INTEGER PROGRAMMING APPLIED TO THE MAP LABEL PLACEMENT PROBLEM

STEVEN ZORASTER
Zycor, Inc., Austin, Texas, USA

ABSTRACT This paper describes a point label placement program that uses a mathematical optimization algorithm to determine the best position for each label. The program detects all label overplots, moves labels to new positions to resolve overplot problems, and deletes labels when absolutely necessary. All tasks are performed without human intervention. The program is designed for use in production mapping application in the oil industry where thousands of labels must be placed, and hundreds of label conflicts resolved on a single map in everyday operations. This function is performed accurately and efficiently by this program, independent of the number of labels involved. Based on success in this application, it is reasonable to consider the use of optimization techniques to help solve other problems in automated cartography, including label placement for linear features and the selection of features to be displayed on a map.

INTRODUCTION

Name label positioning is a labor-intensive process in manual cartography. Even cartographic production systems that create maps using automated techniques can demand considerable manual intervention to resolve conflicts between labels. Therefore, efficient algorithms to place labels and resolve label overplotting are highly desirable.

Several authors have described label placement algorithms implemented in research environments (Hirsch 1982, Ahn and Freeman 1983, Ahn 1984, Basoglu 1984, Pfefferkorn, et al. 1985). Production systems are in use at several major petroleum companies, but the placement algorithms utilized are highly specialized and largely undocumented.

Existing experimental algorithms are based on heuristic procedures which support the general guidelines for name label placement described in the cartographic literature (Imhof 1975, Robinson et al. 1978). These algorithms are characterized by sophisticated data structures and complex rule following logic. The greatest drawback to most proposed solutions is an inability to take a global view of the problem during the placement of labels and the resolution of overplots. Existing techniques implicitly assume that decisions which are based on local symbol and label patterns will affect only small areas of the map. This is not always true in practice.

When the labeling problem is considered as part of an overall map compilation process, it is natural to think in the same terms used in mathematical optimization theory. The goal of label placement is to maximize (in various proportions and according to some norm) the readability, information content, and artistic quality of a map. This must be done subject to certain constraints. For example, two labels should not overplot each other.

This optimization model has been exploited by Cromley (1984), who has developed an overplot resolution algorithm for point symbols based on a linear programming relaxation procedure and interactive detection of overplots. Starting with Cromley's ideas, a fully automated algorithm has been developed which uses integer programming to resolve label overplotting on the basemaps used in the petroleum exploration industry.

Petroleum industry basemaps are dominated by two kinds of labels, those

STEVEN ZORASTER is Manager of the Research Division of Zycor Inc., Austin, Texas, USA 78741-3818.
MS *submitted 2 January 1986*

associated with wells and those associated with seismic shot points. Well symbols and their labels are subject to the same general rules that apply to point labeling in other cartographic applications. The labels applied to seismic shot points follow rules peculiar to this application. Shot points are organized along seismic lines which are generally straight and the orientation of the labels is usually perpendicular to the local direction of the seismic line rather than parallel to one of the map borders. This placement is explained in more detail in subsequent sections. For these two types of labels, our algorithm performs label placement and overplot resolution using mathematical optimization techniques that are easy to understand, control, and maintain.

EXAMPLES

The capabilities and procedures of this program are best observed on small problems. As explained later in more detail, the overplot detection and resolution process is iterative. A pattern of label positions is tried and all overplots are noted. Then an optimization algorithm picks the best possible combination of label positions that does not repeat the already detected overplots. New overplots may be created at each iteration, but eventually a solution is produced for which no overplots exist.

Figure 1 shows, from top to bottom, the initial label placement, six intermediate configurations, and the final label placements for a well labeling problem. In this case, eight possible label positions were available for each of 3 wells. The first choice position for each label is above and to the right of the well symbol. Alternate positions in order of preference are indicated below with numbers around a well marked by an x.

4	2	1
6	x	3
8	7	5

Figures 2 and 3 show the initial label placement and final configuration for a problem in seismic shot point labeling. 61 labels are involved. Deletion of shot point labels was an acceptable option for this problem and no overplotting of shot points or line segments was allowed. Three iterations were required to achieve an optimal labeling. Ten labels were deleted from the final solution.

LABEL PLACEMENT OPTIONS

If labels can be placed anywhere on a map and at any orientation, then the automated resolution of overplots is an extremely difficult task. By limiting the number of positions possible for each label, the problem is made easier without significantly affecting the final map quality. Most label placement algorithms assume that placement options are restricted in some manner.

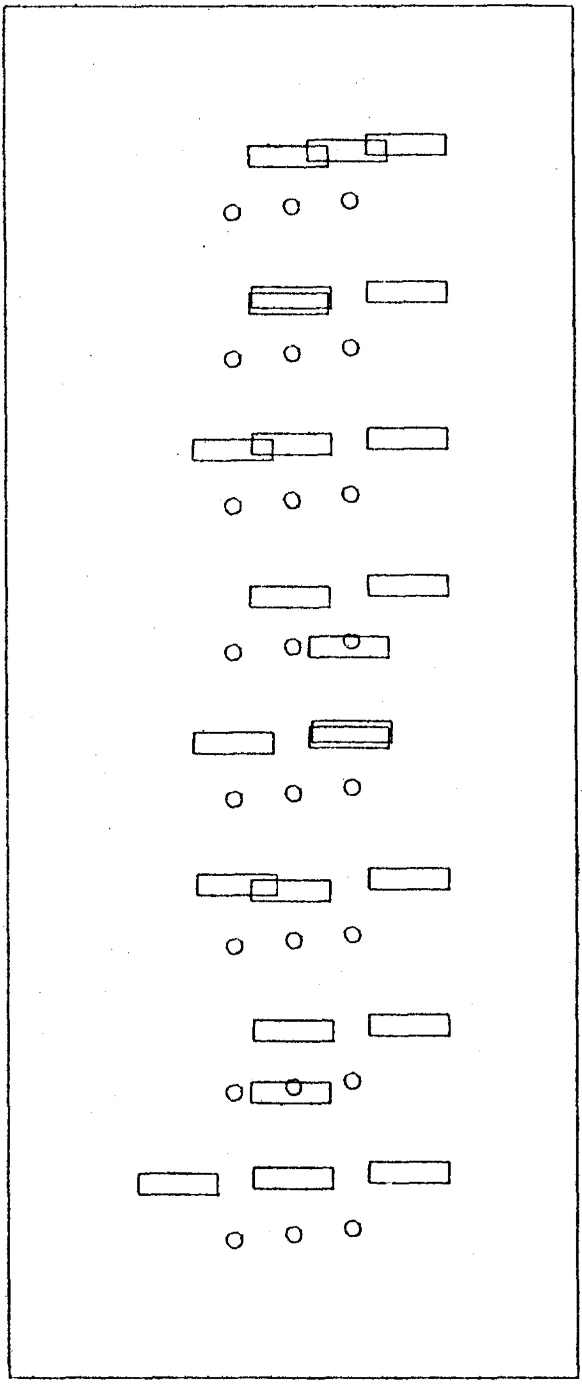

FIGURE 1. *From top to bottom, initial label placement, intermediate configurations, and final label placement for three wells.*

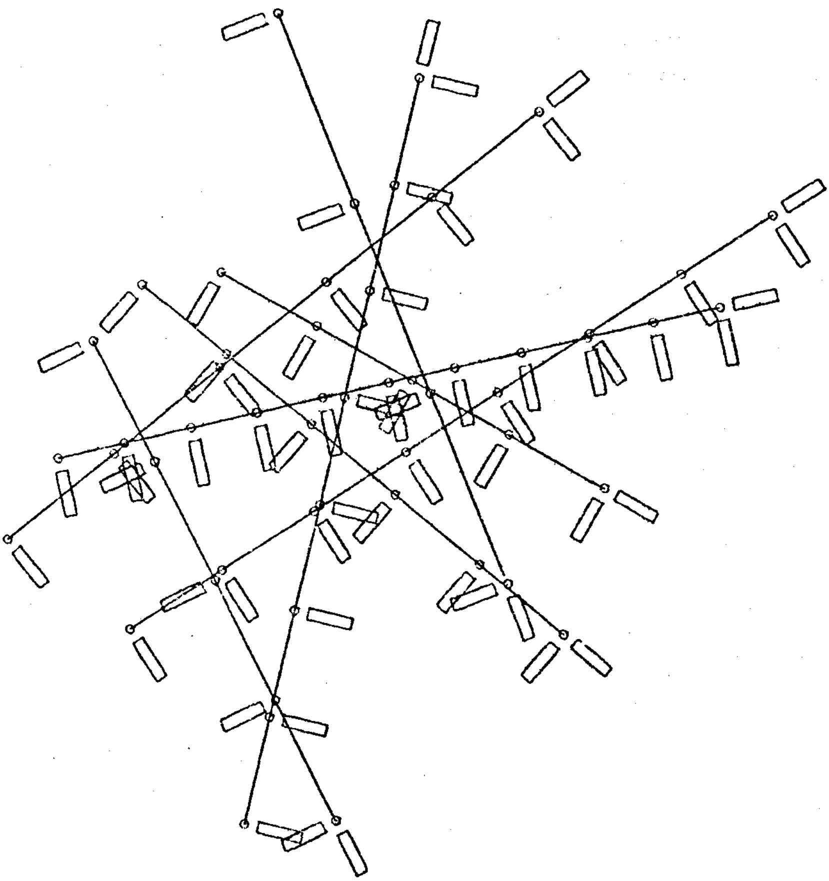

FIGURE 2. *Initial label placement for small seismic line labeling problem. 61 labels are involved.*

Placement of seismic labels

Limiting the number of optional label positions for seismic data is natural. Shot point labels should be plotted perpendicular to a survey line and at a specified distance from the line to make clear which shot points they are associated with. Since labels are usually requested along only one side of a line, there are two alternate positions for each label: on the other side of the line, or deleted from the map. Deletion is an acceptable option because there are many shot points plotted on a single map and information for one deleted label can usually be interpolated from surrounding labels.

Each seismic line has an associated identifying label. Line name labels appear oriented along a continuation of the line at either end. For short lines, labels may be requested only at one end. For long lines, labels must appear at both ends.

Placement of well labels

Limiting the number of placement options for individual point symbols is more difficult to justify because there are few hard rules for the positioning of labels

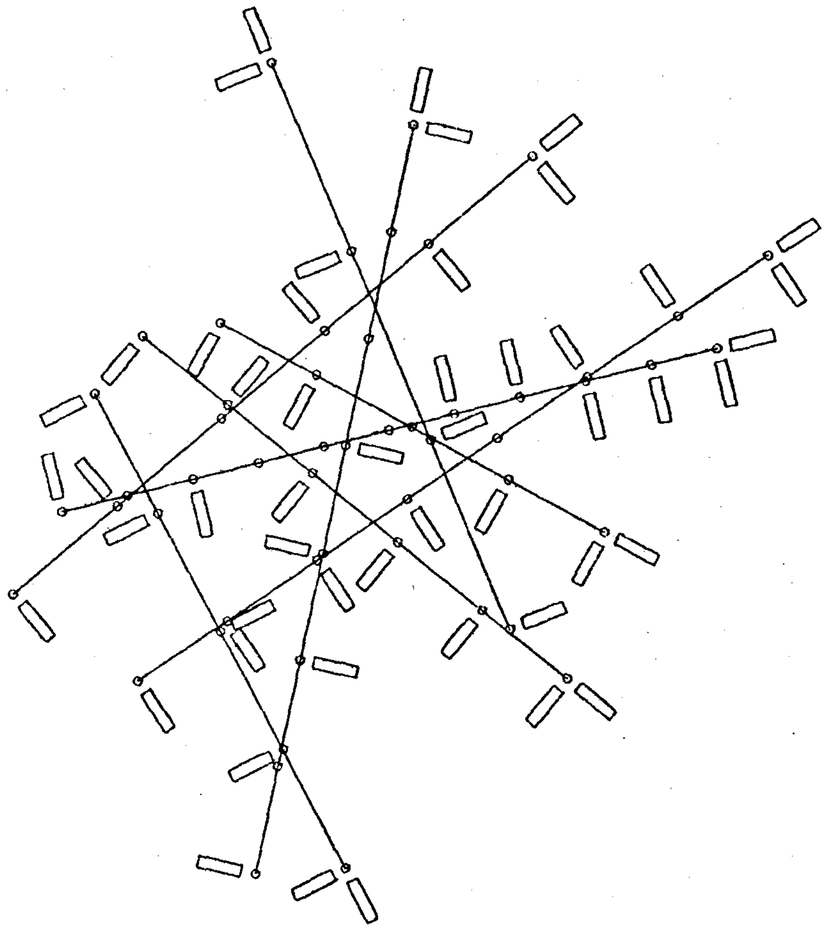

FIGURE 3. *Final label placement for small seismic line labeling problem. 10 labels have been deleted from final configuration.*

with respect to isolated symbols. Still, certain label positions are considered superior to others and consistent patterns must be maintained in order for the map user to quickly associate each label with its correct symbol. In general, the positioning of a label either above, or above and to the right of a symbol is considered the best choice.

Once we begin moving labels, it may become difficult to associate a well symbol with its label since all labels and well symbols are drawn using the same fonts and the same size text. To make the connection clear we often use special well symbols that point to the correct label.

OVERPLOT RESOLUTION

Simple overplot problems can be resolved without complex algorithms. The strength of our overplot resolution program is in its ability to resolve complex problems, in which movement of one label impacts several other labels and may start a chain of events involving scores or hundreds of labels. When complex problems are encountered, a table of conflicts is developed for resolution by an optimization algorithm.

Linear programming and integer programming

Linear programming is a mathematical optimization technique used to minimize (or maximize) the value of a linear objective function subject to linear constraints on the values that can be assumed by problem variables (Chvatal 1983). Given a N element row vector c^T, a M by N matrix A, and a M element column matrix b, the standard linear programming problem is to choose the N elements of the column vector x to minimize

$$c^T x,$$

subject to the constraints

$$Ax \leq b.$$

Integer programming problems are linear programming problems with the added constraint that the elements of x must be integer. $0-1$ integer programming problems require that the elements of x be either 0 or 1. Our overplot resolution program uses $0-1$ integer programming to resolve complex overplot problems.

Many linear programming algorithm implementations are available in the form of software subroutine libraries. Only a few of these implementations handle integer or $0-1$ variables efficiently. For example, documentation for one of the more popular optimization libraries suggests that the number of $0-1$ variables be restricted to less than 200 (Marsten 1985). Fortunately, the structure of the overplot problem may be exploited to obtain efficient solutions using either special purpose software or a mixture of special purpose software and standard algorithms. This problem is discussed in more detail later in this paper.

The overplot problem formulated as a $0-1$ integer programming problem

Assume that each of the K labels on a map can be placed in one of N_k possible positions. in the mathematical formulation of the overplot resolution problem, each position for label k will correspond to a single variable $X_{i,k}$ ($i = 1, 2, \ldots, N_k$) which can take on the values of 0 or 1. Only one variable corresponding to each label is allowed to be 1. This restriction is enforced by the following constraints:

$$\sum_{i=1}^{N_k} X_{i,k} = 1 \qquad k = 1, 2, \ldots, K. \tag{1}$$

If an overplot is possible between the i-th label position for label k and the j-th label position for label m then the following constraint needs to be enforced:

$$X_{i,k} + X_{j,m} \leq 1. \tag{2}$$

Each position for each label will have a penalty $W_{i,k}$ associated with its use. Normally the optimal position will have a penalty of 0 and other positions will have positive penalties proportional to the difficulty caused by attempting to associate a well label in that position with its correct symbol. Our goal is to attempt to minimize the total penalty represented by

$$\sum_{k=1}^{K} \sum_{i=1}^{N_k} W_{i,k} X_{i,k}.$$

subject to the constraints (1) and (2).

Other constraints are possible. For example, we can limit the number of shot point labels that can be deleted from each seismic line. To do this, we add a third subscript to seismic line label variables to indicate which line they are associated with. Then, assuming that the third position choice for each label is deletion, if we do not wish to delete more than P_s labels from line s which has $R_s(>P_s)$ labels we simply add the following constraint:

$$\sum_{k=1}^{R_s} X_{3,k,s} \leq P_s.$$

By adjusting the penalties corresponding to particular label positions, special effects can be achieved. For example, alternate well label positions for older, shallower, or less productive wells can be assigned relatively small penalties making it more likely that the labels for newer, deeper, or more productive wells will appear in their optimal placements.

If a label only slightly overplots a well symbol it may be better to increase the penalty function associated with that position, rather than to forbid it entirely by adding a constraint to the problem. Then the label will be moved to an alternate position only if moving it doesn't overplot other labels. Similar procedures can be used when a label overplots part of a seismic line.

Other objective functions are also possible. For example, it would be easy to choose the label placement combination that minimizes the worst label placement for any single label while satisfying all constraints.

ALGORITHM IMPLEMENTATION

There are interesting technical problems associated with the implementation of any sophisticated cartographic algorithm. Two of the more significant problems encountered during development of this label placement algorithm are discussed in this section.

Overplot detection

The strategy for identifying label overplots is important. Presenting the optimization algorithm with all potential overplots would create unnecessary work, because most potential constraints are never limiting factors in the optimization calculations. As pointed out by Cromley (1985), it is better to proceed in an iterative manner. First, all overplots produced by placing labels in their preferred positions are used to develop pairwise conflict constraints. Then, the label placement is optimized with respect to these known constraints by use of the integer programming algorithm. The result is a new label placement that will produce new overplots. These new overplots are added to the program constraint list and the problem solved again.

To detect overplots efficiently, label placement data must be partitioned or sorted. Sophisticated schemes to organize spatial data, such as quad trees or K-D trees, were considered but rejected as requiring more time to create than would be saved during overplot detection for most problems. Instead, simpler techniques were adopted. Efficient algorithms that find all crossings of N line segments in a rectangular region in N*LOG(N) time are available. These techniques are sometimes based on the space sweep paradigm (Mehlhorn 1984). An implementation of the sweep paradigm, modified for application to rectangles constructed around each label, is used in our algorithm.

Practical integer programming algorithms

Integer programming problems are generally difficult to solve. Special techniques based on the structure of the constraints and of the objective function along with careful problem formulation are required to solve large problems (Johnson et al. 1985). A label placement problem with thousands of variables and hundreds of constraints is a large integer programming problem.

One method that is used to solve large 0 – 1 programing problems is the 'greedy' heuristic, in which all variables are initially set to 0 and variables are set to 1 in successive iterations according to some criterion until no further variables can be set to 1 (Fox and Scudder 1985).

Another approach to this type of problem is to use a standard linear programming algorithm to solve the optimization problem without the 0 – 1 restrictions on the variables. All solution variables for this relaxed problem will be between 0 and 1. Intelligent strategies are then applied to fix certain of the non-integer variables to either 0 or 1, after which the problem is again solved without restrictions on the remaining variables. This process is repeated until all variables are 0 or 1.

A more mathematically complex technique involves creating a Lagrangian relaxation of the original problem and adjusting the Lagrangian multipliers in an iterative manner according to the 'subgradient optimization' method. When using this algorithm, the penalties for label variables corresponding to unsatisfied constraints are increased at each iteration, while the penalties for label variables corresponding to constraints that are not binding are decreased. The amount of change is reduced periodically in a manner that guarantees algorithm convergence (Held, et al. 1974).

Our research has not determined which algorithm is most efficient for this type of integer programming problem. Our best results have usually been based on Lagrangian relaxation, but the algorithm choice may depend on the size of the problem, the type of data being processed (wells or seismic shot points or both), and on how important it is to obtain the exact optimal answer. On most maps, coming reasonably close to the optimal label placement is sufficient.

RESULTS

If deletion of any label is acceptable, then this iterative optimization approach to label placement will always produce a map free of overplots. On maps with moderately crowded point symbols, high-quality results can be expected without deletions. These high-quality results can be obtained quickly even for very large problems.

Figures 4 and 5 show the initial and final label placements for a map with 273 seismic shot point labels. The final configuration was obtained in 5 relaxation iterations requiring a total of 70 CPU seconds on a VAX 750. 170 pairwise overplot conflicts were detected and resolved. Of course, the number of relaxation iterations required and the total CPU time are functions of the label density and of the number of placement options available for each label. Problems with over 2000 seismic shot point labels and more than 700 pairwise label conflicts have been solved in less than 6 CPU minutes on a VAX/750. The computer program used to

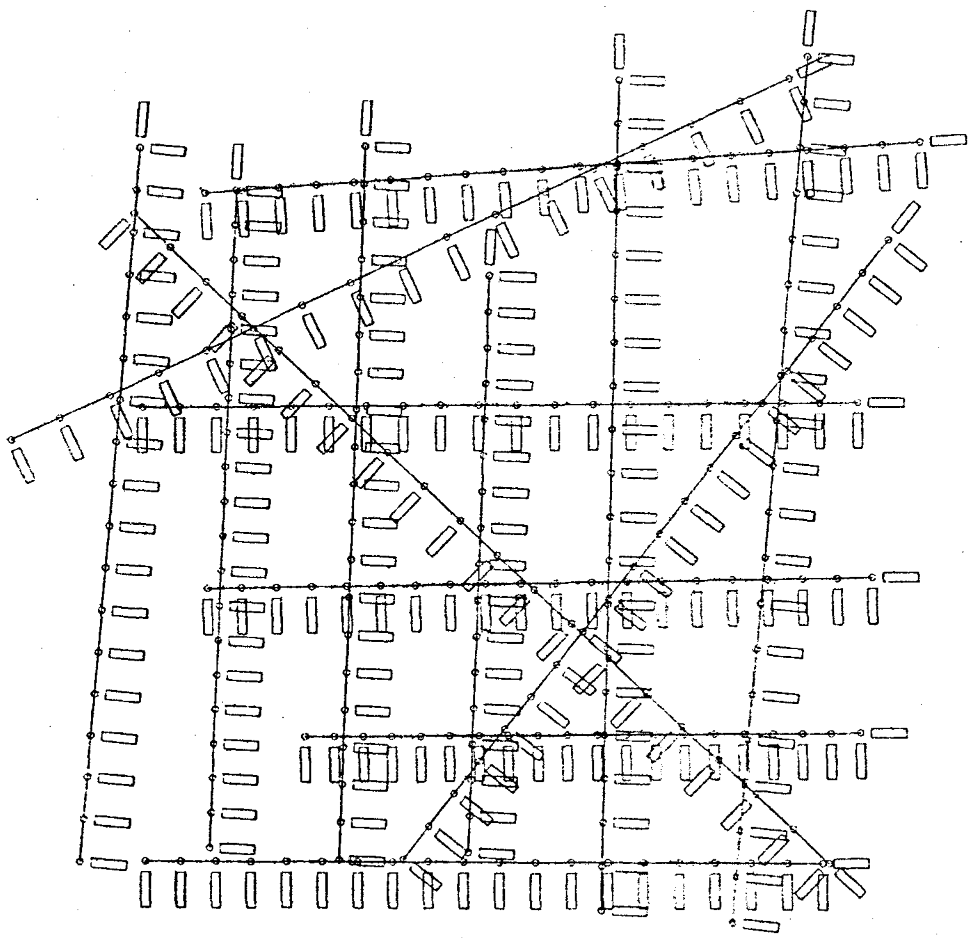

FIGURE 4. *Initial label placement for medium size problem involving 14 seismic lines and 273 labels.*

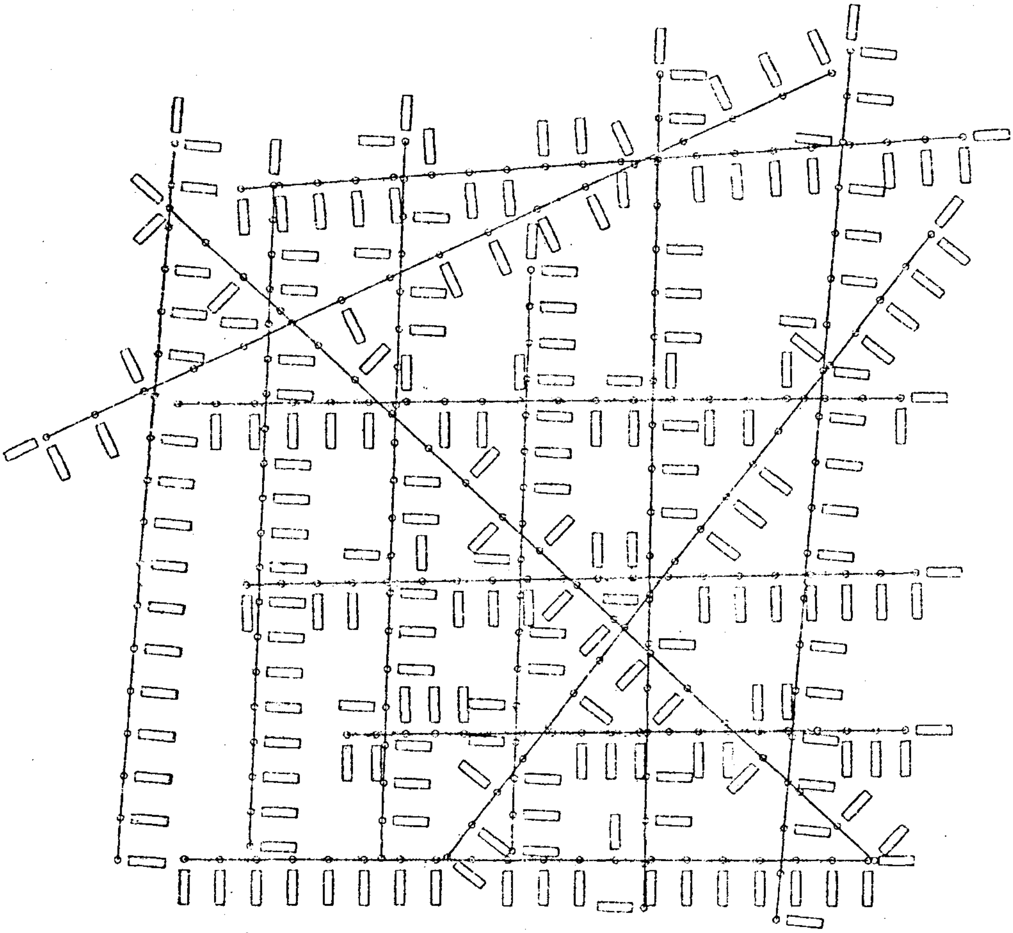

FIGURE 5. *Final label placement for the seismic labeling problem presented in Figure 4. 56 labels have been deleted.*

solve both of these problems, as well as the other examples in this paper, was coded in FORTRAN 77.

DISCUSSION

Optimization techniques are not often used to solve cartographic problems. Much recent research on the solution of difficult cartographic problems has focused on the use of artificial intelligence techniques. Artificial intelligence may work well for some applications, but the label placement problem seems to be solved more easily with traditional mathematical algorithms.

The techniques described in this paper can also be used to assist in the placement of labels on linear features. For example, several choices could be identified for the placement of a contour label. These choices could be assigned penalties, constraints based on conflicts of contour labels with labels for other features developed, and a choice between positions could be made in exactly the same way and at the same time that choices are made for point features.

The selection and placement of features for map display could also be based on optimization techniques. Each feature that is a candidate for display can have an objective function coefficient assigned which depends on whether it could be positioned in its geographically correct location and on its inherent information content. Tables of conflicts with other candidate features could be developed, and the choice of which features to display and where to display them could be based on maximizing the resulting objective function, subject to pairwise interference constraints and constraints on the maximum number of features that could be displayed on the map.

REFERENCES

AHN, J. 1984. Automatic map name placement system. *Image Processing Laboratory Technical Report 063*. Electrical, Computer, and Systems Engineer Department, Rensselaer Polytechnic Institute, Troy, New York.

AHN, J. and FREEMAN, H. 1983. A program for automatic name placement. *Proceedings*, AUTO-CARTO VI, vol. 2, 444–453.

BASOGLU, U. 1984. A new approach to automated name placement systems. PhD. Dissertation. Department of Geography, University of Wisconsin-Madison; available through University Microfilms International.

CHVATAL, V. 1983. *Linear programming*. W.H. Freeman and Company.

CROMLEY, R.C. 1985. An LP relaxation procedure for annotating point features using interactive graphics. *Proceedings*, AUTO-CARTO VII, 127–132.

FOX, G.E., and SCUDDER, G. 1985. A heuristic with tie breaking for certain 0–1 integer programming models. *Naval Research Logistics Quarterly*, vol. 32, 613–623.

HELD, M., WOLFE, P., and CROWDER H. 1974. Validation of subgradient optimization. *Mathematical Programming*, vol. 6, 62–88.

HIRSCH, S.A. 1982. An algorithm for automatic name placement around point data. *The American Cartographer*, vol. 9, no. 1, 5–17.

IMHOF, E. 1975. Positioning names on maps. *The American Cartographer*, vol. 2, no. 2, 128–144.

JOHNSON, E.L., KOSTREVA, M., and SUHL, V. 1985. Solving 0–1 integer programming problems arising from large scale planning models. *Operations Research*, vol. 33, no. 4, July-August, 803–819.

MARSTEN, R. 1985. The XMP Mathematical Programming Library. Department of Management Information Systems, University of Arizona, Tucson, Arizona.

MEHLHORN, K. 1984. *Data structures and algorithms 3: multi-dimensional searching and computational geometry*. Springer-Verlag.

PFEFFERKORN, C., et al. 1985. ACES: A cartographic expert system. *Proceedings*, AUTO-CARTO VII, 399–407.

ROBINSON, A.H., R.D. SALE and J.L. MORRISON, 1978. *Elements of cartography*, Fourth Edition, John Wiley and Sons.

RESUME L'article décrit un programme de placement des écritures à l'aide d'un algorithme d'optimisation mathématique qui détermine la meilleure position de chaque écriture. Ce programme trouve tous les chevauchements d'écriture, replace les écritures pour éliminer le chevauchement ou élimine les écritures si c'est absolument nécessaire. Tout se fait sans intervention humaine. Le programme a été conçu pour la cartographie de la production de l'industrie pétrolière, dont chaque jour chaque carte nécessite des milliers d'écritures et la résolution de centaines de cas de chevauchement. Ce programme remplit cette fonction avec précision et efficience, indépendamment du nombre d'écritures impliquées. Ce succès laisse croire qu'il est raisonnable de considérer l'utilisation de la technique d'optimisation pour aider à solutioner d'autres problèmes en cartographie automatisée, y comprise le positionnement des écritures et la sélection des éléments à montrer sur la carte.

ZUSAMMENFASSUNG Der Artikel beschreibt ein Programm zur Schriftplazierung von Punktsignaturen. Es benutzt einen mathematischen Optimierungsalgorithmus, um die beste Position für jede Beschriftung zu finden. Das Programm entdeckt alle Schriftüberschneidungen, verlagert die Schriftplazierung und streicht notfalls Beschriftungen. Alle Aufgaben werden ohne menschliche Einwirkung durchgeführt. Das Programm wurde für die Kartenherstellung innerhalb der Ölindustrie entworfen, wo auf einer Karte Tausende von Punktsignaturen beschriftet und Hunderte von Plazierungskonflikten gelöst werden müssen. Diese Funktion wird durch das Programm genau und effizient ausgeführt, unabhängig von der Zahl der Beschriftungen. Der Erfolg dieser Anwendungen sollte es ermöglichen, die Optimierungs-verfahren zur Lösung anderer Probleme in der rechnergestützten Kartographie zu verwenden, einschliesslich der Schriftplazierung für Liniensignaturen und der Auswahl darzustellender Kartenzeichen.

RESUMEN Este trabajo describe un programa de colocación de rótulos que utiliza la optimización matemática de un algoritmo para determinar la mejor posición de cada uno de ellos. Este programa detecta toda sobreposición de rótulos, los cambia a nuevas posiciones para resolver problemas de sobreposición y elimina los rótulos cuando sea absolutamente necesario. Se ejecutan todas estas tareas sin intervención humana. El programa está diseñado para uso en la mostración – cartográfica de la producción en la industria petrolera donde, en las operaciones diarias y en un solo mapa, hay que colocar miles de rótulos y resolver cientos de conflictos de ubicación de rótulos. Este programa funciona eficientemente y con precisión, no importa el número de rótulos involucrados. En base del éxito de esta aplicación, es razonable considerar la utilización de técnicas de optimización para resolver los problemas en la cartografía automatizada, incluyendo la colocación de rótulos para razgos lineares y la selección de razgos que desplegarán en el mapa.