Towards a Sequent Calculus for Negation as Failure

James Harland, Royal Melbourne Institute of Technology

1 Introduction

We consider the problem of giving a sequent calculus interpretation of the Negation as Failure (NAF) rule[2, 10, 6]. In essence, this consists of giving proof rules which can be used to demonstrate that a given formula can be *disproved*, i.e. that there can be no proof of the formula. The notion of being able to determine the existence of such disproofs seems to be at the heart of the Negation as Failure rule, and the problem of determining disproofs has been studied for both classical propositional logic [11] and intuitionistic propositional logic [9]. As both of these logics are decidable, it is not surprising that it is possible to determine sequent calculi for disproofs in each case, although it is a lot simpler in the classical case than in the intuitionistic one — the system given in [9] relies on a careful analysis of the intuitionistic propositional system so that the contraction rule can be eliminated entirely [3], which is a subtle and intricate problem.

Whilst both of these approaches suffice for the detection of disprovable formulae, the Negation as Failure rule, as used in logic programs, requires both that a negation-as-failure connective be introduced, thus making the proof and disproof systems interdependent, and that variables be used in programs and goals. Both of these features require a significant amount of work. For the first feature, the standard provability relations for classical and intuitionistic logic are monotonic (i.e. provability is preserved under the addition of formulae), whereas the Negation as Failure rule is anti-monotonic (i.e. provability is preserved under deletion of formulae). Hence the addition of a Negation as Failure connective to either classical or intuitionistic logic requires that the new logic be *static*, i.e. neither monotonic nor anti-monotonic. Thus the usual rule of weakening will need to be modified in order to accommodate Negation as Failure, and so we investigate a way of making the system static rather than monotonic (or anti-monotonic).

The introduction of variables makes the process of discovering disproofs significantly more difficult. For example, consider the program $\forall x p(f(x))$. Here we have that $p(f(a))$ succeeds, but $p(a)$ fails. The difficulty is mainly due to the contraction rule, as it is well known that in classical logic the only place in a proof in which the use of contraction is strictly necessary is in the application of the $\forall$-L and $\exists$-R rules, i.e. when a variable may be instantiated. However, it is interesting to note that this very problem arises in Prolog systems, which, in essence, implement the $\forall$-L rule (and in the presence of Negation as Failure), and is dealt with by unification techniques. Thus it seems useful to attempt to give an account of proof rules which use unification.

We discuss some approaches to the two problems above, and consider some further properties of the sequent calculi LK for classical logic, and LJ for intuitionistic logic in order to incorporate a Negation as Failure connective. The approach in this paper differs from that of [4], as in this case we are interested in a general NAF connective rather than one restricted to atomic formulae and a particular fragment of intuitionistic logic. The current approach may be considered an update of [5].

It should be said that in many ways, proof theory seems a natural place to study the question of Negation as Failure, in that proof theory deals with the structure and properties of proofs, and as such is more "syntactic" in nature than model theory, and hence seems more oriented towards implementations. In particular, it would seem that the properties of interest to implementors should be most apparent in proof theory.

In some ways, this quest is rather quixotic, in that there can be no complete axiomatization of the complement of the provability relations of either classical or intuitionistic predicate calculus. However, we are more interested in what is *disprovable* rather than *unprovable*, and where the (no doubt fuzzy) boundaries may be. It should also be said that this analysis, such as it is, seems to provide more questions than it does answers.

We will often need to consider a dual form of sequent to the usual $\Gamma \rightarrow \Delta$, to account for "disprovable" sequents. We write such sequents as $\Gamma \not\longrightarrow \Delta$ (such sequents are called anti-sequents in [9]). We denote by $\Gamma \vdash^+ \Delta$ that the sequent $\Gamma \longrightarrow \Delta$ has a proof, and by $\Gamma \vdash^- \Delta$ that the sequent $\Gamma \not\longrightarrow \Delta$ has a disproof.

2 Weakening Rules and Monotonicity

As mentioned above, it is straightforward to derive a system of disproofs complementary to propositional LK proofs, and more intricate but still possible for propositional LJ proofs. This task is made significantly simpler by the lack of interdependence between the positive and negative systems. However, Negation as Failure is generally used in logic programs interdependently with the positive system. For example, consider the logic program below, where we write $\sim F$ for the negation-as-failure of F.

$$p \subset \sim q$$
$$q \subset \sim r$$

Clearly p is provable if q is disprovable, and q is provable if r is disprovable. The interdependence of the two systems is needed in order to perform the expected inferences on the above program.

As mentioned above, one of the key subtleties in intermixing the two systems is that the provability system is *monotonic* and the disprovability system is *anti-monotonic*. This means that when the two systems are intermixed, the resulting system will need to be neither monotonic nor anti-monotonic, and hence *static* (or what is sometimes referred to as *cautiously monotonic*).

As the monotonicity of $\vdash^+_{LK}$ is due to the presence of the weakening rules, we need to examine its role carefully. However, we do not wish to abandon the weakening rule altogether, as we do not appear to require that deduction be *relevant*, i.e. we still desire that $p, q \vdash^+ p$ (and hence that our system coincides with the standard provability relation in the absence of Negation as Failure). We also desire to retain the dual to the contraction rule, and so it seems too drastic to remove the W-L rule entirely.

Hence it is useful to subdivide the uses of the weakening rules into the cases below.

$$\frac{\Gamma, F \longrightarrow \Delta}{\Gamma, F, F \longrightarrow \Delta}\ \text{E-L} \qquad \frac{\Gamma \longrightarrow F, \Delta}{\Gamma \longrightarrow F, F, \Delta}\ \text{E-R}$$

$$\frac{\Gamma, \longrightarrow \Delta}{\Gamma, F \longrightarrow \Delta}\ \text{where } F \notin \Gamma \quad \text{M-L} \qquad \frac{\Gamma \longrightarrow \Delta}{\Gamma \longrightarrow F, \Delta}\ \text{where } F \notin \Delta \quad \text{M-R}$$

Clearly the instances E-L and E-R (named E for expansion) we would wish to retain, as this corresponds to introducing an extra copy of a formula which already exists in the sequent. However, the ones marked M-L and M-R (named M for monotonic) are problematic in this context, as we wish to permit only certain

forms of weakening, and in particular to restrict it in such a way that no information is "lost" in the middle of a proof.

A further observation that may be made is that a slightly more powerful rule than the expansion rule is of interest. In essence, the expansion rule is the precise dual to contraction; a formula that appears in a sequent may be used an arbitrary number of times in a proof. However, the following occurrence of weakening, which is not an instance of expansion, is also of interest:

$$\frac{p \subset q, q \longrightarrow q}{p, p \subset q, q \longrightarrow q} \text{ W-L}$$

This appears to be a reasonable inference, as the weakening rule only adds a formula which is provable anyway. Hence it would seem that the most appropriate and precise form of the weakening rules for our purposes are the ones below, which we refer to as the *provable weakening* rules:

$$\frac{\Gamma \longrightarrow F, \Delta \quad \Gamma \longrightarrow \Delta}{\Gamma, F \longrightarrow \Delta} \text{ PW-L} \qquad \frac{\Gamma, F \longrightarrow \Delta \quad \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow F, \Delta} \text{ PW-R}$$

Note that both of these rules are admissible in LK, as in each case the conclusion can be derived from the premiss $\Gamma \longrightarrow \Delta$ by weakening. Note also that in a system which is neither monotonic nor anti-monotonic, the two premisses in each case are independent (i.e. that the provability of $\Gamma \longrightarrow \Delta$ does not imply the provability of $\Gamma \longrightarrow F, \Delta$ and vice-versa, and respectively for the other case). The expansion rules are clearly instances of the provable weakening rules, in the case that the premiss containing F is initial.

Intuitively, we may think of these rules as stating that a formula F may be added to the antecedent (resp. succedent) if both the original sequent and the sequent with the formula added to the succedent (resp. antecedent) are provable. There is a clear similarity between these two rules and the cut rule; in particular, the cut rule can act as a form of "anti-weakening" for both of these rules, in that a formulae introduced to the proof by an occurrence of provable weakening can always be eliminated later by an occurrence of cut. For this reason we think of the provable weakening rules not as structural rules per se, but as a dual of sorts to the cut rule.

It should also be noted that the provable weakening rules have the subformula property, as every formula in the conclusion of either rule appears in the premiss of the rule. Hence there is no conceptual problem in using these rules "backwards", i.e. for proof search.

For these reasons it seems that the appropriate way to intermix the positive and negative systems is to use a form of the axioms corresponding to an upwards permutation of the M-rules, to determine an appropriate introduction rule for $\sim$ and the quantifiers, and to consider the properties of the *indirect* rules of provable weakening and cut.

3 Rules for the Negation as Failure Connective

It is simple to see what the logical rules (i.e. those that introduce a connective) for the standard connectives should be, both for propositional LK^+ and LK^-, with the latter coming from the well-known reversibility properties. The remaining question is how to determine the rules for the Negation as Failure connective $\sim$. This will involve the intermixing of LK^+ and LK^- into a new system, which we will denote as $\mathrm{LK}^{\pm}$.

It is straightforward to come up with the appropriate rules when there is only one formula in the succedent. Clearly if we have that $\Gamma \vdash^{-}_{LK} F$, we would expect to be able to derive $\Gamma \vdash^{+}_{LK} \sim F$ by Negation as Failure, and dually, we expect that $\Gamma \vdash^{+}_{LK} F$ enables us to derive $\Gamma \vdash^{-}_{LK} \sim F$.

It is more intricate when there are multiple formulae in the succedent. Intuitively, if we have derived the sequent $\Gamma \longrightarrow \Delta$, then we know that at least one of the formulae in Δ is derivable from those in Γ, but we may not be able to specify which formula it is. Hence, when it comes to the right introduction rule for $\sim$, all that we can infer from the provability of $\Gamma \longrightarrow F_1, \ldots, F_n$ is the provability of $\Gamma \not\longrightarrow \sim F_1, \ldots, \sim F_n$, i.e. that the negation-as-failure of the succedent is disprovable from the antecedent.

For the other direction, i.e. the derivation of provability information from $\Gamma \not\longrightarrow \Delta$, the introduction rule is a little more subtle. From the derivation of $\Gamma \not\longrightarrow \Delta$, we know that each of the formulae in Δ has a disproof from Γ, and hence we can infer $\sim F$ for each $F \in \Delta$. Hence we arrive at the two rules below.

$$\frac{\Gamma \longrightarrow F_1, \ldots, F_n}{\Gamma \not\longrightarrow \sim F_1, \ldots \sim F_n} \sim\text{-R}- \qquad\qquad \frac{\Gamma \not\longrightarrow F_1, \ldots, F_n}{\Gamma \longrightarrow \sim F_i} \sim\text{-R}+$$

Note that from $\Gamma \not\longrightarrow F_1, \ldots, F_n$ we may use n occurrences of the $\sim$-R+ rule to derive $\Gamma \longrightarrow \sim F_i$ for each $1 \leq i \leq n$, and hence we may derive $\Gamma \longrightarrow \sim F_1 \wedge \ldots \wedge \sim F_n$.

Note also that the $\sim$-R+ rule does not have the subformula property. This is important for the proof search interpretation of the rules, as without this property, it is necessary to "guess" the missing formulae. However, in this case, a reasonable alternative is to always choose $n = 1$, as if one is trying to derive $\Gamma \longrightarrow \sim F$, it seems only reasonable to attempt to derive $\Gamma \not\longrightarrow F$, rather than a more complicated succedent.[1]

This suggests that an intuitionistic system is also of interest, and particularly as classical and intuitionistic logics coincide on Horn clauses [8]. It is not hard to show that the incorporation of the $\sim$ rules above into the propositional LJ preserves certain basic properties. Presumably the incorporation of these rules into the CRIP system [9] would yield a more interesting system, but that is still under consideration.

The problem of left introduction rules for $\sim$ is still problematic.

4 Indirect Rules

At this point, it seems natural to ask about the role of the indirect rules (i.e. cut and provable weakening) in the negative system. We saw above how provable weakening and cut are duals in the positive system; are there analogous rules in the negative and/or mixed systems?

When considering the role of the cut rule in LK, we may reason as follows. For there to be no proof of a sequent $\Gamma \longrightarrow \Delta$, then it cannot be the case that both $\Gamma \longrightarrow F, \Delta$ and $\Gamma, F \longrightarrow \Delta$ are provable, as then by the cut rule we would have that $\Gamma \longrightarrow \Delta$ is provable. Hence we must have that for all formulae F, one of $\Gamma \longrightarrow F, \Delta$ and $\Gamma, F \longrightarrow \Delta$ is not provable. Hence from the cut rule we may derive the rules below:

[1] This question is related to the matter of the axioms for disproofs; given an axiom $\Gamma \not\longrightarrow \Delta$, then it is clear that $\Gamma \vdash^{-}_{LK} A$ for each $A \in \Delta$. However, space restrictions prevent us from discussing this point further.

$$\frac{\Gamma, F \not\longrightarrow \Delta \quad \Gamma \longrightarrow F, \Delta}{\Gamma \not\longrightarrow \Delta}\ \text{pull} \qquad \frac{\Gamma \not\longrightarrow F, \Delta \quad \Gamma, F \longrightarrow \Delta}{\Gamma \not\longrightarrow \Delta}\ \text{pull}$$

A similar analysis of provable weakening gives us the rules below.

$$\frac{\Gamma \longrightarrow F, \Delta \quad \Gamma \not\longrightarrow \Delta}{\Gamma, F \not\longrightarrow \Delta}\ \text{PW-L–} \qquad \frac{\Gamma \not\longrightarrow F, \Delta \quad \Gamma \longrightarrow \Delta}{\Gamma, F \not\longrightarrow \Delta}\ \text{PW-L–}$$

$$\frac{\Gamma, F \longrightarrow \Delta \quad \Gamma \not\longrightarrow \Delta}{\Gamma \not\longrightarrow F, \Delta}\ \text{PW-R–} \qquad \frac{\Gamma, F \not\longrightarrow \Delta \quad \Gamma \longrightarrow \Delta}{\Gamma \not\longrightarrow F, \Delta}\ \text{PW-R–}$$

Note that the pull[2] rules are the negative version of the cut rule, as these do not have the subformula property, whereas the provable weakening ones do have the subformula property. Note again that in a system which is neither monotonic nor anti-monotonic, the premisses of the provable weakening rules are independent (in fact, the rules themselves only really make sense in a non-monotonic system).

There are corresponding versions of these rules for the intuitionistic system, derived from these rules by the usual restriction that there is at most one formula in the succedent.

The question then arises of whether these rules are admissible, which is still open. As the provable weakening rules have the subformula property, it can be argued that this is less important for the provable weakening rules than for the pull rules.

5 Variables

One of the well-known features of LK and LJ is that the contraction rule needs to be used in conjunction with ∀-L in order to maintain completeness. For example, consider the sequent $p(a), \forall x p(x) \supset p(f(x)) \rightarrow p(f(f(a)))$, which is clearly provable. However, in order to prove this sequent we need to instantiate the bound variable x to two different values at different points in the proof. This effect is often achieved by altering the ∀-L rule to be as follows:

$$\frac{\Gamma, F[t/x], \forall x F \rightarrow \Delta}{\Gamma, \forall x F \rightarrow \Delta}$$

In this way a copy of the quantified formula is kept, as well as the instantiated form.

A refinement of this idea was given by Cerrito [1]. In such cases, rather than having to "guess" the term t, the variable x is replaced with a *witness* variable W_i, and modifications made to the axioms (and to the ∀-R and ∃-L rules) to allow the choice of the term t to be made lazily (i.e. at the axioms, rather than when the quantifier is reduced). This may be thought of as using unification techniques to avoid the need for an explicit contraction rule, and it is shown in [1] how such a system is sound and complete with respect to systems containing the normal rules[3].

However, as far as inferring negative information is concerned, this system is certainly an improvement, but is still a little imprecise. For example, consider the sequent $q(b), p(a) \rightarrow \exists x p(x)$. Clearly this sequent is provable, and using the technique of Cerrito, the sequent would be reduced to $q(b), p(a) \rightarrow p(W)$, when it is

[2] The terminology comes from the game of cricket. The cut and pull are similar shots played on different sides of the wicket.

[3] The proof in [1] is actually for linear logic, but as remarked in [1], the same technique should be applicable to any "reasonable" system, including classical and intuitionistic logics.

discovered that $W = a$ provides the appropriate witness. Also, it is clear that there is no proof of $q(b), p(a) \rightarrow \forall x p(x)$, as the new constant which is substituted for x does not unify with a. Now consider the negative sequents $q(b), p(a) \not\rightarrow \exists x p(x)$, which should not be derivable, and $q(b), p(a) \not\rightarrow \forall x p(x)$, which should be derivable. In the latter case, it seems reasonable to replace x with a witness variable, and have the axiom case find a substitution for it so that there is *no* common formula between antecedent and succedent, i.e. that the negative sequent gets reduced to $q(b), p(a) \not\rightarrow p(W)$, and we find that the substitution $W \leftarrow b$ results in an unprovable sequent. This suggests that the former sequent should be reduced to $q(b), p(a) \not\rightarrow p(c)$, where c is a new constant, and we find that this sequent is indeed unprovable, and hence $q(b), p(a) \not\rightarrow \exists x p(x)$ is derivable, which is absurd. Hence the precise way in which the quantifiers are to be used needs careful analysis.

The main problem is that the use of a new constant is not precise enough, and hence we need a more precise method for quantifiers of existential strength [1] (i.e. $\forall$ in an antecedent and $\exists$ in a succedent). Similar remarks apply to the system of Lincoln and Shankar [7].

A promising approach is to us a notion of "uninstantiable" variables rather than new constants. For example, we reduce the sequent $\forall x p(f(x)) \rightarrow \forall x p(x)$ to $p(f(x_e)) \rightarrow p(x_u)$, where x_e is an instantiable (i.e. existential) variable, and x_u is an uninstantiable variable. The trick is then to determine the appropriate conditions on variables and substitutions to capture the exact notions of provability and disprovability in such situations. The analsis of [7] would presumably be useful here.

6 Conclusions and Further Work

The full disprovability system for both the classical and intuitionistic case is still under investigation, with the main problem being the precise form of the quantifier rules. Preliminary versions of these rules, such as in [5], have all proved to be unsatisfactory, although it is possible to get some basic results, such as showing that for the Horn clause fragment no goal is both provable and disprovable, that if the body of a clause is provable, then so is the head of the clause and that replacing a literal with its "provability value" (i.e. deleting clauses containing $\sim A$ where A is provable, deleting from the body of a clause an atom A where A is provable, etc.) does not change the consequences of the program. Unfortunately, space restrictions prevent the formal statement of these results here. Ultimately, the precision of the negative forms of the quantifier rules depends on the quality of the analysis of their role in the original system.

The other main area of investigation is the indirect rules, and in particular whether they are admissible or not. In part, this relates to the question of exactly what the notion of disprovability should be. For example, given the program $p \supset p$ and the goal p, it is well known that SLD-resolution will not terminate, but that other evaluation methods, such as bottom-up execution and top-down execution with memoing will terminate. Hence the latter methods correspond to a stronger notion of disprovability. It is interesting to note that it is possible to derive $p \supset p \not\longrightarrow p$ from the intuitionistic version of the PW-L– rule. Hence the system of [9], in which this same "anti-sequent" is derivable, is presumably of interest for the eliminability of the indirect rules. Also, this raises the possibility that SLD-resolution corresponds to a system without indirect rules, whereas the other methods correspond to one with indirect rules.

A related point is that in LJ contraction is generally necessary before applying the the $\supset$-L rule, just as for $\forall$-L. Given that the original motivation for linear logic was the decomposability of intuitionistic implication, there are some intriguing possibilities for this kind of analysis in the linear sequent calculus.

Being able to show that a formula is disprovable, and hence unprovable, is presumably of interest for showing the consistency of a given theory. Hence the existence of a formula F such that $\sim F$ is provable will show that the given theory is consistent. This kind of property is neatly handled in LK by the multiple conclusions; in LJ, it is less obvious how such inferences may be made.

7 Acknowledgments

The author would like to thank Serenella Cerrito, Jean-Yves Girard, Nevin Heintze, John Lloyd, Dale Miller and David Pym for interesting discussions related to this material. Suggestions from some anonymous referees have also been very helpful.

The support of the Australian Research Council through the Machine Intelligence Project and of the Centre for Intelligent Decision Systems is gratefully acknowledged.

References

1. S. Cerrito, Herbrand Methods in Linear Logic, Proceedings of the Joint International Conference and Symposium on Logic Programming, Washington DC, November, 1992.
2. K. Clark, Negation as Failure, in *Logic and Databases*, H. Gallaire and J. Minker (eds.), Plenum Press, 1978.
3. R. Dyckhoff, Contraction-free Sequent Calculi for Intuitionistic Logic, *Journal of Symbolic Logic* 57:3:795-807, 1992.
4. J. Harland, *On Hereditary Harrop Formulae as a Basis for Logic Programming*, Ph.D. thesis, Department of Computer Science, University of Edinburgh, July, 1991. Published as Technical Report CST-81-91.
5. J. Harland, *Towards a Static Proof System for Negation as Failure*, Technical Report 92/10, Department of Computer Science, University of Melbourne.
6. J. Jaffar, J-L. Lassez and J. Lloyd, Completeness of the Negation as Failure Rule, *Proceedings of the International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, 1983.
7. P. Lincoln and N. Shankar, Proof Search in First-order Linear Logic and Other Cut-free Sequent Calculi, to appear in the Proceedings of the Ninth IEEE Symposium on Logic in Computer Science, Paris, July, 1994.
8. D.A. Miller, G. Nadathur, F. Pfenning and A. Scedrov, Uniform Proofs as a Foundation for Logic Programming, *Annals of Pure and Applied Logic* 51:125-157, 1991.
9. L. Pinto and R. Dyckhoff, Loop-free Construction of Counter-models for Intuitionistic Propositional Logic, Proceedings of the Gauss Symposium, Munich, August, 1993.
10. R. Reiter, On Closed World Databases, in *Logic and Databases*, H. Gallair and J. Minker (eds.), Plenum Press, 1978.
11. M. Tiomkin, Proving Unprovability, *Proceedings of the IEEE Symposium on Logic in Computer Science* 22-26, Edinburgh, July, 1988.

This article was processed using the LaTeX macro package with LLNCS style