SIAM J. DISC. MATH.
Vol. 7, No. 4, pp. 531–543, November 1994



# PREDICTING CAUSE-EFFECT RELATIONSHIPS FROM INCOMPLETE DISCRETE OBSERVATIONS *

E. BOROS[†§], P. L. HAMMER[†§], AND J. N. HOOKER[‡]

**Abstract.** This paper addresses a prediction problem occurring frequently in practice. The problem consists in predicting the value of a function on the basis of discrete observational data that are incomplete in two senses. Only certain arguments of the function are observed, and the function value is observed only for certain combinations of values of these arguments. The problem is considered under a monotonicity condition that is natural in many applications. Applications to tax auditing, medicine, and real estate valuation are discussed. In particular, a special class of problems is identified for which the best monotone prediction can be found in polynomial time.

**Key words.** cause-effect relationship, monotone regression, incomplete observations

**AMS subject classifications.** 62J02, 06A10, 90C09

## 1. Introduction.

The problem of establishing cause-effect relationship based on incomplete observations was studied in [7]. In this paper we address the problem of finding a good approximation of an unknown discrete function on the basis of a set of observations, which is incomplete in two senses. We observe the values of only some of the arguments of the function and we observe the function value only for certain combinations of values of these arguments. Our goal is to predict the value of the function for any combination of values for these arguments. Such problems occur frequently, and we begin with some examples.

Suppose that a tax bureau must decide which tax forms to audit, with the goal of auditing only those for which the increased return justifies the auditing expense. The observation set consists of data records corresponding to forms audited in the past. Each record indicates whether the audit was justified and lists some attributes of the taxpayer. The problem is to determine, on the basis of these attributes, when future tax forms should be audited.

A taxpayer's attributes form a vector $x$ of Boolean (0-1) variables, where each $x_j$ indicates whether a certain threshold is exceeded. For example, one $x_j$ may have the value 1 when the taxpayer claims too many dependents for a person of his age, and another may have the value 1 when he claims too many charitable contributions relative to his income.

The observation set partially defines a function $g$ whose value is 1 when auditing was justified, and 0 otherwise. Whether an audit was justified depends not only on the recorded attributes used to make the auditing decision, but on a number of hidden factors as well. The arguments of $g$ therefore consist of a vector $x$ of the recorded

* Received by the editors April 16, 1991; accepted for publication (in revised form) July 16, 1993.

† Rutgers Center for Operations Research (RUTCOR), Rutgers University, P.O. Box 5062, New Brunswick, New Jersey 08903.

‡ Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. This author's research was supported by Air Force Office of Scientific Research grant 91-0287 and Office of Naval Research grant N00014-92-J-1028.

§ This author's research was partially supported by Air Force Office of Scientific Research grants 89-0512B and F49620-93-1-0041 and Office of Naval Research grants N00014-92-J-1375 and N00014-92-J-4083.

attributes and a vector $y$ of hidden attributes. So, each data record can be written in the form $(x, g(x,y))$. In practice, the function $g$ is only partially defined because its value $g(x,y)$ is not given for every possible vector $(x,y)$ of attributes. We cannot even assume there is an observation for every possible $x$.

Our problem is to predict the value of $g(x,y)$ on the basis of $x$ alone. To do this, we derive an *approximation* $f$ of $g$ that is a function of $x$ only. Deriving $f$ requires that we interpolate a value when $x$ has not been observed. However, even when $x$ has been observed, we may need to reconcile different observed values of $g(x,y)$ for different $y$'s.

We focus on problems in which $f$ is *monotone*, in the sense that $f(x) \not\succ f(x')$ whenever $x \preceq x'$, where $\prec$ is a partial order on the outcomes and $\prec$ a partial order on the $x$'s. In the tax auditing example, the two outcomes are linearly ordered ($0 \prec 1$), and $\prec$ is defined by $x \preceq x'$ whenever $x_j \leq x'_j$ for all $j$. Monotonicity is a reasonable assumption for this problem, because every 1 among the attributes is another reason for suspecting that the form should be audited. If one form with certain suspicious traits should be audited, then another form with these and still other suspicious traits should certainly be audited.

We also generalize our approach to problems in which more than two outcomes are possible, where these outcomes are partially ordered. The tax auditor, for example, might classify forms as needing no audit, needing a second look to determine the desirability of an audit, and needing an audit right away. These outcomes happen to be linearly ordered in a natural way (in the order listed), and it is reasonable to assume that $f$ is monotone.

The attributes as well as the outcome can have more than two values. Suppose, for example, that we have a battery of 45 biochemical tests for carcinogenicity, each of which can have outcomes *negative*, *indefinite*, and *positive*. We do not apply every test to a given chemical; when a test is not applied, we say that the test "result" is *no data*. The 45 test results are viewed as attributes of the chemical tested. Since the tests are not foolproof, we first apply them to a number of chemicals that have been previously tested clinically for carcinogenicity, with possible outcomes *harmless*, *undetermined* (indicating inconclusive clinical experience), *potentially dangerous* (meaning that the chemical causes cancer when ingested in very large doses), and *dangerous*. The problem is to find a function $f$ that predicts the outcome of clinical trials for a new chemical.

To check for monotonicity, we must impose a partial ordering on the attribute vectors and on the outcomes. The four possible values of each attribute $x_j$ submit to a partial order: *negative* is less than *indefinite* and *no data*, which are less than *positive*; but *indefinite* and *no data* are incomparable. We can therefore say that $x \preceq x'$ when $x_j \preceq x'_j$ for all $j$. The four outcomes have a similar partial ordering: *harmless* is less than *potentially dangerous* and *undetermined*, both of which are less than *dangerous*; but we may be unable to order *potentially dangerous* and *undetermined* with respect to each other. It is reasonable to assume that $f$ is monotone with respect to these orderings.

Whenever $f$ incorrectly predicts the outcome for a given observation, a penalty is incurred, where the size of the penalty depends on the correct and predicted outcomes. For us, the best approximation $f$ is one that minimizes the total penalty. In [4] we investigate the issue of defining "best approximation" from a statistical point of view and distinguish the analysis described here from logit and categorical data analysis.

If there is no restriction on $f$, finding the best approximation is relatively easy, since we can treat each $x$ separately. That is, for any fixed $x$, $f(x)$ should have the value

that minimizes the sum of the penalties over observations of the form $(x, y)$. If $x$ occurs in no observations, $f(x)$ can be set to an arbitrary value. However, if $f$ is required to be monotone (or to have some other restrictive property), the problem cannot in general be decomposed this way, and the observed values of $f(x)$ for unobserved $x$ are in general restricted. In [4] we consider some other possible restrictions on $f$ (when it is a Boolean function). More examples for similar problems can be found in [6], [21], [22].

In this paper, we show that a network flow model can be used to determine the best approximation $f$ when the partial ordering of the outcomes belongs to a special class of partial orders, including interval orders. An *interval order* is one in which every element can be associated with an interval of real numbers, such that $\alpha \prec \beta$ if and only if the upper end of the interval associated with $\alpha$ is smaller than the lower end of that associated with $\beta$. A special case of an interval order is one that is layered, in the sense that the elements are partitioned into a sequence of sets $S_1, \ldots, S_m$, such that all the elements in each $S_i$ are incomparable, but everything in $S_i$ is less than or equal to everything in $S_{i+1}$. The outcomes of the clinical trials for carcinogenicity have this sort of ordering. A linear order, as in the tax auditing problem, is, of course, an interval order.

Another instance that might call for an interval ordering is estimating the value of a piece of property. Suppose that we suspect that certain combinations of property attributes justify assessing the value to be within certain ranges. The ranges may overlap, but the outcomes nonetheless have an interval order. If higher-ranked attribute values are more desirable, it may be reasonable to assume monotonicity.

We begin in the next section with a precise statement of the problem and a small example. In §3 we show that finding the best monotone approximation is equivalent to a generalization of the "maximal closure" problem on a directed graph [16]. When there are two ordered outcomes, the generalization coincides with the maximal closure problem. In §4 we describe a special class of partial orders, the so-called aligned orders. We prove that aligned orders can be recognized in polynomial time and that interval orders are aligned. Finally, in §5 we show that the generalized maximal closure problem can be reduced to the maximal closure problem when the outcomes form an aligned order. Since the maximal closure problem can be solved via a minimum cut computation, we can solve the best approximation problem likewise. The complexity of our algorithm is $O(n^3m^3)$, where $n$ is the number of distinct input attribute vectors $x$ and where $m$ is the number of possible outcomes.

**2. Problem statement and main results.** Let $(X, \prec)$ be a finite partially ordered set of (observed) attribute vectors $x$ and $(V, \prec)$ a finite partially ordered set of outcomes. The mapping $\mu : X \times V \longmapsto \mathbb{Z}$ indicates the number of times an attribute vector $x$ resulted in a given outcome; that is, $\mu(x, \alpha)$ is the number of observations $(x, y)$ for which $g(x, y) = \alpha$. The penalty function is $c : V \times V \longmapsto \mathbb{Z}$, where $c_{\beta,\alpha}$ is the penalty for predicting outcome $\beta$ when the observed outcome is $\alpha$. The problem is to find a *V-monotone* function $f(x)$, i.e., one for which $f(x) \not\succ f(x')$ whenever $x \preceq x'$, so as to minimize the total penalty,

$$\varepsilon[f] = \sum_{x \in X} \sum_{\alpha \in V} c_{f(x),\alpha} \mu(x, \alpha). \tag{2.1}$$

As a matter of fact, we want to define the function $f(x)$ only on the set $X$ of observed attribute vectors. For an unobserved value $x'$, we can let $f(x')$ take any value for which $f(x') \not\succ f(x)$ for all observed values $x$ satisfying $x' \preceq x$ and for which $f(x) \not\succ$

$f(x')$ for all observed values $x$ satisfying $x \preceq x'$. The above problem is called the *monotone approximation problem*, and an instance of this problem is denoted by $\mathcal{M} = (X, V, c, \mu)$.

Let us observe that the objective function in the above problem is separable in the sense that it can be written as

$$\varepsilon[f] = \sum_{x \in X} h_x(f(x)),$$

where

$$h_x(f(x)) = \sum_{\alpha \in V} c_{f(x),\alpha} \mu(x, \alpha),$$

and the function $h_x$ depends only on $x$ but not on other elements of $X$. The problem of separable optimization subject to precedence (-like) constraints, in general, can be formulated as follows:

$$\min \sum_{x \in X} h_x(f(x))$$

$$\text{s.t. } f(x) \in V \quad \text{for all } x \in X, \quad \text{and}$$

(SOP)

$$f(x) \not\succ f(y) \quad \text{whenever} \quad x \prec y.$$

Here $(X, \succ)$ and $(V, \succ)$ are partially ordered sets, $X$ is a finite set, and $h_x : V \to \mathbb{R}$ are given real-valued functions for $x \in X$.

If $(V, \prec)$ is the set of reals with the usual order $<$ and if $h_x$ are convex functions, (SOP) is known in the statistical literature as the *isotonic regression* problem; see, e.g., [2]. There is a bewildering variety of algorithms created to solve this problem in various special cases, many of which are not polynomial. An equivalent problem was considered in [14], [19] in the context of inventory/production systems. An $O(|X|^4)$ algorithm was provided in these papers for the case when the functions $h_x$ are special convex differentiable functions for $x \in X$. For general convex functions and for the case of handling general upper and lower bounds, an algorithm of similar complexity is given in [18]. For a somewhat more general class of objective functions, a different algorithm of the same complexity is presented in [5]. The methods of the last two papers extend to the case when $(V, \succ)$ is a linearly ordered discrete (finite or infinite) set. In the further specialized case, when $(X, \succ)$ is a linear order and $h_x$ are convex quadratic real functions, an $O(|X|)$-time algorithm is provided in [3].

None of the above methods can handle the problem (SOP) in polynomial time if the objective function is not special (usually convex, or something almost as restrictive). The authors are not aware of any results for the case where $(V, \succ)$ is not linearly ordered.

In this paper, we address the problem (SOP) in the case when $(V, \succ)$ is a finite partially ordered set and no restrictions are imposed on the functions $h_x, x \in X$. The main result of this paper is that the problem (SOP) can be solved in polynomial time in the case when $(V, \succ)$ belongs to a special class of partially ordered sets. We say that a partial order $(V, \succ)$ is *aligned* if there is a linear extension of $\succ$, i.e., a labeling $V = \{\nu_0, \ldots, \nu_v\}$ of the elements of $V$ satisfying the following two conditions:

TABLE 1
*A sample observation set.*

| | | | $\mu(x,\alpha)$ | | | | $\sum_\beta c_{\alpha\beta}\mu(x,\beta)$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $h$ | $u$ | $p$ | $d$ | $h$ | $u$ | $p$ | $d$ | $f_0$ | $f_1$ | $w_{xh}$ | $w_{xu}$ | $w_{xp}$ | $w_{xd}$ |
| 1 | $N$ | $N$ | 2 | 1 | 0 | 0 | 2 | 4 | 5 | 8 | $h$ | $h$ | −2 | −2 | −3 | −1 |
| 2 | $I$ | $N$ | 1 | 1 | 0 | 1 | 5 | 4 | 5 | 5 | $u$ | $u$ | −5 | 1 | 0 | −1 |
| 3 | $I$ | $U$ | 1 | 0 | 0 | 0 | 0 | 2 | 2 | 3 | $h$ | $u$ | 0 | −2 | −2 | 1 |
| 4 | $P$ | $U$ | 1 | 1 | 2 | 0 | 6 | 4 | 3 | 9 | $p$ | $p$ | −6 | 2 | 3 | −8 |
| 5 | $I$ | $P$ | 0 | 0 | 1 | 2 | 8 | 5 | 4 | 2 | $d$ | $d$ | −8 | 3 | 4 | −1 |

(C1) $\forall \nu_i \prec \nu_j \quad \exists$ an index $k$ such that $i \le k$, $k+1 \le j$, and $\nu_k \prec \nu_{k+1}$;

(C2) $$\nu_k \prec \nu_{k+1} \Rightarrow \nu_j \prec \nu_{k+1} \quad \forall j \le k.$$

We show that (SOP) can be solved in $O(|X|^3|V|^3)$ time if $(V,\succ)$ is aligned. We also show that, given a partial order $(V,\succ)$, we can recognize in $O(|V|^2)$ time whether it is aligned, and we prove that interval orders are aligned.

We use the following as a running example throughout the remainder of the paper. Suppose that we wish to determine the carcinogenicity of chemicals on the basis of only two tests. We have clinical data for five chemicals (Table 1). $N, U, I$, and $P$, respectively, denote the possible test results (*negative*, *no data*, *indefinite*, and *positive*), and $h, u, p$, and $d$ denote the four possible outcomes (*harmless*, *undetermined*, *potentially dangerous*, and *dangerous*). The table displays the number $\mu(x,\alpha)$ of times chemicals with attribute vector $x$ lead to clinical result $\alpha$.

To define the penalty for error, let us rank the outcomes as follows: $h$ has rank 1, $u$ and $p$ have rank 2, and $d$ has rank 3. Then we might say that the penalty for an error is 1 more than the distance between the ranks of the predicted and observed outcome when the outcomes are different, and 0 otherwise. For instance, predicting a potentially dangerous ($p$) chemical to be harmless ($h$) brings penalty 2, and predicting it to be undetermined ($u$) brings penalty 1. The total penalty $\sum_\beta c_{\alpha\beta}\mu(x,\beta)$ for predicting that chemicals described by $x$ have clinical outcome $\alpha$ appears in the middle of the table. For instance, the penalty for predicting that chemicals with test results $(N,N)$ are dangerous is 8, because they were observed to be harmless on two occasions and to have undetermined effect on one occasion.

If there are no restrictions on the approximating function $f$, the best approximation is clearly the function $f_0$ shown in the table. We set $f_0(N,N) = h$, for instance, since predicting $h$ results in the least penalty ($\sum_{\beta\in V} c_{h,\beta}\mu((N,N),\beta) = 2c_{h,h} + c_{h,u} = 2$). In general, $f_0$ is defined such that the sum $\sum_{\beta\in V} c_{f_0(x),\beta}\mu(x,\beta)$ is minimal for $x \in X$. In this example, however, $f_0$ is not monotone, since $f_0(I,N) \succ f_0(I,U)$ even though $(I,N) \preceq (I,U)$. Our task is to find the best monotone approximation.

**3. Maximal $V$-partitions.** Let $G = (N, A, w)$ be a directed graph, where $N$ denotes the set of nodes and where $A$ denotes the set of arcs. A subset $C \subseteq N$ of the nodes is called a *terminal set* of $G$ if $(x,y) \in A, x \in C$ implies that $y \in C$. Clearly, the intersection of two terminal sets is again a terminal set. For a given subset $S$ of the nodes, its *closure* $\mathrm{cl}(S)$ is defined as the minimal terminal set containing $S$. In

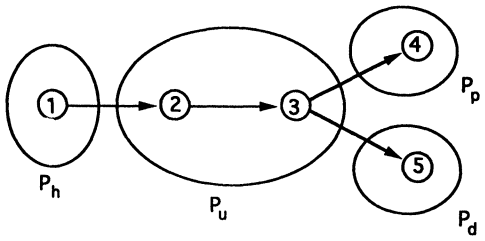


FIG. 1. *Illustration of a V-chain.*

other words, $\mathrm{cl}(S)$ contains exactly those vertices $y$ of $G$ that are either belonging to $S$ or accessible from $S$ by a directed path of $G$. Given real weights $w_x, x \in N$ associated with the nodes, a *maximal closure* of the graph $G$ is a terminal set $C$ for which $w(C) = \sum_{x \in C} w_x$ is maximal.

To an instance $\mathcal{M} = (X, V, \mu, c)$ of the monotone approximation problem, we associate a directed graph $G_{\mathcal{M}} = (X, A)$ in which the observed attribute vectors are the nodes and where there is an arc from $x$ to $y$ whenever $x \prec y$.

As an example, consider the graph $G_0$ in Fig. 1, which has nodes corresponding to the five values of $x$ in Table 1. (Ignore for now the large sets circled in Fig. 1.) $G_0$ contains an arc $(x, x')$ whenever $x \preceq x'$. The closure of a set $C$ of nodes ($x$'s) is the set of all $x$'s greater than or equal to the $x$'s in $C$. For instance, the closure of $\{3\}$ is $\{3, 4, 5\}$.

Picard [16] introduced the problem of finding a maximal closure in a given directed graph and showed that the maximal closure can be found as the source side of a minimum cut in an associated network. This problem has many interesting applications, e.g., [1], [11]–[13], [15], [17], [20], [23].

In this section, we introduce a generalization of this problem, which is shown to be equivalent to the problem of $V$-monotone approximation and which is reducible in certain cases to the problem of maximal closures.

Let us consider a partially ordered finite set $(V, \prec)$ and let $\alpha || \beta$ denote the fact that $\alpha$ and $\beta$ are incomparable elements of $V$. A partition $\mathcal{P} = \{P_\alpha | \alpha \in V\}$ of the nodes of $G$ (i.e., $\bigcup_{\alpha \in V} P_\alpha = N$) is called a *V-partition* if

$$\forall (i, j) \in A, \quad i \in P_\alpha \quad \text{and} \quad j \in P_\beta \quad \text{imply that } \alpha \not\succ \beta. \tag{3.1}$$

For example, suppose that we let $V = \{h, u, p, d\}$, as in Table 1. One possible $V$-partition is given by the circled sets in Fig. 1, that is, $P_h = \{1\}, P_u = \{2, 3\}, P_p = \{4\}$, and $P_d = \{5\}$.

Note that the sets in this $V$-partition are determined by the function $f_1$, displayed in Table 1. Namely, the set $P_d$ is the set of nodes classified by $f_1$ as dangerous $(f_1(x) = d), P_p$ is the set of nodes classified as potentially dangerous $(f_1(x) = p), P_u$ is the set of nodes classified as undetermined $(f_1(x) = u)$, and $P_h$ is the set of nodes classified as harmless $(f_1(x) = h)$.

Note that the function $f_1$ that defines the above $V$-partition of Fig. 1 is $V$-monotone. In fact, we have the following result.

THEOREM 1. *Let $G_{\mathcal{M}} = (X, A)$ be the directed graph associated with a given instance of the monotone approximation problem. Then there is a one-to-one correspondence between the V-monotone functions and the V-partitions of G.*

*Proof.* Let us assume first that $f : X \longmapsto V$ is a $V$-monotone function. For $\alpha \in V$, let

$$P_\alpha = \{x \mid f(x) = \alpha\} \tag{3.2}$$

and let $\mathcal{P}_f = \{P_\alpha | \alpha \in V\}$. We show that $\mathcal{P}_f$ is a $V$-partition.

Now condition (3.1) is implied by the monotonicity of $f$, as follows. Let $(x, y) \in A$ be an arbitrary arc with $x \in P_\alpha$ and with $y \in P_\beta$. We now have $\alpha = f(x)$ and $\beta = f(y)$ by (3.2). The monotonicity of $f$ implies that

$$f(x) \not\succ f(y). \tag{3.3}$$

It follows then that either $\alpha \preceq \beta$ or $\alpha || \beta$, and hence (3.1) follows.

For the converse relation, let $\mathcal{P} = \{P_\alpha | \alpha \in V\}$ be a $V$-partition of $G_\mathcal{M}$. Then the function $f_\mathcal{P}$, defined by

$$f_\mathcal{P}(x) = \alpha \quad \text{if } x \in P_\alpha, \tag{3.4}$$

is clearly an $X \longmapsto V$ mapping, and property (3.1) immediately implies its $V$-monotonicity. □

Let $w_{x,\alpha}$ be given real weights for $x \in N$ and $\alpha \in V$ and let the weight $w(\mathcal{P})$ of the $V$-partition $\mathcal{P}$ be defined as

$$w(\mathcal{P}) = \sum_{\alpha \in V} \sum_{\beta \succeq \alpha} \sum_{x \in P_\beta} w_{x,\alpha}. \tag{3.5}$$

Then a *maximum-weight V-partition* of the weighted graph $G = (N, A)$ is a $V$-partition $\mathcal{P}$, which maximizes $w(\mathcal{P})$.

Let us again consider the directed graph $G_\mathcal{M} = (X, A)$ associated with an instance $\mathcal{M} = (X, V, \mu, c)$ of the monotone approximation problem. Let us now define weights $w_{x,\alpha}$ for $x \in X$ and $\alpha \in V$ by the following set of equations:

$$\sum_{\beta \preceq \alpha} w_{x,\beta} = -h_x(\alpha) \left( = -\sum_{\beta \in V} c_{\alpha,\beta} \mu(x, \beta) \right) \quad \text{for all } \alpha \in V. \tag{3.6}$$

It is easy to observe that, if we arrange the elements of $V$ according to an arbitrary linear extension of $\prec$, then the coefficient matrix on the left-hand side of (3.6) becomes a lower triangular matrix with 1's in the main diagonal. Hence, for any integral parameter vectors $c$ and $\mu$, the system of equations (3.6) has a unique integral solution.

We show below that the problem of monotone approximation is equivalent to a maximum-weight $V$-partition problem in the associated graph.

We know that finding a $V$-partition is equivalent to finding a monotone approximation $f$, but, to equate the maximal $V$-partition problem to the monotone approximation problem, we must relate the weight of a $V$-partition to the penalty associated with an approximation. This is done in the example of Table 1 as follows. The negative of the weight $w_{xh}$ contributed by each $x$ lying in $P_h$ to a $V$-partition is the penalty that results from predicting chemicals with attributes $x$ to be harmless. The negative of the total weight $w_{xh} + w_{xu}$ contributed by each $x$ in $P_u$ is the penalty for predicting chemicals with attributes $x$ to be undetermined, and similarly for $w_{xh} + w_{xp}$. The negative of the total weight $w_{xh} + w_{xu} + w_{xp} + w_{xd}$ contributed by each $x$ in $P_d$ is the penalty for predicting $x$ to be dangerous. The resulting weights $w_{x\alpha}$ for each outcome $\alpha$ appear on the right side of Table 1. Suppose, for instance, that the predictions are given by $f_1$. The penalty incurred for each $x$ is shown in boldface in the middle of the table. The penalty for each $x$ is the same as the negative of the sum of the boldface weights that appear at the end of that row. It is clear that finding a maximal-weight

$V$-partition over these weights is equivalent to finding predictions that minimize the penalty.

THEOREM 2. *Given an instance $\mathcal{M} = (X, V, c, \mu)$ of the monotone approximation problem and an arbitrary $V$-monotone function $f : X \longmapsto V$, let $\mathcal{P}_f = \{P_\alpha | \alpha \in V\}$ be the $V$-partition of the graph $G_{\mathcal{M}}$, defined by* (3.2). *Then*

$$\varepsilon[f] = -w(\mathcal{P}_f), \tag{3.7}$$

*where $w_{x,\alpha}, \alpha \in V$ is the unique solution of* (3.6) *for every $x \in X$ and where $w(\mathcal{P}_f)$ is defined by* (3.5).

*Proof.* We prove (3.7) as follows.

$$\begin{aligned} w(\mathcal{P}_f) &= \sum_{\alpha \in V} \sum_{\beta \succeq \alpha} \sum_{x \in P_\beta} w_{x,\alpha} \\ &= \sum_{\alpha \in V} \sum_{\beta \succeq \alpha} \sum_{f(x)=\beta} w_{x,\alpha} \\ &= \sum_{x \in X} \sum_{\alpha \preceq f(x)} w_{x,\alpha} \\ &= \sum_{x \in X} \left( -\sum_{\beta \in V} c_{f(x),\beta} \mu(x, \beta) \right) \\ &= -\sum_{x \in X} \sum_{\beta \in V} c_{f(x),\beta} \mu(x, \beta) \\ &= -\varepsilon[f]. \end{aligned} \tag{3.8}$$

Here, the second line comes from the first one by the definition of $\mathcal{P}_f$, the third line comes by a simple rearrangement, and the fourth equation is obtained by applying (3.6). For the last equation, we simply applied (2.1). □

COROLLARY 1. *For a given instance $\mathcal{M} = (X, V, c, \mu)$ of monotone approximation, let $w = (w_{x,\alpha} | x \in X, \alpha \in V)$ be determined by* (3.6). *Then finding the best $V$-monotone fit to $\mathcal{M}$ is equivalent to finding the maximum-weight $V$-partition in the associated graph $G_{\mathcal{M}}$.*

**4. Aligned partial orders.** In this section, let us consider aligned partial orders $(V, \prec)$, i.e., partial orders satisfying conditions (C1) and (C2).

Let us first prove that this property of a partial order can be recognized in polynomial time.

LEMMA 1. *Given a partial order $(V, \succ)$, one can either find a labeling $V = \{\nu_0, \nu_1, \ldots, \nu_v\}$ satisfying conditions* (C1) *and* (C2) *or conclude that $(V, \succ)$ is not aligned in $O(|V|^2)$ time.*

*Proof.* For an $\alpha \in V$, let $S_\alpha = \{\beta | \beta \prec \alpha\}$. Furthermore, let $\gamma \in V$ such that $|S_\gamma|$ is maximal.

We show that $(V, \succ)$ is aligned if and only if the induced relation $(V \setminus S_\gamma, \succ)$ is an empty relation and the induced partial order $(S_\gamma, \succ)$ is aligned. The recognition algorithm and its complexity follows immediately from this.

Let us first assume that $(V \setminus S_\gamma, \succ)$ is an empty relation and that the induced partial order $(S_\gamma, \succ)$ is aligned. Let $S_\gamma = \{\nu_0, \ldots, \nu_{|S_\gamma|-1}\}$ be an appropriate labeling of the elements of $S_\gamma$ satisfying conditions (C1) and (C2). It is easy to verify that, by

defining $\nu_{|S_\gamma|} = \gamma$ and taking an arbitrary labeling $\{\nu_{|S_\gamma|+1}, \ldots, \nu_v\} = V \setminus (S_\gamma \cup \{\gamma\})$, we can obtain a labeling of $V$ that satisfies conditions (C1) and (C2).

To see the converse direction, let us assume that $(V, \succ)$ is an aligned partial order and the labeling $V = \{\nu_0, \ldots, \nu_v\}$ satisfies conditions (C1) and (C2). Let $k$ be the largest index for which $\nu_{k-1} \prec \nu_k$. Then $S_{\nu_k} = \{\nu_j | j = 0, \ldots, k-1\}$ follows from condition (C2). Furthermore, condition (C1) implies that there are no indices $j > i \geq k$ such that $\nu_j \succ \nu_i$. It follows then that $S_{\nu_i} \subseteq S_{\nu_k}$ for any $i \neq k$, and hence $|S_{\nu_k}|$ is maximal. It is also clear that the induced partial order $(S_{\nu_k}, \succ)$ is also aligned (with the same labeling) and that the induced partial order $(\{\nu_k, \ldots, \nu_v\}, \succ)$ is an empty relation.

The only thing left to show is that the claim remains true for any other index $k'$ for which $|S_{\nu_{k'}}|$ is also maximal. Let us observe that, in this case, $k' > k$ and $S_{\nu_{k'}} = S_{\nu_k}$ follow from properties (C1) and (C2). Therefore, by interchanging the elements $\nu_k$ and $\nu_{k'}$, we obtain another good labeling of $V$, which proves the lemma. □

We show next that interval orders are aligned.

LEMMA 2. *Interval orders are aligned.*

*Proof.* Let us recall (see [9]) that $(V, \prec)$ is an interval order, by definition, if there are real intervals $[a_\alpha, b_\alpha] \subseteq \mathbb{R}$ for $\alpha \in V$ such that $\alpha \prec \beta$ if and only if $b_\alpha < a_\beta$. Let us consider such a set of intervals realizing $(V, \succ)$.

It is clear that

$$S_\alpha = \{\beta \,|\, b_\beta < a_\alpha\}.$$

It follows then that $S_\alpha \subseteq S_\beta$ if $a_\alpha \leq a_\beta$; therefore $|S_\alpha|$ is maximal if and only if $a_\alpha$ is maximal. Let $\alpha \in V$ be such that $a_\alpha$ is maximal. Then, for every $\beta \in V$, either $b_\beta < a_\alpha$ (in which case, $\beta \in S_\alpha$) or $a_\beta \leq a_\alpha \leq b_\beta$. This implies that there is a common point $a_\alpha$ of all intervals associated to the elements of $V$ that are not in $S_\alpha$, and hence there cannot be any relation between these elements.

Since $(S_\alpha, \succ)$ is again an interval order and since $|S_\alpha| < |V|$, we can continue, and the recognition algorithm described in the previous lemma will not fail. □

It is easy to see that aligned orders are not necessarily interval orders. For example, let us consider $V = \{\nu_0, \nu_1, \nu_2, \nu_3, \nu_4\}$, in which $\nu_0 \prec \nu_2, \nu_0 \prec \nu_4, \nu_1 \prec \nu_2$, and $\nu_1 \prec \nu_3$. This partial order is clearly aligned, but it is not interval, since neither $\nu_0 \prec \nu_3$ nor $\nu_1 \prec \nu_4$.

**5. Maximal $V$-partitions via maximal closures.** In this section, we show that if $(V, \prec)$ is an aligned partial order, then a maximal weight $V$-partition can be determined in polynomial time.

Let $G = (N, A)$ be a given directed graph, let $(V, \prec)$ be an aligned order, and let $V = \{\nu_0, \ldots, \nu_v\}$ be a labeling of the elements of $V$ satisfying conditions (C1) and (C2). Furthermore, let $w_{x,\alpha}$ be given reals for $x \in N$ and $\alpha \in V$.

It follows then that, for any $V$-partition, $\mathcal{P} = \{P_\alpha | \alpha \in V\}$ and, for any index $i$ for which $\nu_{i-1} \prec \nu_i$,

$$\hat{P}_{\nu_i} = \bigcup_{j=i}^{v} P_{\nu_j}. \tag{5.1}$$

In the example, we can let $\{\nu_0, \nu_1, \nu_2, \nu_3\} = \{h, u, p, d\}$ as an appropriate labeling. As we can observe easily, this partial order is not only aligned but is also an interval order.

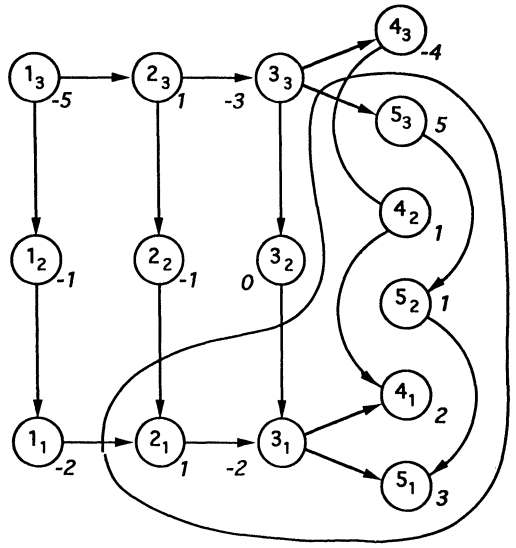


FIG. 2. *Maximal closure formulation of an approximation problem.*

Let $k_{\mathcal{P}}(x)$ be defined for any $x \in N$ as the index of $f_{\mathcal{P}}(x)$ in this linear order, where $f_{\mathcal{P}}$ is again the function defined by (3.4). In other words, $f_{\mathcal{P}}(x) = \nu_{k_{\mathcal{P}}(x)}$ for every $x \in N$.

Let us define a directed graph, $\hat{G} = (\hat{N}, \hat{A})$ associated with $G$ and $(V, \prec)$, as follows. The graph $\hat{G}$ contains $v|N|$ nodes, and each node $x \in N$ has $v$ "relatives" in $\hat{N}$, denoted by $x_1, x_2, \ldots, x_v$. We include the arcs $(x_i, y_i)$ for $(x, y) \in A$ and for indices $i$ with $\nu_{i-1} \prec \nu_i$ in the partially ordered set $(V, \prec)$. We also add the arcs $(x_i, x_{i-1})$ to $\hat{A}$ for every $x \in N$ and $i = 2, \ldots, v$. In other words, we place $v$ copies of the node set $N$ on top of each other and we include all the arcs of $G$ in the $i$th level, if $\nu_{i-1}$ and $\nu_i$ are comparable; otherwise the $i$th level is an empty graph. In addition to these, we include a directed arc from each vertex to its lower relative, too. The graph $\hat{G}$ associated with the graph of Fig. 1 appears in Fig. 2.

We define real weights $\hat{w}_{x_i}$ associated with every vertex $x_i \in \hat{N}$ as follows:

$$\hat{w}_{x_i} = \left( \sum_{\beta \preceq \nu_i} w_{x,\beta} - \sum_{\beta \preceq \nu_{i-1}} w_{x,\beta} \right) \quad \text{for } i = 1, \ldots, v, \text{ and } x \in N. \tag{5.2}$$

The weights $\hat{w}_{x_i}$ for our example also appear in Fig. 2.

THEOREM 3. *There is a one-to-one correspondence between the V-partitions of $G$ and the terminal sets of $\hat{G}$. Furthermore, if $\mathcal{P} = \{P_\alpha | \alpha \in V\}$ is a V-partition of $G$, and if $\hat{C}$ is the corresponding terminal set of $\hat{G}$, then*

$$w(\mathcal{P}) = \sum_{x \in N} w_{x,\nu_0} + \hat{w}\left(\hat{C}\right), \tag{5.3}$$

*where $\hat{w}$ is the vector defined by* (5.2).

*Since $\sum_{x \in N} w_{x,\nu_0}$ is a constant, it follows that finding a maximal V-partition in $G$ is equivalent with finding a maximum-weight terminal set in $\hat{G}$.*

*Proof.* Let us consider first a $V$-partition of $G, \mathcal{P} = \{P_\alpha | \alpha \in V\}$, and define a subset of the nodes of $\hat{G}$ as follows:

$$\hat{C}_{\mathcal{P}} = \bigcup_{i=1}^{v} \left\{ x_i \,\middle|\, x \in \bigcup_{j=i}^{v} P_{\nu_j} \right\}. \tag{5.4}$$

We show that $\hat{C}_{\mathcal{P}}$ is a terminal set in $\hat{G}$.

For $x_i \in \hat{C}_{\mathcal{P}}$, it follows that $x_{i-1} \in \hat{C}_{\mathcal{P}}$ by definition (5.4). Let us consider an arc $(x_i, y_i) \in \hat{A}$ of $\hat{G}$, where $x_i \in \hat{C}_{\mathcal{P}}$. By the definition of $\hat{G}$, $(x_i, y_i) \in \hat{A}$ only if $\nu_{i-1} \prec \nu_i$. Therefore it follows from (5.1) and (3.1) that $\bigcup_{j=i}^{v} P_{\nu_j}$ is a terminal set of $G$. By definition, $x_i \in \hat{C}_{\mathcal{P}}$ implies that $x \in P_{\nu_j}$ for some $j \geq i$, i.e., that $x \in \bigcup_{j=i}^{v} P_{\nu_j}$. From $(x_i, y_i) \in \hat{A}$, it also follows that $(x, y) \in A$ and therefore $y \in \bigcup_{j=i}^{v} P_{\nu_j}$, thus implying that $y_i \in \hat{C}_{\mathcal{P}}$.

Conversely, let us consider a terminal set $\hat{C}$ of $\hat{G}$. Let $P_{\nu_0} = \{x | x_1 \notin \hat{C}\}$, let $P_{\nu_i} = \{x | x_i \in \hat{C} \text{ but } x_{i+1} \notin \hat{C}\}$, for $i = 1, 2, \ldots, v-1$, and let $P_{\nu_v} = \{x | x_v \in \hat{C}\}$. We show that $\mathcal{P} = \{P_{\nu_i} | i = 0, 1, \ldots, v\}$ is a $V$-partition of $G$ such that $\hat{C} \equiv \hat{C}_{\mathcal{P}}$.

Since $\hat{C}$ is a terminal set of $\hat{G}$, it follows that, for every $x \in N$, either $x_1 \notin \hat{C}$, or $x_v \in \hat{C}$, or there is a unique $0 \leq i < v$ such that $x_i \in \hat{C}$ but $x_{i+1} \notin \hat{C}$. Therefore $\mathcal{P}$ is a partition of $N$.

Let us now consider an arbitrary arc $(x, y) \in A$ of $G$ for which $x \in P_{\nu_i}$ and $y \in P_{\nu_j}$, and let us assume that $\nu_j \prec \nu_i$. Then, on the one hand, $x_i \in \hat{C}$, but $x_{j+1} \notin \hat{C}$, by the above definition of $\mathcal{P}$. On the other hand, by property (C1), there must be an index $k$ such that $j \leq k, k+1 \leq i$, and $\nu_k \prec \nu_{k+1}$. Then the arc $(x_{k+1}, y_{k+1})$ belongs to the graph $\hat{G}$, and thus the vertices $x_i, x_{i-1}, \ldots, x_{k+1}, y_{k+1}, y_k, \ldots, y_{j+1}$ form a directed path from $x_i$ to $x_{j+1}$ in $\hat{G}$, which starts in the terminal set $\hat{C}$ and ends outside of it. This contradiction shows that $\nu_j \not\prec \nu_i$, i.e., that $k_{\mathcal{P}}(y) \not\prec k_{\mathcal{P}}(x)$ for any arc $(x, y) \in A$ of $G$, and thus it proves that $\mathcal{P}$ is a $V$-partition of $G$.

Finally, (5.3) follows easily by (5.2) and (5.4):

$$\begin{aligned} \hat{w}\left(\hat{C}\right) &= \sum_{x_i \in \hat{N}} \hat{w}_{x_i} \\ &= \sum_{x \in N} \sum_{i=1}^{k_P(x)} \hat{w}_{x_i} \\ &= \sum_{x \in N} \left( \sum_{\alpha \preceq f_{\mathcal{P}}(x)} w_{x,\alpha} - w_{x,\nu_0} \right) \\ &= w(\mathcal{P}) - \sum_{x \in N} w_{x,\nu_0}. \qquad \square \end{aligned} \tag{5.5}$$

The maximal closure for Fig. 2 is encircled in the figure. Note that this solution corresponds to the function $f_1$ defined in Table 1. Thus $f_1$ is a best approximation to the data.

We conclude by indicating in Table 2 the possible values of $f_1(x)$ for both observed and unobserved values of $x$. The observed values are in boldface. Note, for instance, that $f_1(N, U)$ cannot be $d$, since $(N, U) \preceq (I, U)$, $f_1(I, U) = u$, and $d \succ u$. In this example, the value of $f_1(x)$ is completely determined for only one unobserved $x$, namely, $x = (P, P)$.

TABLE 2
*Function values consistent with an optimal approximation.*

| $x$ | $f_1(x)$ |
|---|---|
| $(\mathbf{N},\mathbf{N})$ | $\mathbf{h}$ |
| $(U,N)$ | $h,u,p$ |
| $(N,U)$ | $h,u,p$ |
| $(\mathbf{I},\mathbf{N})$ | $\mathbf{u}$ |
| $(N,I)$ | $h,u,p,d$ |
| $(U,U)$ | $h,u,p$ |
| $(\mathbf{I},\mathbf{U})$ | $\mathbf{u}$ |
| $(P,N)$ | $u,p,d$ |
| $(U,I)$ | $h,u,p,d$ |
| $(I,I)$ | $u,p,d$ |
| $(N,P)$ | $h,u,p,d$ |
| $(\mathbf{P},\mathbf{U})$ | $\mathbf{p}$ |
| $(P,I)$ | $u,p,d$ |
| $(U,P)$ | $h,u,p,d$ |
| $(\mathbf{I},\mathbf{P})$ | $\mathbf{d}$ |
| $(P,P)$ | $d$ |

**Acknowledgments.** We thank the referees for the many helpful comments and suggestions made on an earlier version of this paper. We thank Dr. Paul B. Kantor of the School of Communication, Information and Library Studies, Rutgers University for kindly providing us with the example on carcinogenecity of chemicals, and Dr. Alex Kogan of Rutgers Center for Operations Research, Rutgers University for the helpful remarks and questions on partial ordered sets.

## REFERENCES

[1] M. L. BALINSKI, *On a selection problem,* Management Sci., 17 (1970), pp. 230–231.

[2] R. E. BARLOW, D. J. BARTHOLOMEW, D. J. BREMNER, AND H. D. BRUNK, *Statistical Inference under Order Restrictions,* John Wiley, New York, 1972.

[3] M. J. BEST AND N. CHAKRAVARTI, *Active set algorithms for isotonic regression; A unifying framework,* Math. Programming, 47 (1990), pp. 425–439.

[4] E. BOROS, P. L. HAMMER, AND J. N. HOOKER, *Boolean regression,* Research Report RRR 19-91, RUTCOR, Rutgers University, New Brunswick, NJ, 1991.

[5] E. BOROS AND R. SHAMIR, *Convex Separable Minimization over Partial Order Constraints,* Research Report RRR 27-91, RUTCOR, Rutgers University, New Brunswick, NJ, 1991.

[6] C. CARTER AND J. CATLETT, *Assessing credit card applications using machine learning,* IEEE Expert, Fall 1987, pp. 71–79.

[7] Y. CRAMA, P. L. HAMMER, AND T. IBARAKI, *Cause-effect relationships and partially defined Boolean functions,* Ann. Oper. Res., 16 (1988), pp. 299–325.

[8] B. FAALAND, K. KIM, AND T. SCHMITT, *A new algorithm for computing the maximal closure of a graph,* Management Sci., 36 (1990), pp. 315–331.

[9] P. C. FISHBURN, *Interval Orders and Interval Graphs,* John Wiley, New York, 1985.

[10] P. L. HAMMER AND S. RUDEANU, *Boolean Methods in Operations Research and Related Areas,* Springer-Verlag, Berlin, Heidelberg, New York, 1968.

[11] P. HANSEN, *Quelques Approches de la Programmation Non Linéaire en variables* 0-1, Conference on Mathematical Programming, Bruxelles, 1974.

[12] R. F. HAUCK AND J. M. MALONE, *Optimal Open Pit Mine Contours with Optimal Depletion Scheduling: Theory and Methodology,* U.S. Steel Corp., 1969.

[13] J. W. MAMER AND S. A. SMITH, *Optimizing field repair kits based on job completion rate,* Management Sci., 28 (1982), pp. 1328–1333.

[14] W. L. MAXWELL AND J. A. MUCKSTADT, *Establishing consistent and realistic reorder intervals in production-distribution systems,* Oper. Res., 33 (1985), pp. 1316–1341.

[15] J. D. MURCHLAND, *Rhys's combinatorial station selection problem,* London Graduate School of Business Studies, Transport Network Theory Unit, Report LBS-TNT-68, London, 1968.

[16] J.-C. PICARD, *Maximal closure of a graph and applications to combinatorial problems,* Management Sci., 22 (1976), pp. 1268–1272.

[17] J.-C. PICARD AND M. QUEYRANNE, *Selected applications of minimum cuts in networks,* INFOR, 20 (1982), pp. 394–422.

[18] ———, *Integer Minimization of a Separable Convex Function Subject to Variable Upper Bound Constraints,* Research Report, University of British Columbia, Vancouver, BC, April 1985.

[19] R. ROUNDY, *A 98% effective lot-sizing rule for a multi-product, multi-stage production/inventory system,* Math. Oper. Res., 11 (1986), pp. 699–727.

[20] J. M. W. RHYS, *A selection problem of shared fixed costs and network flows,* Management Sci., 17 (1970), pp. 200–207.

[21] M. J. SHAW AND J. A. GENTRY, *Inductive learning for risk classification,* IEEE Expert, February 1990, pp. 47–53.

[22] K. A. SPACKMAN, *Learning categorical decision criteria in biomedical domains,* in Proc. of the Fifth International Conference on Machine Learning, 1988, pp. 36–46.

[23] H. M. WEINGARTNER, *Capital budgeting of interrelated projects: Survey and synthesis,* Management Sci., 12 (1966), pp. 485–516.