The duplication of G' creates two copies of the dotted box, one which will be linked to vertices originally linked by edges of label 1 and another which will be linked to vertices originally linked by edges of label 2. As a consequence, the cyclic path will be disconnected.

At times the number of paths from a point x in G to some output vertex y might increase substantially with the duplication. We see this phenomenon with the next example. Take the following graph G and consider the duplication illustrated on the right





where the unique path going from x to y in G is transformed into 2^3 paths in $\mathscr{D}(G', G)$. In the example we obtain the exponential effect by exploiting the fact that three distinguished parts of a path in G' are duplicated in such a way that three new defocussing points followed respectively by three new focussing points are formed. Each duplicated part is arranged so to follow one the other. In Section 8 we show that duplication cannot always disrupt cycles and we give sufficient and necessary conditions for this to happen.

In general a path in G' is duplicated into two paths in $\mathscr{D}(G', G)$ only when it lies between inputs and outputs of G' which are not respectively focussing and defocussing points in G. We use the notation $\#paths(V, W, G')$ to indicate the number of paths between the *boundary* points V and W which lie entirely in the subgraph G' of G, i.e. all vertices in the paths with the exception of V and W lie in G'.

Proposition 14. *Let G be a graph, G' be a subgraph of it. Let V, W be two points in the boundary of G', and v, w be a pair of points in G'. Suppose that there is an edge from V to v and edge from w to W. Then the following statements are true*:

1. *Suppose v be a focussing vertex and w be a defocussing vertex. If the edges between V, v and w, W are labelled differently, then*

   $$\#paths(V, W, G'_1 \oplus G'_2) = 0$$

   otherwise

   $$\#paths(V, W, G'_1 \oplus G'_2) = \#paths(V, W, G')$$

2. *If v is not focussing and w is defocussing* (*the same holds if v is focussing and w is not defocussing*) *then*

   $$\#paths(V, W, G'_1 \oplus G'_2) = \#paths(V, W, G')$$

3. *If v is not focussing and w is not defocussing then*

   $$\#paths(V, W, G'_1 \oplus G'_2) = 2 \cdot \#paths(V, W, G')$$

We skip the proof because trivial. One can inspect the symbolic picture given after Definition 4.

In general the number of inputs and outputs in G is *not* preserved by duplication. It is preserved only when inputs and outputs in G' are not inputs and outputs of G as well. In this latter case, the duplication would induce two copies of the terminal points as well.

Before ending this section, let us introduce a definition which establishes when two optical graphs have the same structure of branching points. This concept will be often used in the next sections. It makes sense for both connected and disconnected optical graphs.

Definition 15. Two optical graphs have the same *topological structure* if they can both be reduced to the same optical graph by collapsing each edge between pairs of points of degree at most 2 to a vertex.

We say that an optical graph is *minimal* if it has no edges between points of degree at most 2.

Proposition 16. *Given any optical graph G there is a minimal optical graph G' which has the same topological structure as G. Moreover, there is only one such minimal optical graph.*

Proof. Let G be an optical graph. Define G' to be the optical graph obtained by collapsing to a point *all* edges in G between two points of degree at most 2. This defines a graph which is minimal for G. For the uniqueness, we should notice that by collapsing the edges in whatever order we preserve the topological structure. In fact, any sequence of consecutive edges linking vertices of degree at most 2 will be collapsed to an empty sequence, and this does not depend on the order of collapse. □

If G is an acyclic optical graph, the *height* of a vertex v in G is the maximal length of the paths in G starting from some input vertex of G and ending in v. The *height* of G is the maximal height of its input vertices. Clearly the longest paths in G start from input vertices and end in output vertices of G.

If G is an optical graph of height n, we say that a subgraph G' of G has height $k \leqslant n$ if *all* vertices of height k in G are contained in G' and no vertex of height greater than k is contained in G'.

8. Duplication and cycles

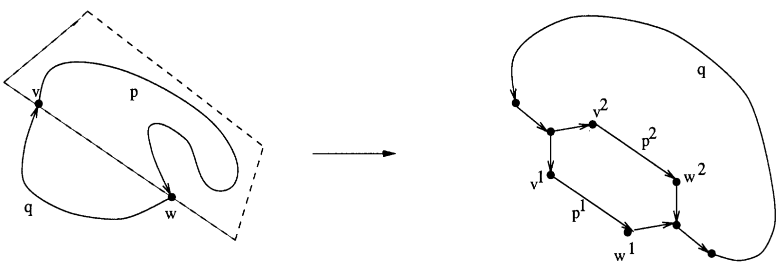
We show here some general facts on graphs with cycles. In Section 7 we have seen that a cycle in an optical graph might be disrupted by duplication. There are times when no attempt of duplication can split a cycle though. Theorem 25 at the end of this section gives necessary and sufficient conditions for the elimination of a single cycle. These conditions are *always* satisfied by graphs of proofs as discussed at the end of Section 11.

In the following we make a distinction between loops and cycles. A *loop* is a cycle which does not pass over the same edge twice. It is clear that if there are no loops in a graph then the graph is acyclic. We say that a loop *meets* another loop (or equivalently that two loops are *nested*) if they have one vertex in common.

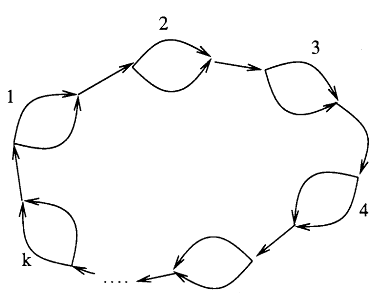
A loop in an optical graph has a *way in* if one of its vertices is a focussing point. A loop has a *way out* if one of its vertices is a defocussing point. Notice that for the focussing point, there is only one of the edges branching in which belongs to the loop and the other *comes in* from some other parts of the graph. Similarly for the defocussing point.

Proposition 17. *Let G be an optical graph containing a loop. If a loop has no way in and no way out, then it cannot be eliminated by duplication.*

Proof. The loop *forms* a connected component L of the graph since by assumption there are no points lying in the loop which are branching points. If $\mathscr{D}$ is applied to some subgraph G' of G then we can see that either G' does not contain any vertex in L, or L completely lies in G', or G' contains a sequence of consecutive edges lying in L. In the first case, the duplication does not involve L and hence the loop is not disrupted in $\mathscr{D}(G', G)$. In the second case, the loop will be duplicated but not disrupted. The third case creates two nested loops in $\mathscr{D}(G', G)$. This is because the pair of points in G' which are input and output of G' and lie in L are neither focussing nor defocussing points. This means that by duplication we will have a transformation which can be illustrated as follows:



where v, p, w have been duplicated on the right hand side and their copies have been denoted by superscripts 1, 2 respectively. Notice that the picture illustrates only one of the possible cases. In fact, it might be that many sequences of L belong to G', say k of them. In this case L will be transformed in a chain of nested loops of the form



This concludes the proof. □

Proposition 18. *Suppose that G is an optical graph with a loop having a way in but no way out. Then there is no application of duplication to G that breaks this loop. The same conclusion holds if the loop has a way out but not a way in.*

Proof. We suppose that the loop has a way in and no way out. (The case where the loop has a way out but no way in is symmetric.) This means that no vertex lying in the loop is a defocussing vertex.

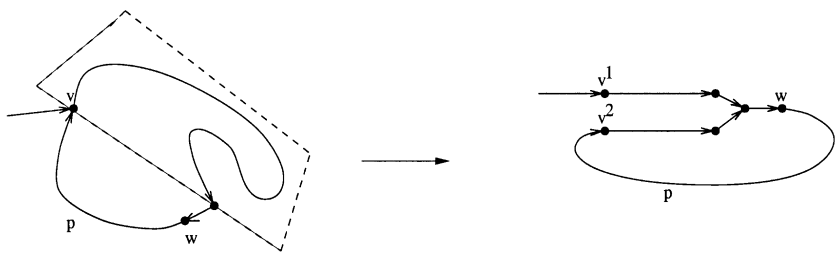
We want to show that for any choice of G', the graph $\mathscr{D}(G', G)$ will still contain the loop.

Suppose first that no input or output vertices of G' lie in the loop. This means that either the loop is all contained in G' or that it lies outside G'. In the first case the action of $\mathscr{D}$ will duplicate the loop without breaking it. In the second case the loop will remain untouched by the duplication.

Suppose now that G' contains an input vertex lying in the loop. Then there will be an output vertex of G' which will also be lying in the loop. This is easy to see. In fact, if a vertex in G' lies in the loop and it is not an output vertex, then its successor in the loop has to be in G'. If we repeat this reasoning by going along the loop vertices, either we end-up in the input vertex from which we started or we meet an output vertex before. In the second case we are done. If we end up in the input vertex where we started, we have a contradiction because input vertices cannot have edges coming in.

Arguing in a similar manner one derives that if G' contains an output vertex lying in the loop then it will contain also an input vertex which lies in the loop. In particular, the output vertex is not a defocussing point because the loop has no way out.

If the input vertex is a focussing point, then by duplicating G' we have the following situation



and therefore $\mathscr{D}(G', G)$ still contains a loop. (Notice that the picture illustrates only the part of the graphs G and G' that is involved in the transformation of the cycle.)

If the input vertex is not a focussing point then the situation is illustrated by the two last pictures in the proof of Proposition 17. As argued in Proposition 17 the sequences of edges belonging to the loop may be several and in this case one needs to notice that the treatment of each of them will fall in one of the cases already described.

This concludes the proof. □

Proposition 19. *Let G be a labelled optical graph containing n loops that pairwise do not meet. Let G' be a subgraph of G such that no loop lies entirely in G'. Then $\mathscr{D}(G', G)$ contains at most n loops that pairwise do not meet.*

Proof. We need only to show that no new loop which is disconnected from the n loops already existing can be created by duplication. This is a consequence of the fact that no *new* sequences of edges are added between vertices which were not already linked in G. To see this latter fact is easy. We start by noticing that a duplication does not change the structure of the subgraphs G'_1, G'_2 (i.e. the two copies of the subgraph G') and it does not change the structure of the graph G outside G'. It only changes the connections between the copies of the boundary points of $G'_1 \oplus G'_2$ in $\mathscr{D}(G', G)$ and the input and output vertices of G'_1, G'_2. From Proposition 14, two such boundary points are linked in G iff they are linked in $\mathscr{D}(G', G)$ also. This proves the claim. $\square$

The following proposition gives conditions under which the number of loops *does not increase* in $\mathscr{D}(G', G)$.

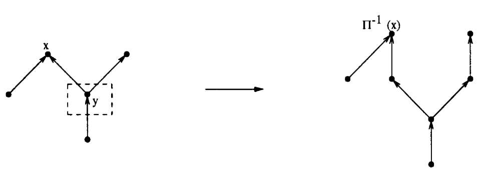
Proposition 20. *Let G be a labelled optical graph containing n loops and G' be a subgraph of G. Suppose that no loop lies entirely in G' and that no pair of vertices v, w in G' have the property that v is an input, w is an output, v is not focussing, w is not defocussing, and v, w lie in a loop in G. Then $\mathscr{D}(G', G)$ contains at most n loops.*

Proof. From the first hypothesis, a loop cannot entirely be duplicated by $\mathscr{D}$. From the second hypothesis on pairs of vertices v, w lying on a loop and Proposition 14 we know that no subsequence lying on a loop is duplicated in $\mathscr{D}(G', G)$. This means that either the loop remains untouched by $\mathscr{D}$ or it will be disrupted. Moreover, no new loops can be created because a pair of points x, y is connected in $\mathscr{D}(G', G)$ if $\pi(x), \pi(y)$ are already connected in G (where $\pi : \mathscr{D}(G', G) \to G$ is the canonical projection). Therefore if a loop exists in $\mathscr{D}(G', G)$, it should exist already in G. This proves the claim. $\square$

We want to study now when the number of loops actually *decreases* in $\mathscr{D}(G', G)$. To see this we first need to prove a technical lemma which will be useful also later in the paper. (We remind the reader that the concept of *topological structure* was introduced in Definition 15.)

Lemma 21 (Blowing Lemma). *Let G be an optical graph. There is an optical graph G' that has the same topological structure as G and has no edge from a defocussing point to a focussing one. Moreover $|G'| \leqslant 3 \cdot |G|$.*

Proof. Take any edge in G which starts from a defocussing point y and ends into a focussing point x. Consider the subgraph G'' defined by the only point y (the choice of y is arbitrary and we could have chosen x instead) and look at $\mathscr{D}(G'', G)$. After the duplication we have that $\pi^{-1}(x)$ is a focussing point whose incoming edge arrives from a copy of y which is not defocussing. This is illustrated below



(The picture represents only the subgraph of G which is relevant to the discussion.) The situation is analogous when x is chosen instead of y. By repeating this transformation over all edges satisfying the hypothesis we obtain a graph free of ‘bad’ configurations. To show that the procedure must terminate in at most the number of branching points of G is easy. It is enough to observe that the operation we described adds at each step a pair of new edges and a pair of new vertices (which are not branching points) to the graph and nothing more, therefore the number of ‘bad’ edges decreases at each step and in particular $|G'| \leqslant 3 \cdot |G|$. □

The graph G' in Lemma 21 is not uniquely defined. This is because the step of the construction in the proof could have been applied to the vertex x instead of the vertex y. The result would be a different graph that would satisfy nevertheless the conclusions of the lemma. We denote $Blow(G)$ any graph obtained as a result of the procedure described in the Blowing Lemma.

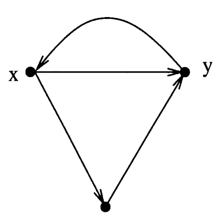
If the graph G we start with has no edges from a defocussing point to a focussing one, then there is no need to apply the procedure and we set $Blow(G) = G$. The following fact holds:

Proposition 22. *Let G be an optical graph which has no edge starting from a defocussing point and ending into a focussing one, and let G' be a subgraph of G. Then $Blow(\mathscr{D}(G',G)) = \mathscr{D}(G',G)$.*

Proof. We inspect the operation of duplication and notice that we should only worry about the *new* nodes (which are either focussing or defocussing points) introduced by duplicating, and by the *new* edges which link these nodes to copies of old nodes. In fact, all other edges in $\mathscr{D}(G',G)$ that start from defocussing points do not end into focussing ones. They are copies of edges in G that by hypothesis does not contain “bad” edges.

Let x be a new node in $\mathscr{D}(G',G)$. If x is a defocussing point, note that the edges departing from it cannot end-up into a focussing node y. This is because y is a copy of an input node of G' that cannot be focussing by hypothesis (otherwise the operation of duplication would not have introduced x). If x is a focussing point, we proceed similarly. We note that the edges arriving into x cannot start from a defocussing node y. This is because y is a copy of an output node of G' that cannot be defocussing by hypothesis (otherwise the operation of duplication would not have introduced x). □

Remark 23. Consider the graph G containing two loops as depicted below



There is an edge going from the defocussing point x to the focussing point y. It is easy to check that there is no subgraph G' for which $\mathscr{D}(G', G)$ is defined and contains one loop less than G. This example illustrates the hypothesis of Theorem 25.

Lemma 24. *Let G be an optical graph with no edge from a defocussing point to a focussing one. Then any assignment of labels $1, 2$ to edges going out from defocussing points and coming in focussing ones transforms G into a labelled optical graph.*

Proof. The hypothesis ensures that any assignment of the labels $1, 2$ to a pair of edges either going in or coming out from a branching point will not interfere with the assignment to a pair of edges of some other branching point. In fact this could be the case for those edges which were connecting a defocussing point to a focussing one and by hypothesis G does not contain such edges. $\square$

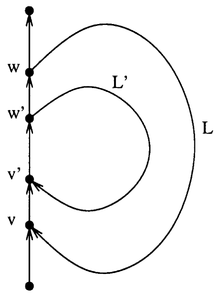
Theorem 25. *Let G be an optical graph containing n loops and assume that no edge in G which starts from a defocussing point ends into a focussing one. There exists a labelling G^* of G such that $\mathscr{D}$ transforms G^* into a graph with $n-1$ loops if and only if there is a loop in G that has a way in and a way out.*

Proof. Suppose that there is a labelling G^* of G such that $\mathscr{D}$ transforms G^* into a graph with $n-1$ loops. By Propositions 17 and 18 we know that for a loop to be eliminated by the $\mathscr{D}$ operation it must have both a way in and a way out. This proves one direction of the statement.

For the other direction, consider the family of loops with a way in and a way out in G. There might be several pairs (v, w) of branching points v, w lying in some loop of G where w follows v, v is focussing and w defocussing. Choose a pair (v, w) and a loop L such that the pair has minimal distance in G. (The *distance* between two vertices v and w in G is the minimal number of edges between these vertices (denoted $d(v, w)$).) We want to show that there is a labelling G^* of G and a subgraph G' of it such that $\mathscr{D}(G', G^*)$ has $n-1$ loops. Notice that G can be labelled because of Proposition 24.

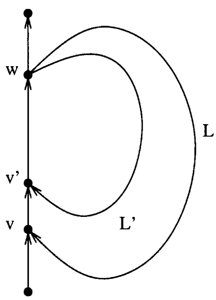
Let X be the set of vertices lying in L between v and w. Define G' to be the subgraph of G that contains the set of vertices X. If y is a focussing point in X and it is not v, then we ask that both its immediate predecessors lie in G'. If y is a defocussing point in X and it is not w, then we ask that both its immediate successors lie in G'. These are the only vertices in G'. The edges in G' are all the edges in G which link the vertices of G'. The vertices v, w are input and output vertices in G' and $\mathscr{D}(G', G^*)$ is defined for all labelling G^* of G because G' satisfies conditions 1–3 in Definition 4.

By hypothesis on the minimality of (v, w) one derives that there are no pairs of branching points (v', w') lying between v and w (notice that v' might be v and w' might be w) which also lie in a loop of G. In fact, suppose that there was a focussing vertex v' and a defocussing vertex w' as illustrated in the picture below



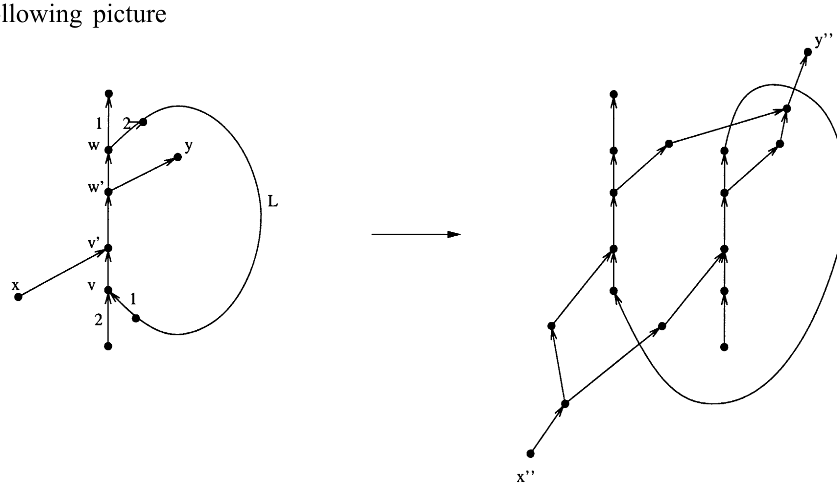
by minimality this is impossible because $d(v,w) \leqslant d(v',w')$ for any pair (v',w') and any loop L' in G. (Notice that we have chosen the loop in such a way that the pair is minimal in G.)

In case the vertex w' coincides with the vertex w as illustrated below,



the pair (v',w) and the loop L' would contradict our choices by minimality. The case where v' coincides with the vertex v is similar. Hence all pairs of branching points (v',w') between v and w do not belong to a loop in G.

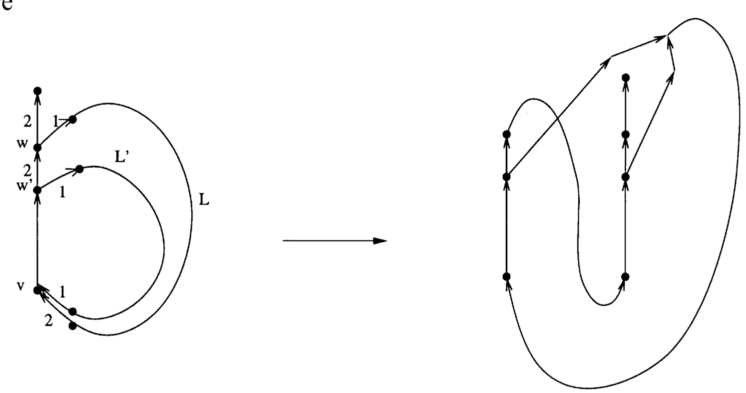
Therefore, fixing (v,w) we label 1 and 2 the edges going in and coming out the loop L from v and w, respectively. We also label $2,1$ the twins edges lying in L. We assign an arbitrary labelling to all other branching points in G. (This is possible because of Lemma 24.) This labelling guarantees the splitting. In fact, let us consider the following picture



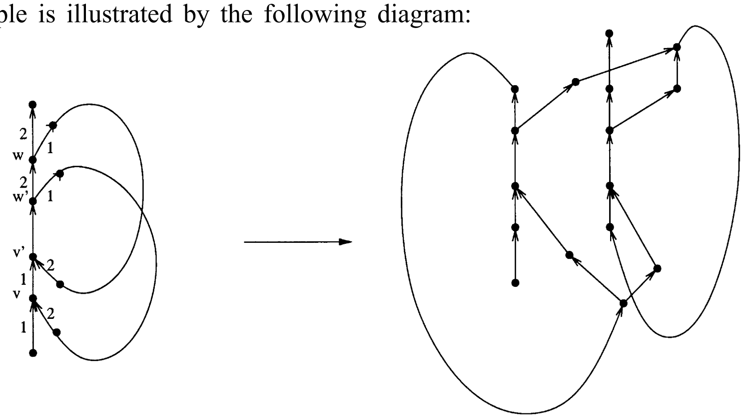
The paths between x'' and y'' (i.e. the images of x and y) will be doubled by duplication but they will not introduce any new nested loops since the paths going out of w' do not end up in any cycle which is coming back to v' (as argued before).

This concludes the proof. □

Note that arbitrary labelling would falsify the theorem as illustrated in the next picture

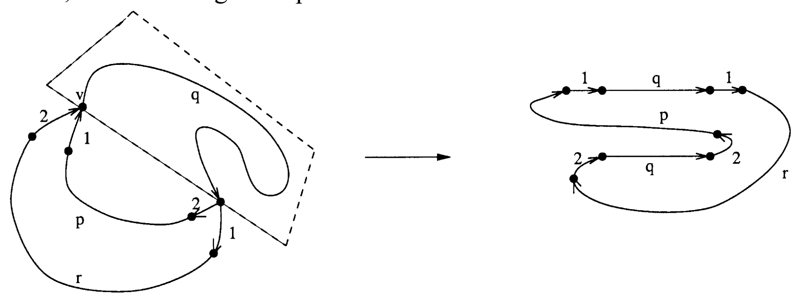


Here we see that two nested loops are duplicated into two nested loops. Another example is illustrated by the following diagram:



where two nested loops are transformed into two loops that do not meet.

If a graph G contains loops with way in and way out, it might be that by duplication the graph is transformed into another where cycles cannot be eliminated anymore. Take, for instance, the following example:



where the resulting loop has no way in and no way out. One might like to find a *global* criterion for the elimination of cycles in optical graphs but we will not do it here.

9. Positive resolution

In this section we define a specific strategy to transform any *acyclic* optical graph G into an H-graph with no focussing vertices (this graph will look like several trees glued together). It will be a transformation which despite its specific nature will bring some light on the relations between the structure of a graph and the growth rate of its expansion. We will give *upper* and *lower* bounds for its resolutions.

Definition 26. Let G be an acyclic optical graph and v be a vertex in G. The *index* of v (denoted *index*(v)) is the maximum number of focussing points lying in a path from some inputs in G to v.

Definition 27. Let G be an acyclic optical graph and p a path in G starting from some input vertex of G and ending in some vertex v. The index of the path p (denoted *index*(p)) is index(v). The index of the graph G (denoted *index*(G)) is the maximum among the indexes of the paths in G.

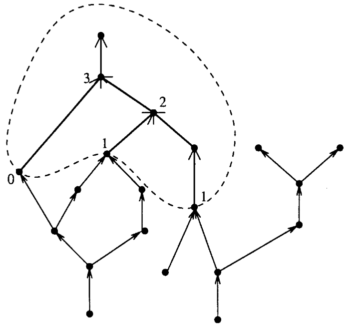
Let n be the index of G. We define now a graph G_n^+ which we call the *positive resolution* of G. By induction on $0 \leqslant i \leqslant n$, we build a graph G_i^+ as follows:

$$G_0^+ = Blow(G)$$

$$G_{i+1}^+ = \mathscr{D}(G_i', G_i^+)$$

where G_i' is the subgraph of G_i^+ defined as follows. Consider all maximal paths in G_i^+ starting from input vertices of G_i^+. For each path p, take the first vertex lying in p such that either it has index 1 in G_i^+ or it has index 0 and it is an immediate predecessor of a vertex of index >1. This vertex is an input vertex of G_i'. As output vertices of G_i' we take the output vertices of G_i^+. The graph G_i' is the subgraph of G_i^+ that contains all those vertices in G_i^+ which lie in some paths of G_i^+ going from an input vertex of G_i' to an output vertex of G_i'. Given any two vertices in G_i', there are as many edges in G_i' between these two vertices as there are in G_i^+. The orientations are also preserved. It is easy to check that all vertices v of index $\geqslant 1$ in G_i^+ lie in G_i'.

To see an example of subgraph G_i' defined out of a graph G_i^+ take the following



where the subgraph G_i' is circled by a dotted line. Note that the inputs of G_i' in the picture are the focussing points of index 1 and the vertex of index 0 which is immediate predecessor of the vertex of index 3.

The graph G_i' is a subgraph of G_i^+ and $\mathscr{D}(G_i', G_i^+)$ is defined, for each $i \leqslant n$, since G_i' satisfies properties 1–3 required by Definition 4. In fact, if v is a focussing point in G_i^+ with $index(v) = 1$, then v belongs to G_i' as input vertex and therefore its immediate predecessors will not lie in G_i'. If v is a focussing point in G_i^+ with $index(v) > 1$, then both its immediate predecessors v_1, v_2 should lie in G_i'. This is because either $index(v_1), index(v_2) \geqslant 1$ (and therefore v_1, v_2 lie in G_i'), or if $index(v_j) = 0$ (for some $j = 1,2$) then by construction v_j should be an input of G_i'. (This proves that property 1 is satisfied.)

If v is a defocussing point in G_i^+ with $index(v) = 0$ then v does not lie in G_i'. This is true because $G_i^+ = Blow(G_i^+)$, for all $i \geqslant 0$. (For $i = 0$ we notice that G_0^+ has been defined by applying the blowing operation to G and for $i > 0$ we argue by induction using Proposition 22). This ensures that v cannot be the immediate predecessor of a node w having an index >1. In fact G_i^+ does not contain edges linking a defocussing point to a focussing one. If $index(v) \geqslant 1$ then its immediate successors belong to G_i' because they lie in some paths between the input vertices of G_i' and the output vertices of G_i^+. (This shows that property 2 holds.)

If $index(G) = n \geqslant 1$ then there is at least one input of G_i' which is a focussing point for all $1 \leqslant i \leqslant n$. This is because $index(G_0^+) = index(G)$ and $index(G_i^+) = n - i$ as we show next. (This shows that property 3 in Definition 4 holds.)

Let us denote with $\pi^i : G_i^+ \to G_0^+$ the projection obtained by composing the canonical projections $\pi_i : G_i^+ \to G_{i-1}^+$, for $n \geqslant i \geqslant 1$ in the obvious way. The following two lemmas illustrate how the index of the graph G_i^+ decreases when i increases.

Lemma 28. *For all* $0 \leqslant i \leqslant n$, $index(G_i^+) = n - i$.

Proof. First let us observe that by the construction in the Blowing Lemma, $index(Blow(G)) = index(G)$ for any acyclic optical graph G. In fact, the blowing operation does not introduce any new focussing points in $Blow(G)$ nor it eliminates them.

Then we want to observe that the definition of G_i^+ $(= \mathscr{D}(G_{i-1}', G_{i-1}^+))$ duplicates the subgraph G_{i-1}' by reducing all focussing points v of index 1 in G_{i-1}^+ to points w (where $\pi_i(w) = v$) with $index(w) = 0$ in G_i^+. In particular, it does not introduce any focussing points since all the outputs of G_{i-1}^+ are also output points for G_{i-1}'. Therefore there are no vertices in G_{i-1}^+ whose counterimage has larger index.

This implies that for all paths p in G_i^+, $index(p) = index(\pi(p)) - 1$. Hence, $index(G_i^+) = index(G_{i-1}^+) - 1$ and in particular, $index(G_i^+) = index(G_0^+) - i = n - i$. □

Lemma 29. *Let x be a vertex in G_i^+ of index $r \geqslant 1$. Then $\pi^i(x)$ has index $i + r$ in G_0^+, for all $n \geqslant i \geqslant 1$. Moreover, if x is a focussing point in G_i^+ then $\pi^i(x)$ is a focussing point in G_0^+.*

Proof. To prove the statement we need to show that all vertices x in G_i^+ with $index(x) = j \geqslant 1$ have the property that $index(\pi_i(x)) = j + 1$ in G_{i-1}^+. By arguing as in Lemma 28 this is true for all vertices which are focussing in G_{i-1}^+. For all other vertices v one applies the definition of index and derives the statement by arguing on the index of the focussing vertices lying in some path between the input points in G_i^+ and v.

The second part of the statement follows from the fact that if x is a focussing point in G_i^+ then $\pi_i(x)$ is a focussing point in G_{i-1}^+. □

Notice that we cannot claim a similar statement for the vertices x of index 0 in G_i^+ because $\pi^i(x)$ can be of index 0 as well. Take for instance any input vertex x of G and notice that each counterimage of x in G_i^+ has index 0.

9.1. *An upper bound for positive resolution*

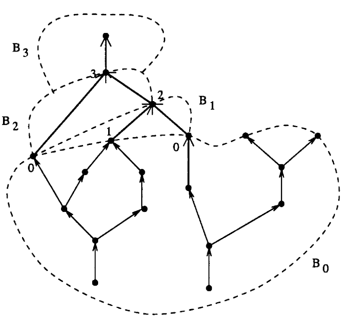
To prove upper and lower bounds for the size of G_n^+ we need to introduce the notion of *decomposition* for an acyclic optical graph. The definition is completely general and need not be applied to the specific graph G_n^+. The idea of this decomposition is to break a graph into 'nice' pieces $B_0,\ldots,B_n$ (where n is the index of the graph) which do not contain any focussing point. The upper bound to the size of G_n^+ will depend on the size of the subgraphs $B_0,\ldots,B_n$, the number of focussing points in G and the index of G.

We define $B_0,\ldots,B_n$ as follows. Let W_i be the set of vertices v in G which are either focussing points with $index(v)=i$, or immediate predecessors of focussing points y with $index(y)>i$ and $index(v)<i$, for $1\leqslant i\leqslant n$. Let W_0 be the set of inputs of G.

For all $0\leqslant i\leqslant n$, the graph B_i is the subgraph of G whose vertices are all nodes in G which lie in some path from a vertex in W_i to a vertex which is either in W_{i+1} or is an output vertex of index i in G. The edges of B_i are all the edges in G linking two of its vertices. The orientation of the edges is preserved.

The subgraphs B_i are usually *not* disjoint and they might be disconnected.

To see an example of decomposition let us consider the following graph:



where B_0 and B_1 are not disjoint and B_1 is disconnected.

Lemma 30. *Let x be a vertex in G_i^+, for $1\leqslant i\leqslant n-1$. The vertex x is an input vertex of G_i' iff $\pi^i(x)\in W_{i+1}$. Moreover, x is an input vertex of G_0' iff $x\in W_1$.*

Proof. Notice first that the input vertices x of G'_0 are exactly the nodes in W_1 by definition. Using Lemma 29 and by chasing the definitions we derive the claim for all $1 \leqslant i \leqslant n-1$. □

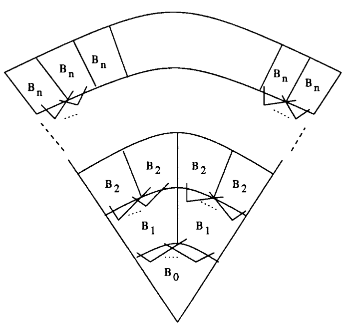
Theorem 31. *Let G be an acyclic optical graph. Let n be the index of G. Then*

$$|G_n^+| \leqslant \sum_{i=0}^{n} 2^i \cdot |B_i| + \sum_{i=1}^{n} 2^{i-1} \cdot |W_i|$$

where $B_0, \ldots, B_n$ is the decomposition of G_0^+.

Proof. Let us begin to notice that G_1^+ is built by patching together two copies of G'_0 and one copy of B_0. Each copy of G'_0 is constituted exactly by the subgraphs $B_1, \ldots, B_n$. (This is to say that all vertices in G'_0 belong to some B_i, for $i = 1, \ldots, n$.)

Inductively G_i^+ has been defined by patching together 2^i copies of G'_i, 2^{i-1} copies of B_{i-1}, 2^{i-2} copies of $B_{i-2}, \ldots, 2$ copies of B_1 and one copy of B_0. Each copy of G'_{i-1} is constituted exactly by the subgraphs $B_i, \ldots, B_n$. (This is to say that all vertices x in a copy of G'_{i-1} in G_i^+ are mapped by π^i into $\pi^i(x) \in B_{j-1}$, for $j = i, \ldots, n$. This is a consequence of Lemma 30.) The graph G_n^+ will look like



The way to patch together different copies of B_j's introduces at most $2^{i-1} \cdot |W_i|$ new vertices at the step i of the construction (by the observation above and Lemma 30).

This means that $|G_n^+|$ can be estimated as claimed. □

Some rougher estimates which we can deduce as corollaries of the theorem are the following

Corollary 32. $|G_n^+| \leqslant \sum_{i=0}^{n} 2^{i+1} \cdot |B_i|$

Proof. This is because $|W_i| \leqslant |B_i|$ (by definition of W_i). □

Corollary 33. $|G_n^+| \leqslant 2^{n+4} \cdot |G|$.

Proof. This follows from Corollary 32 and $|B_i| \leqslant |G_0^+| \leqslant 3 \cdot |G|$ (where the last inequality is Lemma 21). □

Corollary 34. *If for each focussing vertex x in G, $index(x) = index(x_1)+1 = index(x_2)+1$ where x_1, x_2 are immediate predecessors of x in G, then*

$$|G_n^+| \leqslant \sum_{i=0}^{n} 2^i \cdot |B_i|,$$

where n is the index of G and $B_0, \dots, B_n$ is the decomposition of G_0^+.

Proof. From the assumptions the sets W_i contain only focussing points of index i in G. This means that in the construction of Theorem 31 we do not have any new nodes which is introduced to patch together immediate predecessors of focussing points. This means that we can ameliorate the estimates of the theorem as $|G_n^+| \leqslant \sum_{i=0}^{n} 2^i \cdot |B_i|$. □

Corollary 35. *If for each focussing vertex x in G, $index(x) = index(x_1)+1 = index(x_2)+1$ where x_1, x_2 are immediate predecessors of x in G, then*

$$|G_n^+| \leqslant 2^{2k+4} \cdot r$$

where k is the height of G and r is the number of input vertices in G.

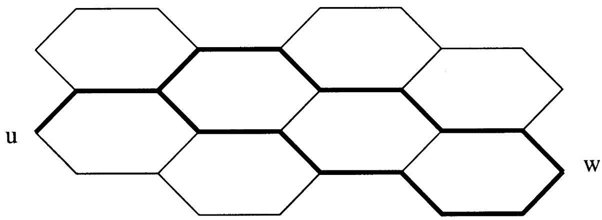
Proof. If k is the height of G and r its input vertices, it is a straightforward calculation to show that the size of G is $\leqslant 2^{k+1} \cdot r$. By Lemma 21 we also know that $|B_i| \leqslant |G_0^+| \leqslant 3 \cdot |G|$. By applying these estimates to Corollary 34 we derive the claim. (Strictly speaking we also use the fact that $k \geqslant n$.) □

9.2. An exponential lower bound for positive resolution

In this section we present the notions of focal pairs, long chain of focal pairs and visibility graph. These concepts have been introduced in [10] to study the exponential divergence which results from the branching within an optical graph G. Here we will use visibility graphs to explain the exponential expansion of resolution graphs and therefore of cut elimination.

Definition 36 (Carbone–Semmes)**.** Let G be an optical graph. By a *focal pair* we mean an ordered pair (u, w) of vertices in G for which there is a pair of distinct paths from u to w. We also require that these paths arrive at w along different edges flowing into w.

An example is illustrated below



where in thicker lines we see two different paths from u to w. The pair (u,w) forms a focal pair.

Definition 37 (Carbone–Semmes). By a *chain of focal pairs* we mean a finite sequence of focal pairs $\{(u_i,w_i)\}_{i=1}^{n}$ such that $u_{i+1}=w_i$ for each $i=1,2,\ldots,n-1$. We call n the *length* of the chain, and u_1 the *starting point* of the chain.

By applying the $\mathscr{D}$ operation we can create new focal pairs. This occurs every time there is a path in the subgraph G' which links an input vertex with an output vertex of G' which are not respectively focussing and defocussing points. This means that the length of a chain of focal pairs in a graph G can actually increase with repeated applications of the operation $\mathscr{D}$.

We introduce now the notion of visibility graph which roughly speaking can be thought to be the 'covering' graph of G at a point v. Intuitively the visibility graph represents the way that G looks from the perspective of v. Its properties (for both cyclic and acyclic graphs) have been thoroughly studied in [10].

Definition 38 (Carbone–Semmes). Let G be an optical graph and v be a vertex. The *visibility* $\mathscr{V}_+(v,G)$ of G from v is a graph whose vertices are the oriented paths of G which start at v. The "degenerate path" which consists of v alone, without any edge attached to it is included and represents a vertex of the visibility graph called *basepoint* of $\mathscr{V}_+(v,G)$. Two vertices p_1,p_2 in $\mathscr{V}_+(v,G)$ are connected by an edge oriented from p_1 to p_2 exactly when the corresponding paths in G have the property that the path associated to p_2 in G is obtained by the path associated to p_1 by adding one more edge in G at the end of it. We attach only one such edge in $\mathscr{V}_+(v,G)$ from p_1 to p_2 and these are the only edges that we attach.

The visibility graph is an oriented graph. It might be infinite, in case the optical graph G contains cycles, but it is *locally finite* however. In fact there is at most one edge going into any vertex and there are at most two edges coming out of any vertex. It is easy to show that the visibility graph is always a tree.

Proposition 39 (Carbone–Semmes). *Let G be an acyclic optical graph and let v be a vertex in G. Suppose that there are no paths in G starting at v with length greater than k. Then $|\mathscr{V}_+(v,G)|\leqslant 2^{k+1}$.*

Proof. (We follow the argument given in [10].) Let S_j denote the set of vertices in $\mathscr{V}_+(v,G)$ which can be reached from the basepoint v by an oriented path of length j, $j \geqslant 0$. We want to estimate the number N_j of elements of S_j. Notice that S_0 consists of only the basepoint, so that $N_0 = 1$. In general we have

$$N_{j+1} \leqslant 2 \cdot N_j$$

for all $j \geqslant 0$. Indeed, the definition of S_j ensures that for each element p of S_{j+1} there is a $q \in S_j$ such that there is an edge in $\mathscr{V}_+(v,G)$ that goes from q to p. There can be at most two p's corresponding to any given q, since $\mathscr{V}_+(v,G)$ is an optical graph, and the inequality above follows from this.

Thus we have that

$$N_j \leqslant 2^j \quad \text{for all } j$$

and the union $\bigcup_{j=0}^{k} S_j$ has at most 2^{k+1} elements. $\square$

Proposition 40 (Carbone–Semmes). *Let G be an acyclic optical graph of size n. Then $\mathscr{V}_+(v,G)$ has at most 2^n vertices.*

Proof. Since G is acyclic then $\mathscr{V}_+(v,G)$ has only finitely many vertices. By Proposition 39 we obtain the desired bound. $\square$

Proposition 41 (Carbone–Semmes). *Let G be an acyclic optical graph. Suppose that v is a vertex in G and that there is a chain of focal pairs in G starting at v and with length n. Then $2^n \leqslant |\mathscr{V}_+(v,G)|$.*

Proof. (We follow the argument given in [10].) Let $\{(u_i,w_i)\}_{i=1}^{n}$ be a chain of focal pairs in G which begins at v and has length n. It suffices to show that there are 2^n distinct vertices in the visibility $\mathscr{V}_+(v,G)$ which project down to w_n under the canonical projection. This amounts to saying that there are at least 2^n different oriented paths in G which go from $v = u_1$ to w_n. This is easy to see, since there are at least two distinct oriented paths α_i and β_i going from u_i to w_i for each $i = 1,\dots,n$, and there are 2^n different ways to combine the α_i's and β_i's to get paths from u_1 to w_n. $\square$

We will see now how the visibility graph corresponds in some precise way to a graph obtained by the elimination of focal pairs through positive resolution.

Theorem 42. *If G is an acyclic optical graph, then*

$$|\mathscr{V}_+(\pi^n(v),G)| \leqslant |\mathscr{V}_+(v,G_n^+)|$$

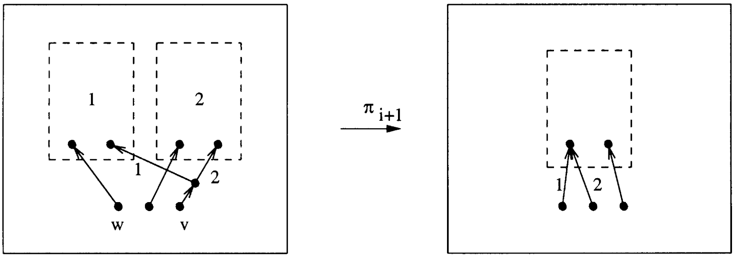
for all vertices v in G_n^+.

Proof. We show that the following holds:

$$|\mathscr{V}_+(\pi_{i+1}(v),G_i^+)| \leqslant |\mathscr{V}_+(v,G_{i+1}^+)|$$

where $\pi_{i+1} : G^+_{i+1} \to G^+_i$ is the canonical projection. Then by composing properly the inequalities we obtain the claim.

To show this inequality let us give the following symbolic picture of G^+_{i+1} (left) and G^+_i (right).



where the output vertices of G'_1, G'_2 are output vertices of G^+_{i+1} as well, and where the input vertices might or might not correspond to focussing points in G^+_i.

It is immediate to see that from the vertex v in the picture

$$|\mathscr{V}_+(v, G^+_{i+1})| = 2 \cdot |\mathscr{V}_+(\pi_{i+1}(v), G^+_i)| + 1$$

and from the vertex w we have

$$|\mathscr{V}_+(w, G^+_{i+1})| = |\mathscr{V}_+(\pi_{i+1}(w), G^+_i)|$$

Therefore the inequality above holds for all vertices x in G^+_{i+1} since the size of the visibility can only increase (this is easily seen by an inspection of the picture). □

From the proposition we just stated it follows that if we have a lower bound for the visibility of a vertex in G then the same lower bound holds for the size of G^+_n. In particular, if the visibility of a vertex in G has exponential size in the size of G, then the positive resolution G^+_n should have exponential size.

Theorem 43. *If G is an acyclic optical graph, then*

$$|\mathscr{V}_+(v, G^+_n)| \leqslant |G^+_n|$$

for all vertices v in G^+_n.

Proof. The graph G^+_n does not contain any focussing point, therefore given any vertex v in G^+_n there is *only one path* from v to a vertex y in G^+_n. This means that there is a bijection from the vertices in $\mathscr{V}_+(v, G^+_n)$ and the vertices in G^+_n which are accessible from v. This bijection can be extended in a natural way to a bijection on edges and it is an isomorphic embedding of $\mathscr{V}_+(v, G^+_n)$ into G^+_n. Hence the claim. □

Corollary 44. *If G is an acyclic optical graph, then*

$$|\mathscr{V}_+(v, G)| \leqslant |G^+_n|$$

for all vertices v in G.

Proof. We combine Theorems 42 and 43. □

Theorem 45. *If for each focussing vertex x in G, $index(x) = index(x_1) + 1 = index(x_2) + 1$ where x_1, x_2 are immediate predecessors of x in G, then*

$$\mathscr{V}_+(\pi^n(v), G) \simeq \mathscr{V}_+(v, G_n^+)$$

for all vertices v in G_n^+.

Proof. From the hypothesis that $index(x) = index(x_1) + 1 = index(x_2) + 1$, for all focussing points x and immediate predecessors x_1, x_2 of x in G, we know that the number of input vertices in each G_i^+ is preserved by duplication. In fact no input vertices of G_i^+ belongs to G_i', for $i = 0, \ldots, n-1$, because an input vertex has index 0 and there are no input vertices that can be immediate predecessors of some focussing vertex of index $i \geqslant 1$.

In particular, for all $i = 0, \ldots, n-1$, all input vertices of G_i' are focussing points. Their output points are the outputs of G_i^+ by construction. Therefore the duplication induced by $\mathscr{D}(G_i', G_i^+)$ preserves the paths as stated in Proposition 11 and no new defocussing or focussing point is introduced in $\mathscr{D}(G_i', G_i^+)$. In particular,

$$\mathscr{V}_+(\pi_i(v), G_i^+) \simeq \mathscr{V}_+(v, G_{i+1}^+)$$

for all v in G_{i+1}^+.

In fact, if v is in G_i' then $\mathscr{V}_+(v, G_{i+1}^+)$ is $\mathscr{V}_+(\pi_i(v), G_i^+)$ since the portion of the graph visited by the visibility is the same for G_i' and G_i^+ by construction.

If v is in $G_{i+1}^+/G_{i,1}' \oplus G_{i,2}'$ then given a path p in G_{i+1}^+ from v to y, and the corresponding path $\pi_i(p)$ in G_i^+ from $\pi_i(v)$ to $\pi_i(y)$, we claim that if q is the path obtained from p by adding an edge from y to y', then there is a path $\pi_i(q)$ in G_i^+ obtained from $\pi_i(p)$ by adding an edge $\pi_i(\varepsilon)$ from $\pi_i(y)$ to some vertex $\pi_i(y')$. This follows by construction of G_{i+1}^+. The only difficult point is given by the fact that in G_{i+1}^+ there might be pairs of distinct points y_1, y_2 which map into the same vertex y in G_i^+. This does not disturb the visibility though. □

From Proposition 41 we can derive an *exponential* lower bound on $\mathscr{V}_+(v, G_n^+)$.

Theorem 46. *Let G be an acyclic optical graph and v be a vertex in G such that there is a chain of focal pairs in G starting at v and with length n. Then*

$$2^n \leqslant |\mathscr{V}_+(y, G_n^+)|$$

for all vertices y such that $\pi^n(y) = v$.

Proof. From Proposition 41 we obtain the inequality $2^n \leqslant |\mathscr{V}_+(v, G)|$ and from Theorem 42 we derive the inequality $|\mathscr{V}_+(v, G)| \leqslant |\mathscr{V}_+(y, G_n^+)|$, for all vertices y such that $\pi^n(y) = v$. By combining the two inequalities we derive the claim. □

9.3. An exponential upper bound for the visibility

In this section we derive an upper bound for the size of a visibility graph from the upper bound proved in Section 9.1.

For any acyclic optical graph G and any vertex v in G let us consider the subgraph G_* of G with input v constituted by all vertices y having a path from v to y, and all edges connecting these vertices.

Let n be the maximal index of a path in G starting at v. Notice that $n = index(G_*)$ because v is the only input vertex of G_*. In particular, the inequality $index(G_*) \leqslant index(G)$ holds.

The graph G_* satisfies the assumptions of Corollary 34 and therefore

$$|G^+_{*,n}| \leqslant \sum_{i=0}^{n} 2^i \cdot |B_i|,$$

where the graph $G^+_{*,n}$ is the positive resolution of G_* and the B_i's define the decomposition of $G^+_{*,0}$ as in Theorem 31 as in Section 9.1. The graph $G^+_{*,n}$ is the visibility of v in G_* and in particular it is the visibility of v in G.

Hence we proved the following:

Theorem 47. *If G is an optical graph and v is one of its vertices, then*

$$|\mathscr{V}_+(v, G)| \leqslant \sum_{i=0}^{n} 2^i \cdot |B_i|$$

where n is the maximal index of a path in G starting at v.

Corollary 48. *If G is an optical graph and v is one of its vertices, then*

$$|\mathscr{V}_+(v, G)| \leqslant 2^{n+2} \cdot |G|$$

where n is the maximal index of a path in G starting at v.

Proof. The statement follows from Theorem 47 and the fact that $|B_i| \leqslant 2 \cdot |G|$. (Remember that the B_i's are defined with respect to G_0^+, where $|G_0^+| \leqslant 2 \cdot |G|$ by the Blowing lemma.) □

This last statement greatly improves the upper bound given in [10]. Moreover one can check the upper bound in Theorem 47 is sharp for a graph G which is a tree (since $index(G) = 0$, and B_0 is G in this case), and for the chain of focal pairs illustrated in the first example of Section 5 (where each $|B_i| = 3$ for $i = 0, \ldots, n-1$, $|B_n| = 1$, and $|W_i| = 1$ for all $i = 0, \ldots, n$).

10. Negative resolution and negative visibility

In the previous sections we confined ourselves to what one sees in a graph from a vertex v in the direction of *positive* orientation. One could just as well look at the *negative* orientation and define notions of negative resolution and negative visibility instead. Here defocussing branch points instead of focussing ones will be considered.

Let *anti-index*(v) be the number of defocussing points from some output vertex in G to the vertex v. Let *anti-index*(G) be the maximum among the anti-indexes of the points in G. We now define the *negative resolution* of G as the graph G_n^- obtained by the following inductive steps:

$$G_0^- = Blow(G)$$

$$G_{i+1}^- = \mathscr{D}(G_i', G_i^-)$$

where the graph G_i' is defined as for G_{i+1}^+, but where the role of the index is played here by the anti-index, inputs by outputs, and focussing vertices by defocussing vertices.

The intuition behind the notion of negative resolution is the same as for the positive resolution. One wants to eliminate all defocussing points from the original acyclic optical graph G and wants to do this by eliminating first the defocussing points of anti-index 1. (Notice that the defocussing points of anti-index 1 in G_i^- are the closest to the output points of G_i^-, as the points of index 1 in G_i^+ were the closest to the input points of G_i^+.)

The results in the previous sections can be restated for negative resolution and we will not do it explicitly here. We want to notice however that the results in Section 9.2 will be dependent on the notion of *negative visibility* $\mathscr{V}_-(v, G)$ which is the counterpart for defocussing points of the visibility graph. (Instead of looking at the positive orientation, one looks at the negative one.)

A slight modification of the strategy which builds positive resolutions consists in considering those subgraphs G_i' which are only required to contain *all* defocussing points in G. In this way, the output vertices of G_i' might be vertices which are not output for G_i^+. Hence, this notion gives a weaker version of the construction of positive resolution and we will refer to it as *positive weak resolution*. The graphs in the third picture of Section 5 illustrates this construction. (We already invited the reader to compare these graphs with the ones in the second picture of the section which correspond to positive resolution.)

An analogous version of *negative weak resolution* is easily defined.

11. Formal proofs and optical graphs

In this section we first introduce a basic interpretation of optical graphs in logical terms. The first result we present shows that for each *acyclic* optical graph one can always find a proof with the same topological structure. Then we will show a series of facts which will emphasize how *special* are logical flow graphs of proofs compared to optical graphs in general. In the next section we interpret *duplication* as the combinatorial counterpart of the operation of cut-elimination which eliminates *contractions*.

Let us start by looking for a proof with a logical flow graph having the same structure of branching points as a given optical graph. We show that it is always possible to find such a proof.

Theorem 49. *For any acyclic optical graph G there is a proof having a logical flow graph with the topological structure of G. Moreover, if a defocussing point in G is followed by a focussing one then the proof will contain a cut on some atomic formula.*

Proof. Let G be an acyclic optical graph. We assume first that G is connected. (The case of G disconnected is treated later.) Let G' be the minimal optical graph with the same topological structure as G, whose existence is claimed by Proposition 16. We will build a proof Π whose logical flow graph has the same topological structure as G' (and G). In particular, focussing points in G' will correspond to F:*times* rules and defocussing points to contractions on the left. We will not have contractions on the right but we will possibly make use of $\wedge$: *right* rules.

The proof Π is built by induction on the height of the subgraphs of G' starting from its input vertices. Let n be the height of G'. The idea is to associate to the subgraphs of G' of height $k \leqslant n$ a set of proofs that will be combined along the construction to form the proof Π. (Notice that there might exist several connected subgraphs of G' of height k.) At stage $k = n$ of the construction we will end up with a proof with the same topological structure as G'.

The end-sequents of the proofs that we build along the construction are of the form $\Gamma \rightarrow F(t)$, for some term t and some collection Γ of formulas of the form $F(s)$. The occurrence $F(t)$ on the right hand side of the end-sequent will be associated to some vertex of the optical graph G'.

Along the construction we mark the formulas in the proof in such a way that they will be related in a unique way to a labelling of the optical graph. More precisely, to each output vertex in a subgraph of G' of height k we assign a formula as a label in such a way that there will be exactly one proof in the set built at stage k which has the same formula appearing on the right hand side of its end-sequent.

Let us enter now into the details of the construction. We consider first all subgraphs of G' of height 1, that is all the input vertices.

To each input vertex in the optical graph we associate a label $F(2)^{(i)}$ and an axiom of the form $F(2) \rightarrow F(2)^{(i)}$, where i is an integer and the superscript (i) denotes the right occurrence of the formula $F(2)$. To each input vertex we assign a *different* value i in such a way that if there are l input vertices in G', i will vary from 1 to l.

Suppose now that all subgraphs of G' of height $k < n$ have been considered. We want to assign a set of proofs to all subgraphs G^* of G' of height $k + 1$ and a label to the vertices of height $k + 1$. This will be done in such a way that for each output vertex of the subgraphs there is exactly one of the proofs in the set whose formula lying in the right hand side of its end-sequent is identical to the label of the vertex. To do this we will consider one by one all vertices in the G^*'s of height $k+1$. When all of them have been considered we will be finished.

There are three possible situations that might occur.

Let v be a vertex of height k in G^* and label $F(t)^{(j)}$. Suppose that v is a defocussing point in G^*. Its two immediate successors in G^* have height $k+1$ and we label them $F(t)^{(r)}$ and $F(t)^{(r+1)}$, where r is the first integer value that has not yet being used in the construction. We also add to the set of proofs two new axioms of the form $F(t)^{(j)}_{(r+2)} \to F(t)^{(r)}$ and $F(t)^{(j)}_{(r+2)} \to F(t)^{(r+1)}$ and we change the label $F(t)^{(j)}$ both in v and in the unique formula labelled $F(t)^{(j)}$ which appears in the right-hand side of the end-sequent of some proof obtained by construction, into the new label $F(t)^{(j)}_{(r+2)}$. (Notice that by doing this the vertices of height $k+1$ in G^* have no subscript in their labels. Just a superscript.)

Suppose that v is a vertex of height $k+1$ and it is a non-branching point in G^*. Then we label it with the label of its immediate predecessor. Nothing will be done to the set of proofs.

Suppose that v is a vertex of height $k+1$ and suppose that v is a focussing point in G^* with its immediate predecessors labelled by $F(t)^{(j)}$ and $F(s)^{(l)}$. We label v with $F(t*s)^{(r)}$, where r is the first integer value that has not yet being used in the construction. We then combine with an $F:times$ rule the two proofs having the right-hand side formula in their end-sequents labelled by $F(t)^{(j)}$ and $F(s)^{(l)}$. These proofs are unique by construction and they are distinct since j and l are distinct. The end-sequent of the proof is of the form $\Gamma \to F(t*s)^{(r)}$ for some collection Γ.

By applying the $F:times$ rule, we might end-up to have on the left-hand side of the end-sequent a pair (maybe more than one) of formulas with the same subscript, say $F(t)^{(i)}_{(k)}$. In this case, we *contract* the two occurrences $F(t)^{(i)}_{(k)}$ and we *cut* the resulting formula with the only proof in the set which has $F(t)^{(i)}_{(k)}$ lying in the right-hand side of its end-sequent. (This proof exist because the subscript (k) can be introduced only after having built it. In particular, it was never used to combine it through a $F:times$ rule with some other proof, because it is denoted by a subscript.)

We repeat this construction for all the pairs of formulas with a common subscript which occur on the left-hand side of Γ.

At the end of the construction (i.e. when all the vertices of height n have been analyzed) we might end-up with a set of subproofs instead of just one. Each one of these subproofs must contain some formula (in the end-sequent) which is denoted by a subscript. In particular, for any such formula $F(t)^{(i)}_{(k)}$ the set should contain exactly *two* subproofs where the formula occurs on the left-hand side of the end-sequents and *one* subproof where the formula occurs on the right-hand side of the end-sequent. These properties are easily checked by induction on the construction.

We want to combine the subproofs in the set with $\wedge:right$ rules to obtain one single proof and we proceed in such a way that the properties above will remain true at each step of the construction. We look at those subproofs in the set whose end-sequent contains on its right-hand side a formula denoted by a subscript, and we start by considering the formula with the largest k, say $F(t)^{(i)}_{(k)}$. By construction there exists two distinct subproofs in the set with $F(t)^{(i)}_{(k)}$ lying on the left-hand side of their end-sequents. (In fact, if the two occurrences where lying in the end-sequent of one single proof they would have already been contracted and cut by some previous step of the construction.) We combine these subproofs via an $\wedge$: *right* rule. On the left-hand side of the end-sequent of the new proof we *contract* the pairs of formulas $F(t)^{(i)}_{(k)}$ and we *cut* the resulting formula with the only proof in the set which has $F(t)^{(i)}_{(k)}$ lying on the right-hand side of its end-sequent. We contract and cut also *all* other pairs of identical formulas which are denoted by a subscript and lie on the left-hand side of the end-sequent.

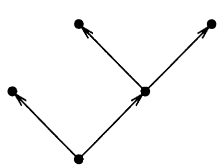
We repeat this step until there are not anymore subproofs whose formula lying on the right-hand side of the end-sequent is denoted by a subscript. This means that we remained with *one* proof only.

To show that the proof we have built has the same topological structure as G' we should notice that contractions and F: *times* applications in Π correspond to those formulas which label branching points of G'. If a branching point in G' is focussing, by construction the F: *times* rule is applied and the correspondence is obvious. If the branching point is defocussing, the construction is a bit more complicated. In fact it introduces two axioms whose formulas on the left are labelled by the same subscript. Later in the construction this pair of formulas will end-up lying in the same sequent. This is ensured by the fact that we will end-up with only one proof at the end and that this proof shall contain as subproofs all the subproofs built along the different stages of the construction. The construction will *contract* the two formulas having the same label and create a defocussing point as desired.

If G is disconnected, we apply the construction described above on all its connected components and we combine all the proofs associated to them with $\wedge$: *right* rules. The logical flow graph of the resulting proof will be clearly disconnected.

This concludes the proof of the first part of the statement. To check the second part is easy. We notice that by construction a focussing point corresponds to the application of an F: *times* rule to some positive occurrence and a defocussing point corresponds to the application of a contraction to negative occurrences. Since negative and positive occurrences should be linked by the assumption on the graph G, this means that there should be a cut-edge along the path in the logical flow graph of the proof going from the focussing point to the defocussing one.

(Notice that the construction asks for a cut-edge also between points in a proof corresponding to configurations as the following



The proof that is produced by this configuration can be simplified to a cut-free one which has the same topological structure as G. The checking of this remark is tedious and we skip it. The main point we want to make here is that no simplification of cuts can be performed to eliminate the cut-edge corresponding to a focussing point followed by a defocussing one, without disrupting the topological structure of the logical flow graph.) $\square$

The proof of the last statement associates to an acyclic graph which contains focussing points a proof *with* cuts on atomic formulas. The presence of focussing points in the graph allows the *implicit* representation of parts of the graph. In fact, by the point of view of the immediate predecessors x, y of a focussing point v in a graph G, whatever follows v is the 'same', in the sense that

$$\mathscr{V}_+(x, G) \simeq \mathscr{V}_+(y, G)$$

More generally, given any two vertices v, w in G we say that G is an *implicit representation* of $\mathscr{V}_+(v, G)$ and $\mathscr{V}_+(w, G)$ when there are paths starting from v and w in G which overlap. Clearly, the overlapping is possible only when v and w have access to a 'common' subgraph of G.

This implicitness can be realized in a proof only through cuts. This aspect of implicitness versus explicitness in graphs of proofs is a reminder of the transformation of circuits into boolean expressions (see [18]).

Theorem 50. *Let $b_1, \ldots, b_k$ be boolean operations and $\mathscr{C}$ be a circuit of size n with inputs $x_1, \ldots, x_r$ defined over $b_1, \ldots, b_k$. There is a proof Π of size $\mathcal{O}(n)$ with cuts on atomic formulas which is formulated in the propositional part of the predicate calculus LK extended with the rules*

$$\frac{\Gamma_1 \to \Delta_1, F(a_1) \ \ldots \ \Gamma_{n(i)} \to \Delta_{n(i)}, F(a_{n(i)})}{\Gamma_{1,\ldots,n(i)} \to \Delta_{1,\ldots,n(i)}, F(b_i(a_1, \ldots, a_{n(i)}))}$$

where $i = 1 \ldots k$ and $n(i)$ is the arity of b_i. Moreover, the underlying graph of $\mathscr{C}$ and the logical flow graph of Π have the same topological structure.

Proof. The proof follows the same lines as the proof of Theorem 49, where the binary $F : times$ rule is replaced now by $n(i)$-ary rules. If for all i, the arities $n(i)$ equal 2, the logical flow graph of a proof in the new calculus would be an optical graph. If $n(i) > 2$ for some i, then one should be careful to extend the notions of optical graphs and topological structure to focussing points of in-degree $n(i)$. This is an obvious extension and there are no serious technical problems to overcome. $\square$

Given an optical graph with cycles there might be no proof with the same topological structure.

Theorem 51. *It is not true that for each optical graph with cycles there is a proof with the same topological structure.*

Proof. Take any optical graph containing a cycle with no way in. Let us suppose that this is the only cycle in the graph and that there is a proof with the same topological structure. By Theorem 25 we know that the cycle cannot be eliminated by duplication. On the other hand, as argued in Section 12 we also know that the procedure of cut elimination transforms a graph of a proof into a new one with the same topological structure except when cuts over contractions are eliminated. In this last case the operation of cut elimination corresponds to duplication over logical flow graphs of proofs, which cannot disrupt the cycle. This gives a contradiction with the fact that cut-free proofs are always acyclic. (We have been a bit sloppy here in our argument concerning cut elimination but the reader can find the details of the transformation in combinatorial terms in Section 12.) □

Remark 52. Proofs with cycles must contain cuts of logical complexity greater than 1. In fact any cyclic path in a proof needs to pass through a cut on a formula which contains a binary logical connective.

Proposition 53. *Let Π be a cut-free proof. There is no focussing point in the logical flow graph of Π which is followed by a defocussing one. In particular, there are no chains of focal pairs of length greater than* 1.

Proof. A focussing point in the logical flow graph of a proof corresponds to a contraction over positive occurrences and a defocussing point corresponds to a contraction over negative occurrences. This means that in the logical flow graph of a proof, a focussing point is followed by a defocussing one only if there is an edge which links positive occurrences to negative ones. This type of edges are the cut-edges and since Π is cut-free the first part of the claim follows. To see the second part of the claim it is easy. We need to notice that a focal pair can follow another focal pair only if the defocussing point of the second follows the focussing point of the first. But for cut-free proofs, this is forbidden by the first part of the statement. □

When we eliminate cuts from a proof we transform it into a proof whose logical flow graph is an optical graph constituted by several (maybe one) acyclic connected components. Each of these components has several inputs and several outputs but no chains of focal points of length greater than 1. They are H-graphs.

To conclude this section let us add some remarks concerning logical flow graphs that contain cycles. In the next section we will show that the operation of duplication $\mathscr{D}$ is the combinatorial formulation of the operation of duplication of subproofs which takes place by treating contractions in cut elimination. In particular, we will see that the topological structure of the logical flow graph of a proof changes during cut elimination exactly when cuts over contractions are eliminated. Bearing this in mind, we can proceed with our remarks.

In Section 8 we have seen how the labelling of an optical graph plays a crucial role in the elimination of cycles. Theorem 25 emphasizes that *not all* labellings of an optical graph would lead to the elimination of its cycles by duplication. Since Gentzen

Cut Elimination theorem ensures that there is a sequence of steps of elimination of cuts which leads to a cut-free proof and since logical flow graphs of cut-free proofs are acyclic [3], we conclude that the logical formalism *LK induces* a labelling of the logical flow graph of a proof which leads to the elimination of cycles through duplication. (The formal check of this assertion is left to the reader.) This is a remarkable point which emphasizes the special nature of the graphs built through formal logical rules.

Another amusing point is that formal logical rules allow to build only those cycles which have both a way in and a way out.

Theorem 54. *Let Π be a proof and L a loop in its logical flow graph. Then L has both a way in and a way out.*

Proof. Let us consider only those procedures of cut elimination which eliminate weak cut-formulas from a proof only at the end. With this restriction we do not lose any generality and moreover we avoid to consider the case where the disruption of a cycle is induced by the cancelation of an entire subproof.

A proof containing cuts on weak formulas only are acyclic. This is easy to prove. It was shown in [4] and the argument exploits the fact that logical paths passing through weak cut-formulas end-up into weak occurrences in axioms and they have no possibility to turn around on axioms as it would be required by a cyclic path. Therefore, this means that L should have been disrupted by duplication at some point during the process of cut elimination (before arriving at the stage where cuts on weak formulas are eliminated). But by Theorem 25 this is possible only if L has both a way in and a way out. □

Remark 55. The logical flow graph of a proof always satisfies the conclusions of the Blowing lemma (Lemma 21), i.e. there is no edge between a defocussing and a focussing point. To see this notice that defocussing points in a logical flow graph are associated to negative occurrences of formulas in a proof. Similarly, focussing points are associated to positive occurrences. Moreover, the only edges in a logical flow graph which connect negative to positive occurrences are axiom-edges. It is routine to check that these occurrences of formulas in an axiom cannot possibly be branching points.

12. Cut elimination and duplication

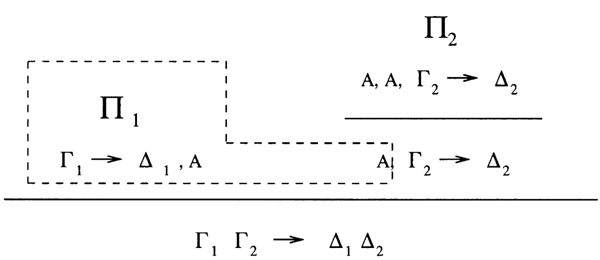
The operation of duplication given in Definition 4 is the combinatorial formulation of the operation applied to logical flow graphs of proofs when cut eliminating *contractions*. All steps of cut elimination except the ones involving contractions, transform a graph into another with the same topological structure. They deform, stretch, shrink the logical flow graph of a proof without disrupting the branching points and/or adding new ones. There is only one exception to this. It is the step dealing with cut-formulas which are *weak* which can disrupt the logical flow graph by cancelling out part of it. We disregard this exception because this step can always be postponed in the process of cut-elimination and performed at the very end of it.

Before treating explicitly the elimination of contractions as a phenomenon of duplication, let us give a closer look to the topological transformations of a graph during cut elimination.

We first consider the elimination of cuts on axioms. Going from (1) to (3), paths are shrunk or extended but that is all. The passage from (2) to (3) is more delicate. In fact with this step one removes a possibly large part of the logical flow graph. The cancelation of the subproof Π_0 can lead to the breaking of cycles, or to the breaking of connections between different formula occurrences in the proof. Chains of focal pairs can be broken and focal pairs can be removed.

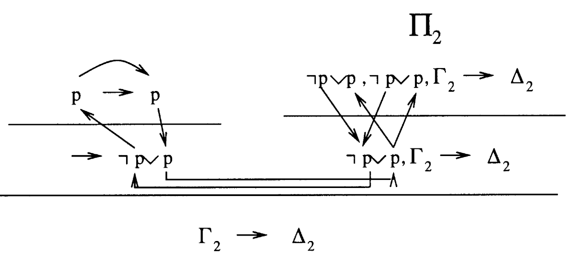
For the case of a cut over formulas introduced by logical rules, going from (4) to (5) requires the shrinking or the extension of paths and the topology of the logical flow graph is not altered. This is the case for all logical connectives. The same holds for quantified formulas while passing from (6) to (7) and for the exchange of cuts with other rules of inference while pushing cuts up in the proof (from (10) to (11)).

Let us now consider the case of contractions. We claim that the passage from (8) to (9) coincides with the operation of duplication introduced in Definition 4. Let G be the optical graph of a proof and suppose that $\mathscr{D}(G', G)$ is the result of the operation $\mathscr{D}$ applied to a proof like (8) with the purpose to eliminate a cut on the contracted formula A. The graph G' corresponds to the subgraph of Π_1, i.e. the subproof we want to duplicate, together with the cut-edges involving the cut-formulas lying in Π_2 in (8) as illustrated by the picture below



The input and output vertices of G' which are focussing and defocussing branch points correspond to negative and positive occurrences of formulas in the contracted cut-formula. All other input and output vertices of G' correspond either to weak formulas or to formulas passing through the side formulas Γ_1, Δ_1.

Let us take a specific proof which illustrates well what we just explained.



In the picture the graph G' has an input vertex which is a focussing point (this corresponds to the negated occurrence of p in $\neg p \vee p$) and an output vertex which is a defocussing point (this corresponds to the positive occurrence of p in $\neg p \vee p$). There are no other input or output vertices in G' and this corresponds to the absence of side formulas Γ_1, Δ_1 in Π_1.

In general a contracted formula might contain occurrences which are not connected through logical paths. For instance, any two distinct propositional letters will have no path connecting them, and this is also true for first order proofs. This means that through the procedure of cut elimination several connected logical flow graphs evolve in *parallel* until all of them have been transformed into H-graphs, i.e. graphs with no focussing point followed by a defocussing one. We are not interested in the interactions of the distinct connected components of a logical flow graph, but rather in the evolution of each of them. In particular we want to measure the size of an H-graph with respect to the geometric properties of the original proof. At this purpose, let us recall Lemma 6.30 of [10].

Lemma 56 (Carbone–Semmes). *Let $\Pi : S$ be a cut-free proof. Let A be the number of axiom-edges in Π, p the number of positive occurrences of atomic formulas in S and n the number of negative occurrences of atomic formulas in S. Then there exists atomic occurrences P, Q in S with P positive and Q negative such that there are at least*

$$\frac{A}{pn}$$

distinct bridges in the logical flow graph of Π which go from Q to P.

Proof. (We follow here the argument given in [10].) Since Π is cut-free we know that the total number of distinct bridges in the logical flow graph of Π equals A. Each of these bridges will go from a negative occurrence in the end-sequent to a positive one. The existence of P and Q as above then follows immediately from a simple counting argument. That is, if P and Q did not exist, then the total number of bridges would have to be strictly less than A, a contradiction. □

Corollary 57. *Let $\Pi : S$ be a cut-free proof whose size is exponential in the size of S. Then the logical flow graph of Π contains a subgraph of exponential size which is an H-graph.*

Proof. From Lemma 56 we know that there is an exponential number of bridges starting from an occurrence Q in S and going to an occurrence P in S. Take the connected subgraph of the logical flow graph of Π which contains these bridges. It is easy to see that it is an H-graph since H is cut-free. Moreover the number of axioms in Π has to be exponential since there are exponentially many bridges starting in P and ending in Q (here we use again the fact that Π is cut-free). Therefore the size of the H-graph is exponential. □

This means that it is *necessary* for the logical flow graph of a proof to develop into a subgraph of exponential size, whenever its cut-free forms are of exponential size. Hence we can hope to use the combinatorial notions introduced in the previous sections to prove that proofs with certain geometrical properties need to evolve into *large* cut-free proofs.

From the examples in Section 5 we have seen that the process of cut-elimination might have no end, depending on the choice of the subproof we want to duplicate. This implies that there are in general infinitely many cut-free proofs which can be generated from the same proof with cuts (the same way as infinitely many H-graphs can be obtained from the same optical graph).

We have also seen how cut-free proofs might reach an arbitrarily large size, compared to the size of the original proof with cuts. We know that cut-free proofs in first-order logic for instance, need to be super-exponentially larger than certain proofs containing cuts (see [16, 17, 22–25]). Our model explains that this is possible. Yet we do not have a concrete treatment of the geometric properties of first order proofs. It is reasonable to think that for first order logic, *binary* duplications will be forced into *k-ary* duplications and that purely geometric considerations will explain super-exponential expansion as well. The conviction here is founded on the result of [5] where it is shown that a proof with cycles can always be reduced to an acyclic proof only elementary larger than the original proof, by changing slightly the proof system. The new proof system (called ALK, being the acyclic version of LK) contains k-ary versions of the quantifier rules which during cut elimination force to duplicate k times the same subproof.

Before concluding this section and start with a thorough analysis of the evolution of focal pairs through the procedure of cut elimination, let us add a few remarks on proofs containing contractions only on atomic formulas. Their logical flow graph is quite simple. In fact it is always acyclic (there are no paths which go back to the same formula and that can help to create a cycle). During cut elimination the subgraphs G' associated to subproofs that have to be duplicated have either an input which is focussing or an output which is defocussing, and for each G' there is only one such a vertex. This is because the contracted cut-formulas are atomic. This implies that for these proofs, we do not need to work with *labels* on optical graphs since there is no matching that should be realized between input and output vertices in G'. Also, for these proofs there is always the possibility to obtain a cut-free proof by eliminating cuts only on those contracted cut-formulas which appear positively in the proof. (Similarly, one can eliminate cuts on contracted cut-formulas which appear negatively and leave inaltered the contractions on positive occurrences.) It is easy to see that this strategy corresponds essentially to the construction of a graph G_n^+ (G_n^-) and roughly speaking it corresponds to the construction of positive (negative) visibilities.

To see a concrete example of this point, let us go back to the proof discussed in Section 3. The only contractions there appeared on the left hand side of the subproofs Π_i. Assume to eliminate first the contraction lying in Π_i and afterwards the contraction on Π_{i+1}, for all $i=2,\ldots,n$. This strategy corresponds to a weak strategy of duplication applied to defocussing points and ultimately gives a negative resolution graph of exponential size as a result.

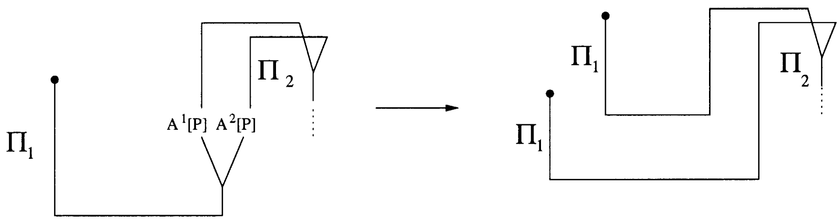
13. Duplication and chains of focal pairs

During the procedure of cut elimination, chains of focal pairs are shortened by duplications and new pairs of focal points might be added to the logical graph of a proof. This discussion started in [3, 4, 9, 10] and we return to it here.

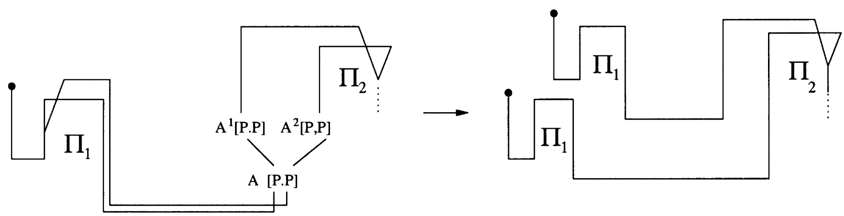
We analyze how focal chains evolve along the transformation from (8) to (9) while cuts on contractions are eliminated, and we study the formation of patterns in logical flow graphs and their rate of growth. In Sections 13.1 and 13.2 we analyze how focal pairs are removed or redistributed in the logical graph by duplications and in Section 13.4 we discuss how new pairs might be formed.

13.1. Removal of focal pairs

Imagine that there is a node in Π_2 of (8) from which a pair of oriented paths emerges and goes through Π_2 until the two paths reach A^1 and A^2, respectively. From there the paths will proceed down through the contraction and across the cut into Π_1. Let us assume that the two paths either converge together at the contraction of A^1 and A^2

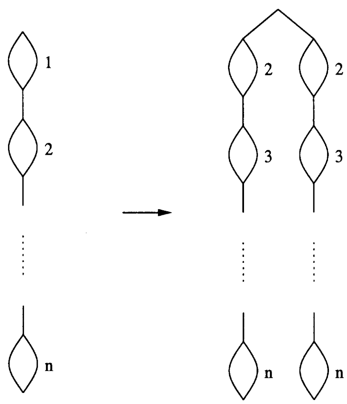


or later in the subproof Π_1



In both cases we assume that the paths end in weak occurrences in the axioms in Π_1. In either case the convergence of the two paths would be broken in the passage from (8) to (9). In particular we would lose the focal pair that we had in the logical flow graph of the original proof.

Suppose now that our pair of paths passes in Π_1 through a chain of focal pairs of length $n-1$ before ending in a weak occurrence in an axiom. In this case we would have a chain of focal pairs of length n in the original proof as a whole, because of the focal pair which begins in Π_2. This chain would not persist after the duplication of the subproofs, but instead we would have two copies of the chain of length $n-1$ from Π_1 after the duplication of subproofs. These two copies would diverge from each other as illustrated in the following picture

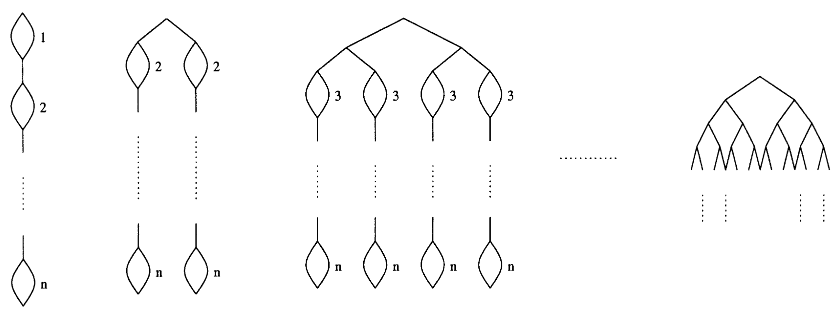


We can see this type of phenomenon concretely by rearranging slightly the first example of Section 3. Suppose we want to prove $\rightarrow F(2^{2^n})$ from the axiom $\rightarrow F(2)$. One can do that by putting together the building blocks

$$\dfrac{\begin{matrix}\Pi_1 \\ \rightarrow F(2^{2^{j-1}})\end{matrix} \quad \dfrac{\dfrac{F(2^{2^{j-1}}) \rightarrow F(2^{2^{j-1}}) \quad F(2^{2^{j-1}}) \rightarrow F(2^{2^{j-1}})}{F(2^{2^{j-1}}), F(2^{2^{j-1}}) \rightarrow F(2^{2^j})}\ F: times}{F(2^{2^{j-1}}) \rightarrow F(2^{2^j})}}{\rightarrow F(2^{2^j})}$$

The occurrence of $F(2)$ in the axiom $\rightarrow F(2)$ behaves somewhat like a weak occurrence, in the sense that paths in the logical flow graph end there and have nowhere else to go.

If one takes the proof with cuts and simplifies all of the cuts over the contractions (i.e. from the bottom of the proof to its top, following (8) and (9)), then one gets in the end a nice binary tree of exponential size in n. That is, one has uniform binary splitting of the branches until almost the very end, where one picks up linear graphs associated to $\rightarrow F(2)$. The evolution from the chain of interesting pairs to a tree (of exponential size) is illustrated below

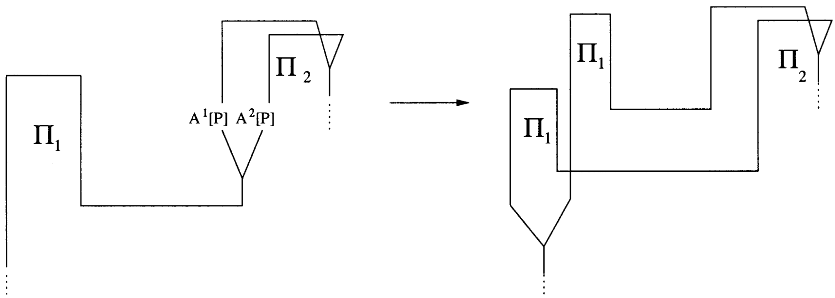


In the terminology of our combinatorial model we obtained the graph G_n^-.

13.2. Redistribution of focal pairs

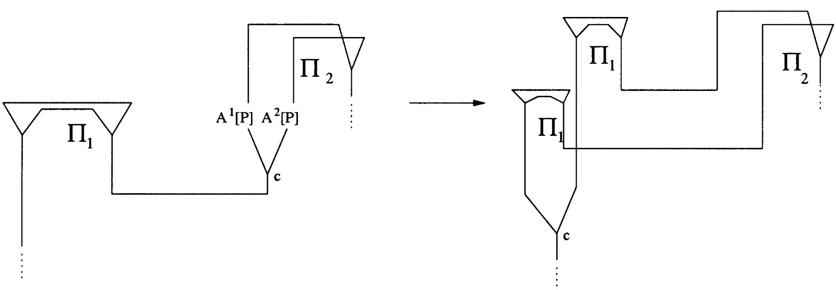
Imagine that we have a pair of paths which begin at some common starting point in Π_2 of (8) and which reach A^1 and A^2 in the contraction. We suppose first that the paths converge to the same point once the contraction is performed. At the end of this section we will consider the case where paths do not converge immediately but wait to converge once they are in Π_1. The phenomenon we will observe is the same.

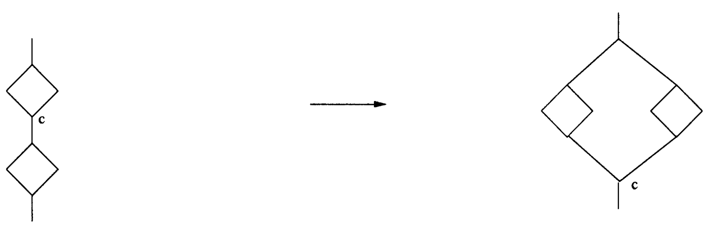
Under our first assumption, the two paths continue along a common trajectory into Π_1 down to a formula in Γ_1 or Δ_1 in the endsequent of Π_1. After duplication the two paths will be reunited in the contractions that occur below the new cuts



In this case the duplication of subproofs would not break apart the original focal pair in (8), it would simply postpone the convergence of the paths.

This kind of process would disrupt a *chain* of focal pairs, however. Suppose now that our paths converge at the contraction and continue on into Π_1, where they run into a second focal pair contained in Π_1 before ending in Γ_1 or Δ_1. This possibility is depicted in the first part of the figure below, and it would give us a chain of focal pairs of length 2 in the original proof, before the duplication of subproofs.



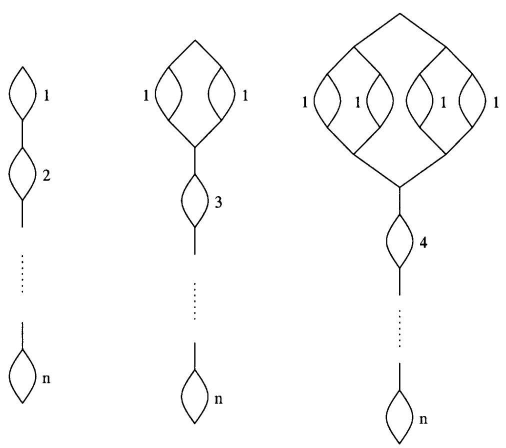


In the duplication of subproofs we eliminate the contraction at which the first convergence takes place. At best we can only postpone the convergence from the original contraction to the ones below the two copies of Π_1 in (9), as in the picture above, but this would not be good enough for maintaining the chain of focal pairs of length 2 in (9). Instead of having two focal pairs, with one following the other, we have a redistribution of the focal pairs in a kind of nesting. This is illustrated in the second part of the picture.

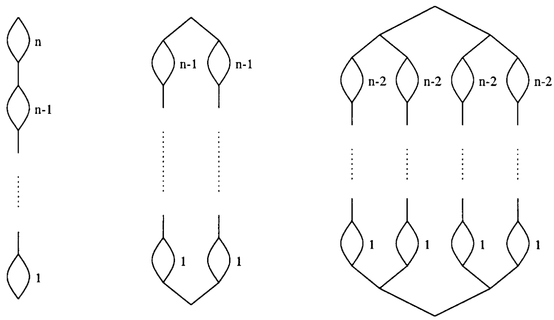
It is clear that our path from the contraction of A^1 and A^2 might continue into a chain of focal pairs of length $n>1$ in Π_1. This would give rise to a chain of length $n+1$ in (8). After the duplication of subproofs we would again lose the chain of length $n+1$ in the larger proof, and we would have two copies of the chain of length n from Π_1.

To see this phenomenon occurring in a concrete example take for instance the proof illustrated in Section 3 of the sequent $F(2) \rightarrow F(2^{2^n})$ which is built by putting together proofs Π_j of $F(2^{2^{j-1}}) \rightarrow F(2^{2^j})$ (for $1 \leqslant j \leqslant n$) using cuts. If we push a cut in this proof above the corresponding contraction by duplicating subproofs as before, then we see exactly the kind of phenomena just described. In the end the logical flow graph will be transformed into an H-graph and, as we already observed in Section 12, the final graph will correspond to the *weak negative resolution.*

There is more than one way to push the cuts up above the contractions in this case. One can start at the "bottom" of the proof (which means starting on the far right-hand side of the picture in Section 3), or at the beginning of the proof, or in the middle. The evolution of the logical graph of the proof is somewhat different if one begins at different points of the proof. For instance, if one starts at the beginning (which means the far left-hand side of the graph), then the systematic duplication of subproofs leads to an evolution of logical flow graphs roughly like the one shown here

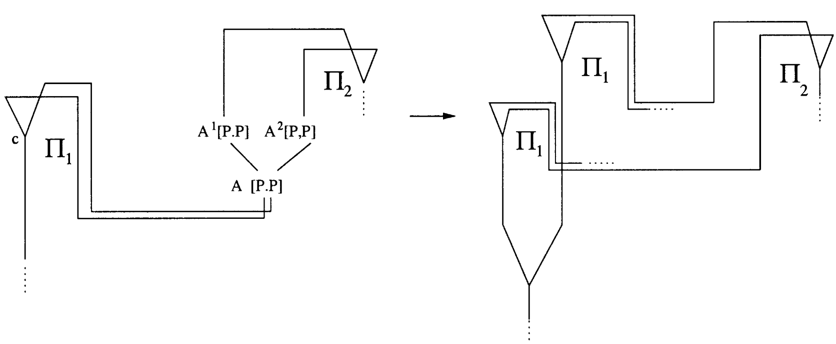


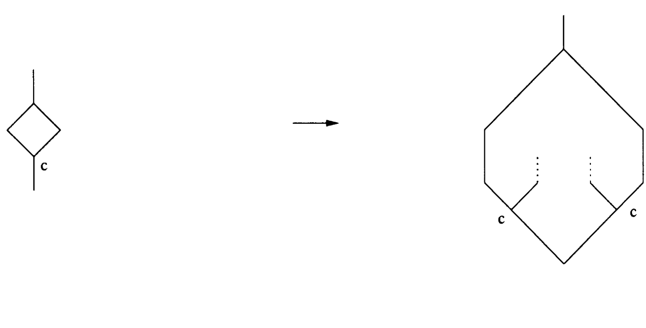
If one begins at the other end of the proof the evolution is illustrated as follows



The final result is the same, independently of whether one chooses to start from the beginning or the end of the original proof, or from anywhere in between. In the end one obtains an H-graph of exponential size. Notice however that if one starts at the beginning of the proof then the whole job is done in $n-1$ steps, i.e., with $n-1$ applications of the operation of duplicating a subproof, as in (8) and (9). The number of steps of the subproof which is duplicated at the $j+1$-th step of the procedure is twice as large as the number of steps of the subproof duplicated at the j-th stage. This means that the procedure is *linear* but that it duplicates *exponentially* large subgraphs. On the other hand, if we start from the end of the proof, we duplicate at each stage a subproof whose number of steps is bounded by the number of steps of the original proof. This leads to an *exponential* procedure which duplicates subgraphs of *bounded* size.

As we promised at the beginning of this section, we consider now the case where a pair of paths which begin at some common starting point in Π_2 of (8) do not converge at the contraction of A^1 and A^2, but wait to converge later on in Π_1.



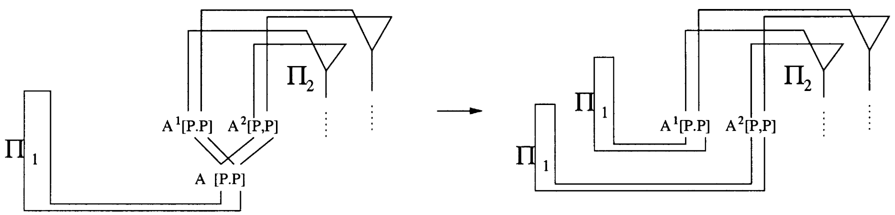


Again we assume that after the paths converge they continue on into a formula in Γ_1 or Δ_1. In this case the focal pair that we have in the original proof (8) persists in (9), but with the convergence postponed as before.

After passing through the point of convergence in Π_1 we might pass through a chain of focal pairs in Π_1, so that in the proof as a whole we have a chain of length $n+1$. As before the first focal pair in this chain would be disrupted by the elimination of the contraction at A^1, A^2, so that the chain of length $n+1$ would not persist in (9). In the end we would have two copies of the chain of length n from Π_1, just as in the first step of the evolution shown in the picture above.

13.3. Removal and redistribution of focal pairs

The phenomena of removal and redistribution of focal pairs can be seen also in case paths from Π_2 go to Π_1 through a cut and then come back out of the same cut into Π_2 again.



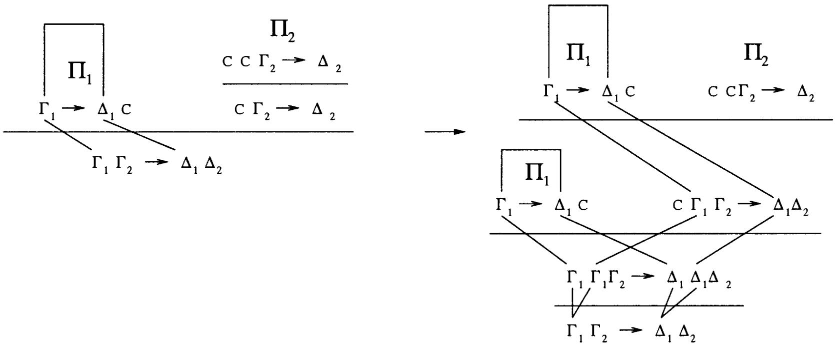
This situation can be analyzed in much the same manner as before. In fact the settings we discussed in Sections 13.1 and 13.2 can both appear here.

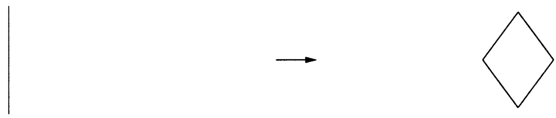
Once the paths go back into Π_2 they may or may not converge again, or encounter additional chains of focal pairs. They might eventually end in some weak formula in Π_2, or go back through the cut into Π_1 again, or go down into the endsequent of Π_2 and continue on into the rest of the proof below. These cases produce different effects on chains of focal pairs.

If the paths end in Π_2 we observe a removal of a focal pair (as illustrated in the picture above). The same happens when the paths go down into the endsequent of Π_2 and continue on into the rest of the proof below. If the paths go back to Π_1 then they might go down into the end-sequent of Π_1 and in this case a new focal pair is formed as argued in Section 13.2. Otherwise, once in Π_1 the paths might come back again to Π_2 through the same cut or end into some weak formulas in Π_1, but the analysis here would start over again.

13.4. Creation of focal pairs

Again we think of putting ourselves back in the situation of (8) and (9), in which we are duplicating a subproof Π_1 in order to split a contraction in another subproof Π_2. Instead of looking at paths that move between Π_1 and Π_2 through cut-formulas, we simply consider an oriented path in Π_1 which begins and ends in the endsequent. In this case the duplication of subproofs leads to a pair of oriented paths in the new proof (9) which have the same endpoints (coming from the contractions below the two copies of Π_1 in the new proof).





In this way a focal pair can be created in the logical flow graph, which is something that we did not see in the earlier cases. Through many repetitions of this process one can create many focal pairs, and a very simple graph in the beginning can be converted eventually into an H-graph of exponential size. Indeed, the effect of this kind of evolution is illustrated in the fourth figure of Section 5 where we were arguing the exponential expansion induced by the presence of disconnected subgraphs in G'.

This process can lead not only to the creation of focal pairs, but also to the *extension* of existing chains of focal pairs. To see this, imagine that we have our proofs Π_1 and Π_2 which are being combined with a cut to make a larger proof Π^*, and that we are duplicating Π_1 in order to simplify the cut over a contraction contained in Π_2, as in (8) and (9). Imagine also that Π^* lives inside of a larger proof Π.

If we have a path p inside Π_1 which begins and ends in the ensequent of Π_1, then we get a focal pair after the duplication of subproofs, as we saw before. However, the path p might continue below the cut in Π which connects Π_1 and Π_2, and in this continuation p might meet additional focal pairs. In this way the duplication of subproofs can lead to the increase in the length of a chain of focal pairs.

More complicated situations are represented by logical flow graphs of proofs which evolve as in the second picture of Section 7.

For this construction however it is important to have some interesting structure in the proof Π below the use of the cut rule by which Π_1 and Π_2 are connected. In particular, there should be cuts below the one under consideration.

Theorem 58. *Let Π be a proof. The bottom-up procedure of cut-elimination (i.e. the procedure which resolves bottom cuts first) does not extend any existing chain of focal pairs in any of the steps of reduction of Π to a cut-free proof.*

Proof. If one simplifies cuts starting from the bottom of the proof and moves upwards, then there would be no more cuts in Π below the one that have been considered last. In particular, if a new focal pair has been created by the procedure, its focal points should have direct paths to the end-sequent. In fact, using the notation in the first picture of this section, the new focal points would lie in Γ_1, Δ_1 and their paths going down towards the end-sequent cannot turn because of the absence of cuts. In this case there would be no possibility of extension of chains of focal pairs. □

13.5. Focal pairs and the size of proofs

With duplication we can have substantial increase in the size of a proof and therefore of its underlying logical flow graph. If we measure the complexity of the logical flow graph of a proof in terms of the length of its longest chains of focal pairs, the following fact tells us that this value is never increased by the bottom-up procedure of cut elimination.

Proposition 59. *Let Π be a proof. Each step of reduction of the bottom-up procedure of cut-elimination never augments the length of chains of focal pairs. The procedure terminates only if there are no chains of length greater than* 1.

Proof. The first part of the statement is a corollary of Theorem 58. Since the procedure of cut elimination stops only when a cut-free proof is obtained, we apply Proposition 53 and derive the second part of the statement. □

The interest on this measure of complexity stands on its geometric nature. In fact, the shortening of long chains of focal pairs and the expansion of the logical flow graph by means of the combinatorial rule $\mathscr{D}$ correspond to the reduction of the logical complexity of the cuts and the increase of the size of the proof through the bottom-up procedure of cut elimination.

The correspondence of our geometric model to the bottom-up procedure of cut elimination is a nice match. For all other procedures we might have some increase in the length of the chains of focal pairs, even though the expansion is still governed by the combinatorial operation $\mathscr{D}$. As we will illustrate in the next section, there are situations where the elimination of cuts produces exponential expansion even though the logical flow graph of the starting proof does not contain any chain of focal pairs. Chains of focal pairs can nevertheless be found at a macroscopic level in the logical graph of a proof.

14. Core proofs and logical core graphs

In this section we introduce the notions of *core proof* and *logical core graph*. Roughly speaking, a core proof is a truncated portion of a proof and the logical core graph is a "macroscopic" version of the logical flow graph of the proof. It traces the flow of occurrences of formulas in a core proof. A germinal version of these notions has been introduced in [10] where it is observed that proofs whose elimination of cuts is exponential might not contain long chains of focal pairs. One would have liked to say that exponential expansion could appear only under the presence of these chains. This is not true in general but perhaps some variant of this intuition might turn out to be true (see Conjectures 60 and 63). In [10] it is shown with some examples that the presence of chains of focal pairs might be recaptured at times in some macroscopic way. We begin with an example and we shall introduce the notions afterwards.

The following example does not contain in its polynomial size proof any chain of focal pairs even though all its cut-free proofs have exponential size. We will see that there is a core proof of this proof whose logical core graph presents chains of focal pairs.

The construction is given by Statman in [22] and shows that for all large m there is a sequent $\Gamma \to \Delta$ of size m having a proof (containing cuts) of polynomial size $\mathcal{O}(m^{1.5})$ and whose cut-free proofs have at least $2^{\sqrt{m}}$ lines. To show the claim Statman exhibits a family of sequents $\Gamma_n \to \Delta_n$ of size $\mathcal{O}(n^2)$, having proofs with cuts of size $\mathcal{O}(n^3)$ and proofs without cuts of size at least $\mathcal{O}(2^n)$.

Let us introduce some notation. Let $c_1, d_1, c_2, d_2, \ldots$ be propositional variables. Define

$$F_i = \bigwedge_{k=1}^{i} (c_i \vee d_i)$$

$$A_1 = c_1$$

$$A_{i+1} = F_i \supset c_{i+1}$$

$$B_1 = d_1$$

$$B_{i+1} = F_i \supset d_{i+1}$$

where in F_i parenthesis are associated from left to right. Let $\Gamma_n \to \Delta_n$ be the sequent

$$A_1 \vee B_1, \ldots, A_n \vee B_n \to c_n, d_n$$

which is clearly of size $\mathcal{O}(n^2)$.

A proof (with cuts) of size $\mathcal{O}(n^3)$ for the sequent $\Gamma_n \to \Delta_n$ is easily built by applying cut and contraction rules in the obvious way to sequents of the form

$$F_i, F_i, A_{i+1} \vee B_{i+1} \to c_{i+1}, d_{i+1}$$

$$F_i, A_{i+1} \vee B_{i+1} \to F_{i+1}$$

(cuts on formulas F_i will be applied for all $i=2\dots n-1$.) The first sequent is provable as follows:

$$\dfrac{\dfrac{F_i \to F_i \; c_{i+1} \to c_{i+1}}{F_i, (F_i \supset c_{i+1}) \to c_{i+1}} \quad \dfrac{F_i \to F_i \; d_{i+1} \to d_{i+1}}{F_i, (F_1 \supset d_{i+1}) \to d_{i+1}}}{F_i, F_i, (F_i \supset c_{i+1}) \vee (F_i \supset d_{i+1}) \to c_{i+1}, d_{i+1}}$$

Call this proof Π_{i+1}. The second sequent is provable as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{matrix}\Pi_{i+1}\\ F_i, F_i, A_{i+1} \vee B_{i+1} \to c_{i+1}, d_{i+1}\end{matrix}}{F_i, F_i, A_{i+1} \vee B_{i+1} \to c_{i+1} \vee d_{i+1} \; F_i \to F_i}}{F_i, F_i, F_i, A_{i+1} \vee B_{i+1} \to F_{i+1}}}{F_i, F_i, A_{i+1} \vee B_{i+1} \to F_{i+1}}}{F_i, A_{i+1} \vee B_{i+1} \to F_{i+1}}$$

In the first proof we notice that the pair of occurrences c_1, d_1 in the left hand side formula F_i has variants in the antecedent of the implication $F_i \supset c_{i+1}$ (i.e. A_{i+1}), and in the antecedent of the implication $F_i \supset d_{i+1}$ (i.e. B_{i+1}). In the second proof the pair of occurrences c_1, d_1 in the left hand side formula F_i has variants in F_{i+1}.

If we look at the paths linking the propositional variables c_1, d_1 occurring in $A_n \vee B_n$ to the propositional variables c_1, d_1 occurring in $A_1 \vee B_1$ of $\Gamma_n \to \Delta_n$ through the proof, we see that there is a pair of paths passing through *all* cut formulas in the proof (there are $n-1$ many of them) and that these paths do not belong to a chain of focal pairs.

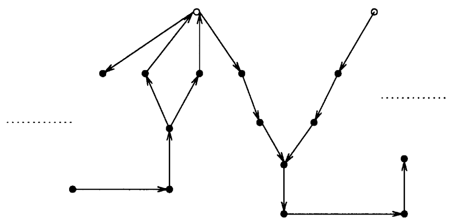
Similarly we can check that all paths passing through the c_i, d_i, for all i behave similarly. They will pass through $n-i$ cut formulas and they will not belong to any long chain of focal pairs. This concludes the example.

We are now ready to introduce some notions. A *core proof* is a tree of sequents which looks almost like a proof with the exception of the labelling of its leaves which do not need be axioms. That is, each sequent in a core proof is either an initial sequent (for which no “justification” is given) or is derived from one or two sequents through the same rules as for proofs.

In practice a core proof presents a piece of a proof where we think of the initial sequents as provable even if we do not furnish their supporting proof explicitly. In [10] this notion was called *proof structure* and was restricted to a truncated portion of a proof containing only cuts and contractions. Here we work in a general setting.

Let us now define the *logical core graph* of a core proof. As for a logical flow graph we link all atomic occurrences in a core proof as indicated by the rules of the calculus. The orientations to the edges is given as for the logical flow graph. We then consider the initial sequents. For each initial sequent and each distinct letter in it we will introduce an *additional vertex*. We then link all occurrences of a given letter in the sequent to the same additional vertex. (Note that in the propositional case there is no ambiguity in determining when two occurrences are occurrences of the same letter. For the predicate case we consider two atomic occurrences to be the same when they are variant of each other, i.e. they are the same up to change of terms.) The orientation is the natural one. Namely there is an edge from an additional vertex to the corresponding positive occurrences in the initial sequent and there is an edge from negative occurrences in the initial sequent to the corresponding additional vertices.

This concludes the definition of logical core graph. Below we illustrate a piece of a graph around an additional vertex. As one can see there might be multiple edges going in or coming out from an additional vertex, and paths linked to it can connect in the graph to other paths coming from additional vertices associated to another initial sequents, for instance.



In the example discussed above it is easy to see that if we consider the initial sequents

$$F_i, F_i, F_i, A_{i+1} \vee B_{i+1} \rightarrow F_{i+1}$$

$$F_{n-1}, F_{n-1}, A_n \vee B_n \rightarrow c_n, d_n$$

and we combine them together using contractions and cuts we obtain a core proof of the sequent $A_1 \vee B_1, \ldots, A_n \vee B_n \rightarrow c_n, d_n$. The logical core graph for the proof will have a chain of n focal pairs passing through the atomic formula c_1 (there will be an identical one for the formula d_1). The chain is shown in the picture below

It appears as a subgraph of the core graph. The circles represent the additional vertices and the defocussing branch points correspond to the contractions on the F_i's (which appear only on the left hand side of the sequent arrows). During the process of cut-elimination it is the simplification of cuts over contractions which leads to the splitting of these defocussing branch points and the exponential expansion in the size of the graph.

It is clear from the definition of core proof that given a proof, one can have several core proofs associated to it. When the initial sequents are axioms, a core proof corresponds to a proof but notice that the logical core graph does not coincide with the logical flow graph. This is simply due to our syntactical choices in the formalization of the sequent calculus. Nothing more. In fact if we were assuming to work in a calculus with *weakening* rules then axioms could be assumed to be sequents of the form $A \rightarrow A$ with A atomic, and the two graphs would coincide (since only one path would cross additional vertices).

To conclude this section we state two conjectures on the structural properties of proofs of tautologies which are "hard" to prove

Conjecture 60. *Let S be a statement of size n with only exponential size $2^{\mathcal{O}(n)}$ cut-free proofs and proofs with cuts of polynomial number of symbols. Given any polynomial size-proof of S, there is a chain of focal pairs of length n in some of the core proofs associated to it.*

A proof is called *reduced* when no binary rule is applied to a weak auxiliary formula in Π, no unary logical rule is applied to two weak auxiliary formulas and no contraction rule is applied to a weak auxiliary formula. This notion was explicitly introduced in [10] and has its origin in [3]. In Lemma 3.2 (pp. 261) of [9] (see also Proposition 6.28 pp. 142 in [10]) it is shown that

Proposition 61. *Let $\Pi : S$ be a proof of k lines. Then there is an effective way to transform Π into a reduced proof $\Pi' : S$ which has at most k lines. If Π contains no cuts then the same is true for Π'.*

The proposition above says that a reduced proof eliminates insignificant structure from a proof. One aspect of this is given in the next result which was proved in [4].

Proposition 62. *Let $\Pi : S$ be a reduced proof where distinguished formulas in axioms are atomic. Let c be the number of contractions and a be the number of axioms in Π. Then $a \geqslant C/2$.*

Conjecture 63. *Let Π be a reduced proof. Suppose that Π contains a chain of focal pairs in one of its core proofs. Then all cut-free proofs obtained from Π by cut-elimination have exponential size.*

Acknowledgements

I am greatful to Misha Gromov and Stephen Semmes for many discussions on the topics involved in this article and others.

Index of symbols (symbol, page)

Alphabetic index (subject, page)

References

[1] S. Buss, The undecidability of k-provability, Ann. Pure Appl. Logic 53 (1991) 72–102.
[2] S. Buss, Personal communication, June 1993.
[3] A. Carbone, Interpolants, cut elimination and flow graphs, Ann. Pure Appl. Logic 83 (1997) 249–299.
[4] A. Carbone, Cycling in proofs and feasibility, Trans. Amer. Math. Soc., in press.
[5] A. Carbone, Turning cycles into spirals, Ann. Pure Appl. Logic 96 (1999) 57–73.
[6] A. Carbone, Some combinatorics behind proofs, submitted for publication.
[7] A. Carbone, S. Semmes, Looking from the inside and the outside, IHES Preprint M/96/44, 1996.
[8] A. Carbone, S. Semmes, Propositional proofs via combinatorial geometry and the search for symmetry, Collegium Logicum, Ann. of the Kurt Gödel Soc., Vol. 3, Inst. Computer Science AS CR Prague Publ. 1999, pp. 85–98.
[9] A. Carbone, S. Semmes, Making proofs without modus ponens: an introduction to the combinatorics and complexity of cut elimination, Bull. Amer. Math. Soc. 34 (1997) 131–159.
[10] A. Carbone, S. Semmes, Geometry, Symmetry and Implicit Representations of Combinatorial Structures, Book in preparation. Preliminary version available at `ftp://math.rice.edu/pub/semmes`, file name `book.ps`.
[11] S. Cook, R. Reckhow, The relative efficiency of propositional proof systems, J. Symbolic Logic 44 (1979) 36–50.
[12] Gerhard Gentzen, Untersuchungen über das logische Schließen I–II, Math. Z. 39 (1934) 176–210, 405–431.
[13] J.Y. Girard, Linear logic, Theoret. Comput. Sci. 50 (1987) 1–102.
[14] Jean-Yves Girard, Proof theory and logical complexity, in: Studies in Proof Theory, Monographs, vol. 1, Bibliopolis, Napoli, Italy, 1987.
[15] A. Haken, The intractability of resolution, Theoret. Comput. Sci. 39 (1985) 297–308.
[16] V.P. Orevkov, Lower bounds for increasing complexity of derivations after cut elimination, J. Sov. Math. 20 (4) (1982).
[17] V.P. Orevkov, Complexity of proofs and their transformations in axiomatic theories, Translations of Mathematical Monographs, vol. 128, American Mathematical Society, Providence, RI, 1993.
[18] Ch. Papadimitriou, Computational Complexity, Addison-Wesley, New York, 1994.
[19] R. Parikh, Existence and feasibility in arithmetic, J. Symbolic Logic 36 (1971) 494–508.
[20] P. Pudlák, The lengths of proofs, in: S. Buss (Ed.), Handbook of Proof Theory, North-Holland, Amsterdam, 1996.
[21] Richard Statman, Structural complexity of proofs, PhD thesis, Stanford University, 1974.
[22] R. Statman, Bounds for proof-search and speed-up in predicate calculus, Ann. Math. Logic 15 (1978) 225–287.
[23] R. Statman, Lower bounds on Herbrand's theorem, Proc. Amer. Math. Soc. 75 (1979) 104–107.
[24] G. Takeuti, Proof theory, Studies in Logic, vol. 81, 2nd ed., North-Holland, Amsterdam, 1987.
[25] G.S. Tseitin, Complexity of a derivation in the propositional calculus, Zap. Nauchn. Sem. Leningrad Otd. Mat. Inst. Akad. Nauk SSSR 8 (1968) 234–259.