Progress has been made in the application of integer programming to logical inference. A new algorithm for verifying logic circuits has been developed. A new regression-based method for generating rules for logical systems has been developed.
MATHEMATICAL PROGRAMMING
AND LOGICAL INFERENCE

Technical Report

December 1990

Grant No. AFOSR-87-0292

Submitted to the
Air Force Office of Scientific Research

by

Egon Balas
Gérard Cornuejols
John Hooker

Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213
ABSTRACT

This is the final technical report for the grant, "Mathematical Programming and Logical Inference," AFOSR-87-0292. The object of this research is to develop new and effective methods for logical inference that are based on mathematical programming.

We investigated fast packing and covering algorithms as well as polyhedral properties of these problems. We identified classes of covering and inference problems that can be solved by linear programming. We also obtained several results in both deductive and inductive logic. In the area of deductive logic, we developed branch-and-cut algorithms for inference in propositional logic, generalized the notion of a Horn problem (widely used in expert systems), designed a new algorithm for verifying logic circuits, found new connections between propositional logical and cutting plane theory, developed an inference method for a generalized belief net ("Bayesian logic"), and proposed new computational methods for Dempster-Shafer theory. In inductive logic, we proposed a new, regression-based method for generating rules for an expert system.
# TABLE OF CONTENTS

1. INTRODUCTION .............................................. 4

2. TECHNICAL RESULTS ......................................... 5
   2.1 Packing and Covering .................................. 5
      2.1.1 Fast Packing and Covering Algorithms
      2.1.2 Polyhedral Results for Covering Problems
      2.1.3 Unions of Polyhedra
   2.2 Problems Soluble with Linear Programming ............ 6
   2.3 Inference in Propositional Logic ....................... 7
      2.3.1 Branch and Cut Methods
      2.3.2 Resolution and Cutting Planes
      2.3.3 Generalized Horn Problems
      2.3.4 Equivalence of Logic Circuits
      2.3.5 Surveying the Field
   2.4 New Inference Methods for Belief Systems .......... 9
      2.4.1 Belief Nets
      2.4.2 Dempster-Shafer Theory
   2.5 A New Approach to Obtaining Rules for Expert Systems .. 10
   2.6 A New Book ........................................... 10

3. REFERENCES .................................................. 12

4. STATUS OF RESEARCH PAPERS ................................. 15
1. INTRODUCTION

This is the final technical report for grant AFOSR-87-0292, "Mathematical Programming and Logical Inference."

The purpose of this research is to search for mathematical structure in the semantics of logics that are useful in artificial intelligence and decision support systems, and to exploit this structure to solve inference problems rapidly. In particular we look for structure that permits us to use the problem-solving machinery of mathematical programming. We believe that these quantitative methods can solve inference problems that have proved difficult or impossible for symbolic methods popular in artificial intelligence, and this belief has been partially confirmed for propositional and probabilistic logic.

We solve inference problems in propositional logic by formulating them as a integer programs, whose structure we exploit to solve with branch-and-cut and other methods. We propose solving problems in belief nets by combining such mathematical programming techniques as column generation and Benders decomposition.
2. TECHNICAL RESULTS

2.1 Packing and Covering

Packing and covering problems are closely related to inference problems in propositional logic.

2.1.1 Fast Packing and Covering Algorithms

The set packing problem is equivalent to the vertex packing problem on the intersection graph $G_A$ of the coefficient matrix of the packing problem at hand, and it is also equivalent to the maximum clique problem on the complement $G_A$ of $G_A$.

In a sequence of papers by Balas and Yu, Balas, Chvatal and Nesetril, and Balas in the mid-eighties, a new type of branch and bound approach was introduced for finding a maximum clique in an arbitrary graph, in which the subproblems generated are polynomially solvable. This is achieved by always choosing subgraphs that belong to some polynomially solvable class. When the subgraphs are triangulated, it is convenient to use the "dual" problem of finding a minimum weighted vertex coloring as an upper bounding device. In [9] Balas and Xue give an $O(n^2)$ algorithm for finding such a vertex coloring, and use it in the framework of a branch and bound algorithm of the above mentioned type to solve the maximum weight clique problem in an arbitrary graph. While earlier methods were typically applied to problems with 50-100 vertices, the new algorithm solves problems on random graphs with up to 1000 vertices.

In another computationally oriented paper [5], Balas and Carrera use a subgradient-based procedure which combines dual ascent with primal heuristics and incorporates cut generating techniques, to solve large sparse real world set covering problems with up to 200-300 constraints and 4000-8000 variables. The algorithm is a vastly improved version of the Balas-Ho approach developed in the late seventies.

2.1.2 Polyhedral Results for Covering Problems

The "deepest" and most effective cutting planes for an integer program associated with an inference problem are the facets of the convex hull of the integer solutions. G. Cornuéjols and A. Sassano wrote a paper [17] describing the 0,1 facets (those with variable coefficients in $\{0,1\}$ and arbitrary right hand side coefficient) for the set covering problem, which is a special case of the inference problem.

Balas and Ng [6] characterized the class of valid inequalities for the set covering polytope with coefficients equal to 0, 1 or 2, and gave necessary and sufficient conditions for such an equality to be minimal and to be facet defining. They showed that all inequalities in the above class are contained in the elementary closure of the constraint set, and that 2 is the largest value of $k$ such that all valid inequalities for
the set covering polytope with coefficients no greater than \( k \) are contained in the elementary closure.

In a companion paper [7], Balas and Ng connected this characterization to the theory of facet lifting. In particular, they introduced a family of lower dimensional polytopes and associated inequalities having only three nonzero coefficients, whose lifting yields all the valid inequalities in the above class, with the lifting coefficients given by closed form expressions.

### 2.1.3 Unions of Polyhedra

A central problem in polyhedral combinatorics is to characterize the convex hull of a union of polyhedra. One such characterization, given by Balas in the mid-seventies, is by a system of linear inequalities in \( q \cdot n \) variables, where \( q \) is the number of polyhedra in the union and \( n \) the dimension of the space containing the polyhedra. When all polyhedra are described by systems that differ only in their right hand side, it is sometimes possible to describe the convex hull of the union by a system whose left hand side is the common left hand side of the individual systems, and whose right hand side is a convex combination of the individual right hand sides. This reduces the number of variables needed for the characterization from \( q \cdot n \) to \( q + n \). Jeroslow, and later Blair, specified certain conditions under which such a simplified representation is possible. In [4], Balas gave a new sufficient condition for this property to hold, which is often easier to recognize. In particular, he showed that the condition is satisfied for polyhedra whose defining systems involve the arc-node incidence matrices of directed graphs, with certain right hand sides. As a special case, he also derived the compact linear characterization of the two terminal Steiner tree polytope due to Ball, Liu and Pulleyblank.

### 2.2 Problems Soluble with Linear Programming

Some inference problems in propositional logic can be solved relatively quickly by linear programming. These include problems that, when formulated as an integer program and the integrality constraints dropped, necessarily have an integer solution. Problems whose coefficients form an "ideal" matrix have this property, and in [16] G. Cornuejols and B. Novick have undertaken to characterize ideal matrices. Their approach is to describe the matrices that are minimally nonideal (i.e., they become ideal if any variable is fixed to 0 or 1). The results have striking similarities with the theory developed over the past twenty years for another important class of matrices, the so-called perfect matrices. There are also important differences. One such difference is a rich variety of small minimally nonideal matrices (whereas there are only three known classes of minimally imperfect matrices).
2.3 Inference in Propositional Logic

An inference (or satisfiability) problem in propositional logic can be formulated as a generalized covering problem. This is a 0-1 integer programming problem whose constraints have the form $ax \geq b$, where the coefficients belong to \{0,1,-1\}. When $b$ is equal to one minus the number of -1's in $a$, the problem represents an inference problem. When, in addition, the coefficients in $a$ belong to \{0,1\}, the problem is a set covering problem. When the coefficients belong to \{0,1\} and $b$ is equal to one less than the number of 1's in $a$, the problem is a set packing problem.

2.3.1 Branch and Cut Methods

One approach to solving the integer program associated with an inference problem is by a branch-and-cut method; that is, a branch-and-bound method that generates cutting planes at some or all nodes of the search tree. In some early work, J. Hooker solved this problem by exploiting the fact that resolution, a well-known theorem proving technique, can be used to generate separating cuts. This approach solved inference problems 1000 or more times faster than ordinary resolution on a large class of randomly-generated problems [21]. He also showed that two particular types of resolution (input and unit resolution), which are used for inference in Horn knowledge bases, in effect generate all propositions that are "rank one" cutting planes for the integer program [24]. In a third paper [26], he and C. Fedjki showed that these cutting planes, as part of a branch-and-cut procedure, lead to an even more effective inference algorithm. It solved hard randomly-generated inference problems more rapidly than what appeared to be the stiffest competition, a very promising branching algorithm developed by R. Jeroslow and J. Wang [28]. The Jeroslow and Wang method, however, was superior on easy problems. No attempt was made to compare the branch-and-cut method with the traditional resolution-based methods, because the latter would run far too long on problems of the size tested.

2.3.2 Resolution and Cutting Planes

One of the fundamental problems of integer programming is to generate all valid cuts for a given set of constraints in 0-1 variables. One approach to solving the problem is to generate a complete set of strongest possible or "prime" cuts, which are cuts that are strictly dominated by no other. (One cut dominates another when all 0-1 points satisfying one satisfy the other.) The problem of generating all prime cuts is a generalization of the problem of generating all prime implications of a set of logical clauses, which can be solved by resolution (as shown by W. V. Quine in the 1950's). J. Hooker showed that resolution can be generalized to generate all prime cuts for an extended type of clause in which at least a specified number of propositions are asserted to be true [20]. In a recent paper [25], he extended this result to a method for generating prime cuts. In particular, he showed that two basic cutting plane operations generate all prime cuts (up to equivalence). Thus one
can solve a fundamental problem of cutting plane theory by taking a logical point of view.

### 2.3.3 Generalized Horn Problems

Horn clauses (disjunctions containing at most one unnegated literal) are very important in artificial intelligence because, for them, the inference problem can be solved rapidly. It is natural to try to extend the notion of a Horn clause to cover a wider class of propositions without forfeiting ease of solution.

S. Yamasaki and S. Doshita [32] found such a class that permits multiple positive literals in a clause provided that, when the clauses are combined, the positives are "nested." V. Arvind and S. Biswas [2] found an $O(n^2)$ algorithm for solving these problems. G. Gallo and Scutellà [18] generalized this work by finding a hierarchy of problem classes (each recognizable in polynomial time), the first of which is the ordinary Horn class, and the second of which is the Yamasaki and Doshita class.

V. Chandru and J. Hooker [14] found a generalization of Horn problems for which the satisfiability (or inference) problem can be solved using the same technique used for ordinary Horn problems (unit resolution), and just as rapidly. Beginning with a result of Chandrasekaran [12], they showed that to every rooted, directed tree there corresponds a family of generalized Horn problems. In particular, ordinary Horn sets correspond to wheels in which the hub is the root. Horn problems that correspond to a given rooted tree are those each of whose clauses can be regarded as specifying flows on the tree that take the pattern of a star subtree centered at the root, plus a single chain. They extended an idea of Aspvall [3] to formulate an $O(n^2)$ recognition algorithm for problems of this form (modulo complementation of variables) when the underlying tree, and the assignment of variables to its arcs, are given. As yet there is no known polynomial-time recognition algorithm when the tree is unspecified. But Chandru and Hooker show how a knowledge base having extended Horn structure can be built in practice by choosing an underlying tree that suits the application.

### 2.3.4 Equivalence of Logic Circuits

An important problem in computer design is to check the equivalence of a newly designed circuit with one known to represent the desired boolean function. Computer firms have been known to spend months of computer time checking a new circuit design by simulating its behavior for all (or many) possible inputs. J. Hooker and his student Hong Yan solved this problem by applying Benders' decomposition to an integer programming model of the problem. Computational results were initially discouraging, because of the large number of Benders cuts that had to be generated. But more recently they discovered that a logical interpretation of the Benders cuts leads to a totally new symbolic algorithm, which looks much more promising. Computational tests are underway.
2.4 New Inference Methods for Belief Systems

2.4.1 Belief Nets.

Belief nets are used to represent uncertainty and causal relationships in expert systems. One popular type of belief net is a Bayesian network [30], which becomes an influence diagram [27][31] when decision nodes are added. In order to use a Bayesian network, however, one must specify an often impractically large number of prior and conditional probabilities.

An alternative approach to representing uncertainty and causal relations is probabilistic logic, which was originally proposed two centuries ago by G. Boole [10][19] and recently reinvented by N. Nilsson [29]. It is much easier to use than Bayesian networks, since one can specify only as many probabilities (or probability ranges) as he knows. The inference problem can be solved as a linear program if one uses column generation methods, which have been proposed independently by three groups: D. Kavvadias and C. H. Papadimitriou; P. Hansen, B. Jaumard and students; and J. Hooker [23].

But probabilistic logic also has a weakness: much of what people know about probabilities is that a given proposition depends on only a few others in the knowledge base and is essentially independent of the rest. These independence assumptions are captured in a Bayesian network but not in probabilistic logic.

K. A. Andersen and J. Hooker [1] obtained the advantages of both probabilistic logic and Bayesian networks by merging them to yield a new type of belief system they call Bayesian logic. Inference poses a substantial computational problem in Bayesian logic, as in Bayesian networks. But Andersen and Hooker showed that applying Benders' decomposition technique to the nonlinear program corresponding to a Bayesian logic problem allows one to use the same column generation techniques that are used in probabilistic logic. They also showed that for a large class of networks (including many that are not "singly connected"), the number of nonlinear constraints needed to encode the independence assumptions grows only linearly with the size of the problem. In particular, the following was proved. Divide the ancestors (predecessors) of a node in a Bayesian network into generations, so that the node itself comprises generation 0, and the parents (immediate predecessors) of all the nodes in generation k comprise generation k + 1. Further divide each generation into sets, no member of which has a common ancestor with a member of another. The resulting sets are ancestral sets, and an ancestral set joined by the parents of all its members is an extended ancestral set. Then if the maximum size of an ancestral set is bounded by a constant, the number of nonlinear constraints required grows linearly with the number of nodes in the network.
2.4.2 Dempster-Shafer Theory

Dempster-Shafer Theory is a well-known mathematical approach to combining evidence. It differs from probabilistic logic and Bayesian systems in that, in the latter systems, one must accumulate all the evidence for or against a given proposition and then assign a probability number that reflects the degree of evidence. In Dempster-Shafer theory, one can assign several probability numbers that reflect different sources of evidence and combine them mathematically. The task computing the combination grows exponentially with the number of sources, however. V. Chandru and J. Hooker propose in a book (now in preparation [15]) a new method, based on a set covering model, for computing the combination.

2.5 A New Approach to Obtaining Rules for Expert Systems

A serious bottleneck in the construction of expert systems is reducing an expert's knowledge to rules. One approach is to analyze a record of the expert's behavior over a long period and extract rules that describe it. The usual approach is to use a clustering algorithm or some similar approach to find rules that are reasonably simple and yet reasonably approximate the expert's behavior. But there is no way to check statistically whether a pattern captured by the rules is genuine or a random effect. E. Boros, P. Hammer and J. Hooker have proposed a new approach that parallels regression theory in statistics [11]. Just as one fits a mathematical formula to numerical data, they fit a logical formula to discrete data. (The approach differs from logit and categorical data analysis.) This permits statistical tests of significance similar to those used in classical regression analysis. Boros, Hammer and Hooker use Bayesian estimation, since maximum likelihood estimation has some undesirable properties. They solve the problem in the form of a pseudo-boolean optimization problem. They also develop fast algorithms for solving special cases of the problem, such as cases in which no negations occur in the problem. Their paper is in preparation.

2.6 A New Book

J. Hooker wrote for Decision Support Systems the first survey of mathematical programming methods in logic [22]. It traces the development and historical roots of the field, identifies the important problems, and proposes directions for future research. J. Hooker and V. Chandru of Purdue University wrote a second survey paper that, unlike the first, discusses types of logic other than propositional logic. It appeared as a chapter in a book on AI in manufacturing [8]. Our conversations with other investigators indicate that these two papers have sparked interest in the field in Japan and Europe as well as the United States. Indeed, papers on the satisfiability problem have mushroomed in the last couple of years.

Chandru and Hooker are extending these two essays into a systematic treatment of the field, Optimization Methods for Logical Inference [15].
It will have chapters on propositional logic (both special cases and the general problem), predicate logic, probabilistic logic and belief systems, and constraint logic programming.
4. REFERENCES


[6] Balas, E., and S. M. Ng, On the set covering polytope: I. All the facets with coefficients in \{0,1,2\}, to appear in Mathematical Programming.


12


## 5. Status of Research Papers

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6] E. Balas, S. M. Ng</td>
<td>One the Set Covering Polytope: I. All the Facets with coefficients in {0,1,2}</td>
<td>published in <em>Mathematical Programming</em> 43 (1989) 57-69</td>
</tr>
<tr>
<td>[7] E. Balas, S. M. Ng</td>
<td>On the Set Covering Polytope: II. Lifting the Facets with Coefficients in {0,1,2},</td>
<td>published in <em>Mathematical Programming</em> 45 (1989) 1-20</td>
</tr>
<tr>
<td>[14] V. Chandru, J. Hooker</td>
<td>Extended Horn Sets in Propositional Logic</td>
<td>to appear in <em>Journal of the ACM</em></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
</tbody>
</table>