ISOTONIC OPTIMIZATION IN TARIFF CONSTRUCTION

BY

W. S. JEWELL

ABSTRACT

The problem of the best fit to set ideal values under general inequality order restrictions is examined for asymmetric, quadratic, absolute, and Chebyshev norms. Special solution procedures are given in terms of network flow algorithms over a network associated with the given isotonic order relations, and the nature of the optimal solutions is characterized for the different norms.

The model is formulated in terms of finding an optimal insurance rate structure over given risk classes for which a desired pattern of tariffs can be specified. The suitability of different norms is considered in the context of corporate profitability, and the relationship to a simple rate relativities model is described.

Keywords: Isotonic Regression, Linear Programming, Quadratic Programming Network Flows Approximation Theory Insurance Ratemaking

1. INTRODUCTION

There are many different problems associated with the determination of rate structures in the insurance industry: the search for patterns in the data; development of underlying causal factors and analysis of their importance; subdivision of a group of existing risk classes into new tariff categories; fitting of experience data to standard formulae or rating bureau schedules; consideration of reserve liquidity, competitive, or legislative factors, and so on.

In this paper we shall examine the problem of determining the best set of premium rates over a finite set of premium classes, where the desired structure of tariffs is not given by a precise causal model, but can be specified in terms of *inequality constraints between tariffs in different classes* which reflect desirable "patterns" or "profiles". This approach allows the decision-maker a great deal of flexibility in specifying the acceptable set of solutions, and to some extent diminishes the need to determine causal factors, since rate structures seem more often to reflect "reasonable" relationships, parallels with existing rate patterns, or competitive factors, than they do some underlying physical law.

The simplest such rate relationship is a partial order between the tariffs; we shall see that this leads to a model in which the feasible solutions can be described in terms of a network optimization problem, similar to a critical path schedule. The optimal solution, or "best fit", of course depends upon the norm chosen, and the ideal solution without structure constraints; we shall examine the appropriateness of three major objective forms in a later section.

The advantage of network formulation for this problem lies in the fact that the structure constraints can be easily visualized, and modified if necessary. For small problems, the computations can be easily carried out "on the network", and are much more revealing than table or tableau formats; for larger problems, efficient computer codes of the network flow optimization type are available.

The mathematical ideas on which this approach is based are not new, having appeared in the literature with the names *isotonic regression* [2], *isotonic optimization* [23], or *majorized network flows* [24]. However the original papers are somewhat obscure, and consider more general and abstract formulations than are needed here. We shall emphasize the algorithmic approach, along the lines of previous programming approaches to approximation theory [8], [18], [14], [10], [25], [26], [22], [16], [19], [5], [4].

2. The Model

Suppose we wish to determine a *premium structure* $y = \{y_1, y_2, \ldots, y_n\}$ for n well-defined risk classes. Assume that from some other model (of profitability, competitive factors, group experience rating, etc.), a set of *ideal premium rates*, $f = \{f_1, f_2, \ldots, f_n\}$, has been determined, and a norm, $E(y) = \| y - f \|$, has been specified to reflect the *undesirability, cost*, or *error* in picking y different from f.

Three objective functions will be considered in the sequel:

$$E(y) = (1/2) \sum_{i=1}^{n} w_i(y_i - f_i)^2 \tag{1}$$

$$E(y) = \sum_{i=1}^{n} w_i \mid y_i - f_i \mid \tag{2}$$

$$E(y) = \max_{i=1,2,\ldots n} w_i \mid y_i - f_i \mid \tag{3}$$

where the $\{w_i\}$ are given positive weights associated with each class. The first is the classical *regression, least-squares* or L_2 norm; the second is the *weighted absolute* or L_1 norm; the last is called variously the *weighted Chebyshev, uniform,* or L_∞ norm [19]. Actually, we shall allow unsymmetric generalizations of these norms [10]; motivations will be discussed in a later section.

3. Isotonic Constraints and the Associated Graph

The concept of structure or pattern implies a certain relationship between premiums in different classes; this is why the ideal solution is not $y = f$. In some situations (such as the determination of classification relativities [13], [1]), one may postulate a certain mathematical form of relationship, and then approximate that form through adjustment of parameters.

In our model, we assume only that a certain *natural ordering* between tariffs is specified by external considerations, for instance that

$$y_j \geq y_i \tag{4}$$

between two classes i and j. An example of this might be in automobile insurance, where one reasonably expects that the categories: "for pleasure only", "less than 10 miles to place of work", "greater than 10 miles to place of work", and "used for business purposes", reflect increased hazard, and therefore should have monotonic rates. Or perhaps competitive factors influence the relationship between different premiums, or there is a concern that an unnatural pattern will induce "moral hazard" on the part of the policyholder.

Of course, if the ideal rates also reflect this ordering, then there is no problem; $y = f$ will be the optimum choice with any norm. However, the $\{f_i\}$ may be determined by small sample data, or by profitability considerations and hence may be contradictory to (4); therein lies the problem of determining the best tariff pattern within feasible structures.

If (4) holds for $j = i + 1$ ($i = 1, 2, \ldots, n - 1$), after possible relabelling, then we speak of a *complete ordering* of the risk classes. This can be visualized as the sequence of directed arcs, called a *path,* shown in Figure 1a; each of the $n - 1$ arcs represents an inequality constraint (4) between two of the n tariffs represented by the nodes.

In the more general case, (4) will hold only between certain given pairs of indices, and we speak of a *partial ordering* between risk classes. The relationships (4) can be represented as a directed graph $G = \{n; A\}$, with the set of nodes $n = \{1, 2, \ldots, n\}$ representing the

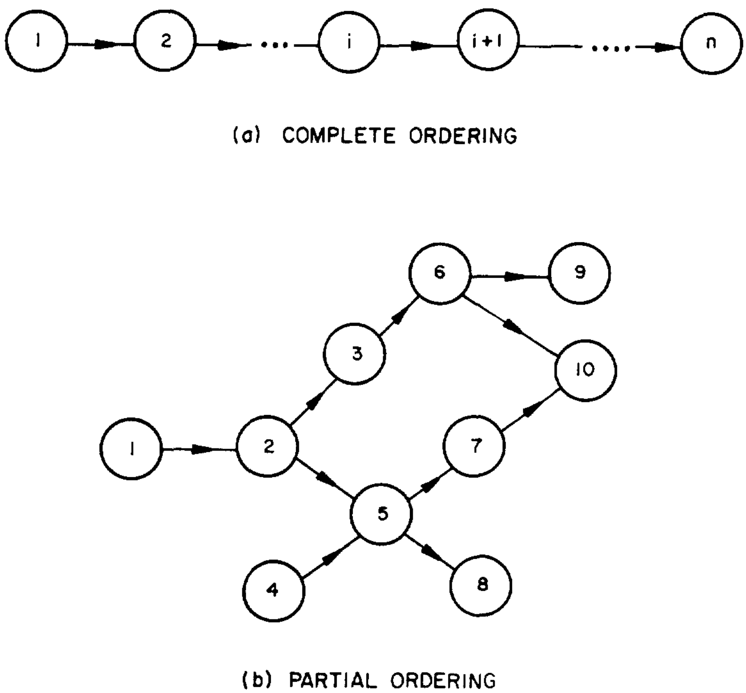


Fig. 1. Graphs Associated With Isotonic Ordering.

risk classes, and the set of directed arcs A representing each ordered set of indices (i, j) in an isotonic relationship (4).

We will assume that the graph G is *connected*; otherwise there is no structural relationship between certain sets of classes, and the problem falls apart into two or more separate ones. Also, the fact that we have a partial order means there are no *directed cycles* (loops of arcs all in the same orientation) within G; (4) would then imply that all the y_i were equal for classes corresponding to nodes in the cycle, and in more general models below might lead to contradictions. It is also well-known that for a partial order, we can arrange the indexing so that $i < j$ across every arc in A.

Figure 1b shows the graph of the partial order:

$$y_1 \leq y_2 \leq y_3 \leq y_6 \leq y_9; \quad y_2 \leq y_5 \leq y_7 \leq y_{10};$$
$$y_4 \leq y_5 \leq y_8; \quad y_{10} \geq y_6.$$

Even this G may not indicate the generality of the model, as it is planar. In a realistic problem, one may have many more arcs, with complicated "connectivity".

It turns out to be possible to generalize the structure constraints somewhat, and stay within the network formulation. Accordingly we shall allow more general forms of (4)

$$y_j - y_i \geq R_{ij}, \quad (i, j) \varepsilon A \tag{5}$$

and/or

$$y_j - y_i \leq S_{ij} \tag{6}$$

for given constants $R_{ij} \leq S_{ij}$. (4) corresponds to $R_{ij} = 0$, $S_{ij} = \infty$; setting $R_{ij} = S_{ij}$ clearly specifies an exact tariff differential.

Finally, we admit absolute bounds $A_i \leq B_i$ on the individual tariffs:

$$A_i \leq y_i \leq B_i, \qquad i \varepsilon n \tag{7}$$

to reflect legislative, competitive, or profitability constraints. We assume $A_i \leq f_i \leq B_i$. Setting $A_i = B_i$ uniquely "anchors" a tariff.

4. Network Optimization

Instead of the tariffs $\{y_i\}$, we shall work with a set of unrestricted error variables

$$v_i = y_i - f_i, \quad i \varepsilon n \tag{8}$$
$$v_i \text{ unrestricted}$$

and, further, will split the error v_i into its positive and negative parts.

$$v_i = u_i^+ - u_i^-, \qquad i \varepsilon n$$
$$u_i^+ \geq 0 \quad u_i^- \geq 0, \tag{9}$$

so that assymmetric norms can be considered.

The *assymetric least-squares isotonic optimization (regression) problem*, corresponding to norm (1), (5), (6), (7), then becomes:

$$\operatorname{Min} E = (1/2) \sum_{i \varepsilon n} [w_i^+ (u_i^+)^2 + w_i^- (u_i^-)^2] \tag{10}$$

subject to (9) and:

$$R_{ij} + f_i - f_j \leq v_j \leq S_{ij} + f_i - f_j \qquad (i, j) \varepsilon A \tag{11}$$

$$A_i - f_i \leq v_i \leq B_i - f_i. \qquad i \varepsilon n \tag{12}$$

Here we have generalized the quadratic weights w_i to assymetric quadratic weights $\{w_i^+ ; w_i^-\}$.

The *assymetric absolute isotonic optimization problem*, corresponding to norm (2), is:

$$\operatorname{Min} E = \sum_{i \varepsilon n} [w_i^+ u_i^+ + w_i^- u_i^-] \tag{13}$$

subject to (9), (11), and (12).

Finally, the *assymmetric Chebyshev isotonic optimization* problem, corresponding to norm (3), can be expressed in terms of a new (nonnegative) variable e:

$$\operatorname{Min} E = e \tag{14}$$

$$\left.\begin{aligned} e - w_i^+ u_i^+ &\geq 0 \\ e - w_i^- u_i^- &\geq 0 \end{aligned}\right\} \quad i \varepsilon n \tag{15}$$

In principle, the mathematical problem is solved at this point. The least-squares optimization is a *quadratic programming problem* [20], [12], [15], [6], while the other two norms are *linear programs* [21], and any general-purpose computer code could be used for numerical solutions.

But there is independent interest in seeing what characteristics the special structure inequalities (11) bring to the solution, and how one may solve small problems by hand. The *mathematical dual programs* will turn out to have an independent network flow interpretation [11], and from duality theory we can better understand the economic price we pay for structural consistency.

The interpretation of general isotonic regression as a network problem has been made by Veinott [24] and others. However, anyone who has worked with critical path scheduling will recognize

(11) as scheduling constraints, and realize that the dual is a flow problem over the associated precedence network. Apparently no one has previously investigated the actual algorithmic features in this formulation.

5. Absolute Isotonic Optimization

Basic Network Model

We consider the absolute norm first, since it is most closely related to network flow theory, and its dual most easily interpreted. We concentrate first on expressing (13), (9), (11), (12) as the dual to an optimal capacitated flow problem [21], [11].

Henceforth, let

$$r_{ij} = R_{ij} - (f_j - f_i); \; r_{ji} = (f_j - f_i) - S_{ij}, \quad (i, j) \in A. \qquad (16)$$

(Note reversal of indices and signs.) If the ideal tariffs, f_i and f_j, satisfy the isotonic constraints (5) and (6) strictly, then the corresponding r_{ij} and r_{ji} are *negative*; if f_i and f_j violate either (5) or (6), this is a conflict of interest, and the corresponding r_{ij} (or r_{ji}, but not both) is *positive*.

Ignore, temporarily, the individual constraints (7). Define a *reference variable* $v_0 = 0$, and henceforth write:

$$u_i^+ = u_{i0} \geq 0; \; u_i^- = u_{0i} \geq 0 \qquad i \in n. \qquad (17)$$

The *absolute isotonic optimization problem* can then be written:

$$\text{Min } E = \sum_{i \in n} [0 \cdot v_i + w_i^+ u_{i0} + w_i^- u_{0i}]$$

$$\begin{array}{llll} v_j - v_i & \geq r_{ij} & (i, j) \in A & \\ -(v_j - v_i) & \geq r_{ji} & & (18) \\ v_i - v_0 \quad + u_{0i} & \geq 0 & & \\ v_0 - v_i + u_{i0} & \geq 0 & i \in n & \\ & v_0 = 0 & & \\ & v_i \text{ unrestricted} & & \\ & u_{i0}, u_{0i} \geq 0. & & \end{array}$$

(We have used the fact that not both v_{i0} and v_{0i} will be positive in a basic solution.)

The reader should have no trouble in verifying that the *mathematical dual* to (19) is the *optimal capacitated flow problem*:

$$\text{Max } D = \sum_{(i,j)} \sum_{\varepsilon A^0} r_{ij} x_{ij}$$
$$\sum_{i \varepsilon n^0} (x_{ij} - x_{ji}) = 0 \qquad j \,\varepsilon\, n^0 \tag{19}$$
$$0 \leq x_{ij} \leq m_{ij} \qquad (i, j) \,\varepsilon\, A^0$$

for an *enlarged network* $G^0 = \{n^0; A^0\}$, constructed from G in the following fashion:

(i) A node 0, corresponding to v_0, is added; $n^0 = n + \{0\}$;
(ii) Two new *feeder arcs*, $(i, 0)$ and $(0, i)$ are added for every $i \,\varepsilon\, n$;
(iii) For every $(i, j) \,\varepsilon\, A$, add the *reverse arc* (j, i) (if S_{ij} finite).

Thus, if there were a original arcs in A, there are $2(a + n)$ arcs in A^0. This network is said to be in *circulation form* because there are no external flow requirements. If there are any profits to be made, they must be from loop flows, passing over arcs with positive r_{ij}.

The *unit flow profits*, r_{ij}, are given by (16) for the original arcs and the reversed arcs; but $r_{i0} = r_{0i} = 0$ for the feeder arcs. Conversely, the *flow capacities*, m_{ij}, are:

$$\left.\begin{aligned} m_{i0} &= w_i^+ \\ m_{0i} &= w_i^- \end{aligned}\right\} \quad i \,\varepsilon\, n \tag{20}$$
$$m_{ij} = \infty \qquad i, j \,\varepsilon\, n$$

Figure 2 illustrates the conversion of a portion of the graph of Figure 1(b) to network flow form.

Nature of Dual Flow Solution.

Clearly, the optimal flow pattern $x^* = \{x_{ij}^*; (i, j) \,\varepsilon\, A^0\}$ always exists, since $x = 0$ is always feasible. It is always finite, since the capacities on the feeder arcs are positive (weights), and the condition $R_{ij} \leq S_{ij}$ guarantees $r_{ij} + r_{ji} \leq 0$, i.e. no profit can be gotten from a "whirlpool" between i and j.

Thus, to make D increase, flow must pass from node 0 through arc $(0, s)$ to some *starting node*, s, thence along a series of (forward or reverse) arcs, *the sum of whose r_{ij} must be positive*, to some *terminal node t*, and thence back to node 0 via $(t, 0)$. The increase in D being proportional to the sum of r_{ij} along this (s, t)-*path* and to the common flow ϕ around this *elementary loop*, it follows that we want to increase $\phi = x_{0s} = x_{t0} = \ldots$ as much as possible, until it is limited by Min (w_s^-, w_t^+).

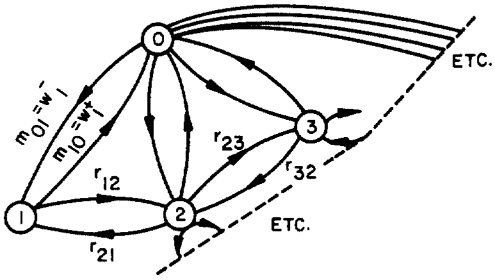


Fig. 2. Conversion of Graph G to Flow Network G^0.

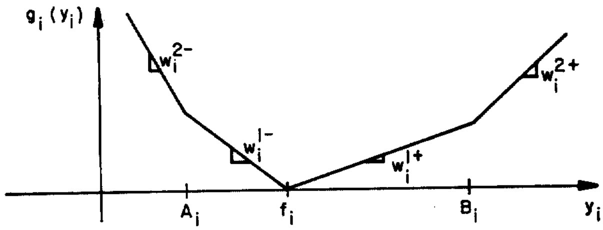


Fig. 3. Piecewise Linear Convex Norm.

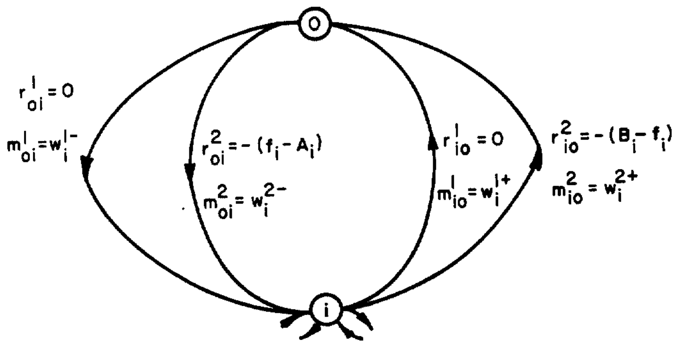


Fig. 4. Feeder Arc Configuration for Piecewise Linear Convex Norm.

Thus, the desirable flow paths include arcs for which $r_{ij} > 0$, i.e. *isotonic relationships unsatisfied* by the ideal f_i and f_j. In the simplest case, these *positive profit* arcs are nonadjacent, and have neighboring arcs which have large negative r_{ij} (well satisfied f_i, f_j relationships), so that combined flows are unprofitable; in other words, infeasible f_i, f_j relationships are *local* and *weak*. Then it is clear that the optimal x^* consists of the union of several disjoint loop flows over one regular and two feeder arcs.

We can understand the nature of the general solution better if we imagine that the *set P of all s, t paths*, P_{st}, for which

$$R(s, t) = \sum_{i,j \varepsilon p_{st}} r_{ij} > 0 \tag{21}$$

has been enumerated; this is a combinatorial task for large G^0 but quite reasonable for small problems. Let ϕ_{st} be the elementary loop flow along $(0, s) + p_{st} + (t, 0)$. Then it can be shown that (19) is equivalent to:

$$\begin{aligned} \text{Max } D = \sum_{p_{st} \varepsilon P} R_{st}\phi_{st} & \\ \sum_{t} \phi_{st} \leq w_s^- & \qquad s, t \;\varepsilon\; n \\ \sum_{s} \phi_{st} \leq w_t^+ & \qquad p_{st} \;\varepsilon\; P \\ \phi_{st} \geq 0. & \end{aligned} \tag{22}$$

In other words, in the general case (large, complex violations of structure by f), the profitable elementary loop flows are competing for entry and exit capacity. Rather larger sequences ("blocks", in [3]) of arcs, some possibly with negative r_{ij}, are included in order to use up adjacent capacity.

This approach can be made the basis of a good feasible starting solution x^0 to (19); one merely begins allocating loop flows in a myopic way, starting with an R_{st}-ordered list of paths. When capacity runs out, one skips the associated paths. Clearly $D(x^0) = D^0 < D^* = D(x^*)$.

Turning now to the relationship between the dual (19), and the primal (18) (A_i, B_i still neglected), we see that it is trivial to get a feasible set of potentials $\{v_i^0\}$. Assume the indices have been chosen so $i < j$ for all $(i, j) \;\varepsilon\; A$.

(i) Set the v_i^0 of lowest index equal to zero;

(ii) Go through the indices in increasing order, and for index j

$$v_j^0 = \underset{(i,\,j)\varepsilon A^0}{\operatorname{Max}_i} \{v_i^0 + R_{ij} + f_i - f_j\}$$

(iii) Repeat (ii) until all v^0 defined.

The v_i^0 can then be adjusted on an *ad hoc* basis to reduce the value of $E\{v^0\} = E^0$.

Network algorithms can be based on starting with either a feasible primal, v^0, with a feasible dual x^0, or with both. The advantage to having both is that the optimal value of the total norm can be bounded via the duality theorem [21]:

$$D^0 \leq D^* = E^* \leq E^0: \tag{23}$$

then, further improvement to v^0 may be judged unnecessary. Further descriptions of algorithms can be found for example in references [21] and [11].

There are important relationships between *optimal flows* x^* and *optimal potentials* v^* given by the *complementary slackness principle* [21]:

(i) If $x_{0i}^* < w_i^-$ and $x_{i0}^* < w_i^+$, then $v_i^* = 0$ $(y_i^* = f_i)$:

(ii) If $v_i^* > 0$ $(y_i^* > f_i)$, then $x_{i0}^* = w_i^+$, or if $v_i^* < 0$ $(y_i^* < f_i)$, then $x_{0i}^* = w_i^-$;

(iii) If $x_{ij}^* > 0$ $[x_{ji}^* > 0]$ for $(i, j) \varepsilon A$, then $v_j^* - v_i^* = r_{ij}$; $y_j^* - y_i^* = R_{ij}$ $[r_{ji}$ or S_{ij}, respectively].

(iv) If $v_j^* - v_i^* > r_{ij}$, $y_j^* - y_i^* > R_{ij}$, then $x_{ij}^* = 0$; similarly for $[r_{ji}, s_{ij}, x_{ji}^*]$.

Finally, we note the sensitivity analysis results, valid in general:

(i) If $y_i^* = f_i$, x_{0i}^* $[x_{i0}^*]$ is the level to which $w_i^-[w_i^+]$ must decrease before y_i^* becomes $< f_i$ $[> f_i]$;

(ii) $x_{ij}^* > 0$ $[x_{ji}^* > 0]$ for $(i, j) \varepsilon A$ is the marginal rate at which $D^* = E^*$ will increase, if R_{ij} is increased [S_{ij} is decreased].

These are valuable in re-examining the formulation, once the first solution has been obtained.

Nature of the Optimal Solution

The nature of the optimal solution to (29) is well-known [21]. In addition to the complementary relations between primal and dual variables, a key concept is that of the *optimal basis* for the optimal extremal solution. For a network of A arcs and N nodes, the basis is a *spanning tree* of $N - 1$ arcs—a connected configuration with no loops. On this tree, the $[v_i^*]$ can be calculated uniquely from tight isotonic relationships $v_j^* - v_i^* = r_{ij}$, and $v_0 \equiv 0$, and the x_{ij}^* can have any feasible value $0 \leq x_{ij}^* \leq m_{ij}$. The remaining $A - N + 1$ arcs are called the *co-tree*; for these arcs, either $x_{ij}^* = 0$ $(v_j^* - v_i^* \leq r_{ij})$, or $x_{ij}^* = m_{ij}$ $(v_j^* - v_i^* \geq r_{ij})$. (If the problem parameters are perturbed slightly to eliminate tieing solutions, the above inequalities can be changed to strict inequalities.) Clearly, not both (i, j) and (j, i) can be tree arcs.

In terms of our original problem, this means that in the optimal solution the n classes represented by the nodes n are *partitioned* into some number $1 \leq r \leq n$ of *solution blocks* $n = \{B_1, B_2, \ldots B_r\}$, each block containing a variable number of nodes. A *singlet block* of only one node k means that either $(0, k)$ or $(k, 0)$ can be placed in the optimal tree, but no other (i, k) or (k, i); $x_{0k}^* = x_{k0}^* = 0$, and $v_k^* = 0$ $(y_k^* = f_k)$. Conversely, if there is only one solution block B_1, then a tree of $n - 1$ relationships (i, j) must be tight $(v_j^* - v_i^* = r_{ij})$ among the nodes n, and this is augmented by one feeder arc $(0, s)$ [or $(t, 0)$] to make a spanning tree for G^0; v_s^* [or v_t^*] $= 0$, and this defines the potentials uniquely over the block. Dual flows must be zero or saturated on all other feeder arcs, and zero on co-tree arcs in B_1.

In the general case of several blocks of differing sizes:

(i) If block k contains N_k nodes, there is a *shrub* (small tree) of $N_k - 1$ arcs within the block, over which the isotonic relations are tight $(v_j^* - v_i^* = r_{ij};\ i, j \in B_k)$;

(ii) Each block k has a reference node, s_k[or t_k], for which $v_{s_k}^*$ $[v_{t_k}^*] = 0$, which then determines the potentials uniquely;

(iii) The union of the r shrubs and the feeder arcs $(0, s_k)$ [or $(t_k, 0)$] gives the n arcs corresponding to the optimal tree for G^0.

(iv) Cotree arcs within blocks, or between blocks (but not feeder arcs) must have $x_{ij}^* = 0$, and $v_j^* - v_i^* \geq r_{ij}$;

(v) Feeder arcs not in the tree must have $x^*_{0i} = w^-_i$ if $v^*_i < 0$, or $x^*_{i0} = w^*_i$ if $v^*_i > 0$.

Other characteristics which come out upon closer examination are:

(a) Some arcs with $r_{ij} < 0$ may have, but not every arc with $r_{ij} > 0$ need have, $x^*_{ij} > 0$;

(b) Nonunique values of $\{v^*_i\}$ and $\{y^*_i\}$ may result if opposing weights $[w^\pm_i]$ are balanced within a block;

(c) Nonunique values of $\{x^*_{ij}\}$ may result if the constants $\{r_{ij}\}$ are conservative around a loop, thus making $v^*_j - v^*_i = r_{ij}$ on an interblock cotree arc.

Individual Constraints, Piecewise Linear Norms

We now return to the problem of individual class constraints, A_i and B_i, in a roundabout manner that will provide additional generality. Suppose that the norm (2) is replaced by the sum of n piecewise-linear convex norms, for each class, as follows:

$$E(y) = \sum_{i-1}^{n} g_i(y_i)$$

$$g_i(y_i) = \begin{cases} w^{1-}_i(f_i - A_i) + w^{2-}_i(A - y_i) & y_i \leq A_i \\ w^{1-}_i(f_i - y_i) & A_i \leq y_i \leq f_i \\ w^{1+}_i(y_i - f_i) & f_i \leq y_i \leq B_i \\ w^{1+}_i(B_i - f_i) + w^{2+}_i(y_i - B_i) & y_i \geq B_i \end{cases} \tag{24}$$

A typical $g_i(y_i)$ is shown in Figure 3. To be convex, $w^{2+}_i \geq w^{1+}_i \geq 0$; $w^{2-}_i \geq w^{1-}_i \geq 0$.

A straightforward application of duality shows that in this case each feeder arc in G^0 is replaced by two feeder arcs; $(i, 0)$ becomes $(i, 0)^1$ and $(i, 0)^2$, etc., the new parameters are:

$$r^1_{0i} = 0;\ m^1_{0i} = w^{1-}_i;\ r^1_{i0} = 0;\ m^1_{i0} = w^{1+}_i;$$
$$r^2_{0i} = -(f_i - A_i);\ m^2_{0i} = w^{2-}_i;\ r^2_{i0} = -(B_i - f_i);\ m^2_{i0} = w^{2+}_i. \tag{25}$$

In this way, the problem remains a network flow problem, with a new configuration as shown in Figure 4.

Loosely speaking, the new arcs permit dual flow in the feeder arcs above the values w^{1+}_i and w^{1-}_i, by imposing a positive unit cost, $+ (f_i - A_i)$ or $+ (B_i - f_i)$. The interested reader can easily work out the other details of duality, or extend the curve to several linear segments.

Finally, to impose strict limits of the type $A_i \leq y_i \leq B_i$, we let the $\{w_i^{2+}\}$ and $\{w_i^{2-}\}$ increase without limit. If there is any elementary loop using path P_{st} for which $r_{0s}^2 + R_{st} + r_{t0}^2 > 0$, then the optimal dual flow will increase without limit, which means that the con straints A_s and B_t on classes s and t have rendered the original isotonic problem primally infeasible. Network computer codes would check this possibility automatically.

6. Chevyshev Isotonic Optimization Problem

As a preliminary to consideration of norm (14), suppose we have a two-variable Chevyshev isotonic problem $y_s \leq y_t$, but $r_{st} > 0$. Then clearly $y_s^* = y_t^*$, and from (15), we find, in the new notation:

$$-v_s^* = u_{0s}^* = \frac{\phi^*}{w_s^-} r_{st}; \quad v_t^* = u_{t0}^* = \frac{\phi^*}{w_t^+} r_{st}$$

where

$$\phi^* = \frac{w_s^- w_t^+}{w_s^- + w_t^+}; \quad E^* = e^* = \phi^* r_{st}. \tag{26}$$

Thus the basic property of this norm is that (at least) two opposing errors of class s and t are equalized.

If we take the dual of (14), (9), (11), (12), (15), we obtain an *optimal flow problem with variable capacities* on the feeder arcs:

$$\begin{aligned} \text{Max } D = &\sum_{(i,j)\,\varepsilon A^0} \sum r_{ij} x_{ij} \\ &\sum_{i \varepsilon n^0} (x_{ij} - x_{ji}) = 0 \qquad j \, \varepsilon \, n^0 \\ &0 \leq x_{0i} \leq w_i^- \cdot z_{0i} \geq 0 \quad i \, \varepsilon \, n \\ &0 \leq x_{0i} \leq w_i^+ \cdot z_{0i} \geq 0 \\ &\sum_{i \varepsilon n} (z_{0i} + z_{i0}) \leq 1. \end{aligned} \tag{27}$$

The last inequality is effectively an equality when $e^* > 0$, and provides for a weighted allocation of dual flow capacity only to those feeder arcs $(0, s)$ and $(t, 0)$ which correspond to equalized weighted errors in the sense of (26).

One can easily develop an algorithm to work directly with (14), (15) and (27) along the lines of other Chebyshev approximation algorithms [5], [10], [14], [17], [18], [22], [23], [25], and [26].

Loosely speaking, it would proceed as follows:

(i) Start with a feasible tariff structure $\{v_i^0\}$, say the optimal solution to the absolute isotonic problem with the same weights, and with $x^0 = z^0 = 0$;

(ii) Identify the current largest weighted error, e^0, say $w_s^-(v_{0s}^0)$ [or $w_t^+(u_{t0}^0)$] and allow the capacity on this arc to increase from zero;

(iii) Increase v_s^0 decrease v_t^0, and all potentials linked together ($v_j^0 - v_i^0 = r_{ij}$) in the same block, by an amount θ. The decrease $e^0 - \theta$ will be limited by:

   (a) an increase in weighted error of another node $t[s]$ in the same block to the same value;

   (b) the attainment of a tight constraint $v_j^0 - v_i^0 = r_{ij}$ at some extremity of the block with a new node k not in the block; or

   (c) the decrease of $e^0 - \theta$ to the fixed maximal weighted error of some node s' or t' in a *different* block;

   (Note that when a potential changes sign, a different weight must be used.)

(iv) In case (a), the optimal solution is attained, with the new node $t[s]$ being the exit [entry] point for dual flow. Equations (26) are satisfied, with ϕ^* being the optimal dual flow around the elementary loop, $D^* = E^*$, and z_{0s}^*, z_{t0}^* adjusted to suit.

(v) In case (b), the structure constraints merge node k (and other nodes in its block) into the current block. One continues as in Step (iii) with the enlarged block;

(vi) In case (c), weighted error equality is between two candidate nodes (s or t) and (s' or t') *not* in the same block, and further decreases of both errors is necessary to determine which block contains the limiting equality pair. A generalized Step (iii) is now performed except that θ and θ' must be in the ratio $w_{s'}^-\theta' = w_s^-\theta$, etc., to keep the weighted maximal errors in the different blocks identical. If further steps (iiic) are reached, one may be working on many different blocks at the same time. However, except in tieing cases (which can be easily perturbed), the optimal solution is still a balance between *some* s and *some* t in the same block.

In general, the Chebyshev norm is characterized by a great deal more freedom than the absolute norm, since only one block defines the two matched maximal weighted errors e^*, and other blocks are free to have their potential shifted up and down within the limits imposed by e^* and the slack between blocks. The advantage to starting with the optimal solution to the absolute norm problem with the same weights is that, when the optimal Chebyshev solution is obtained, a *particular* solution within the freedom just described is found which minimizes the absolute norm of classes outside the limiting block.

The classical approach described above suffices for small problems solved manually. For large problems, a more convenient formulation is gotten by *pricing in* the extra constraints in (27). After some simplification because of the nature of the optimal solution, we find a new *optimal flow problem with parametric costs*:

$$\begin{aligned} \operatorname{Max} F &= \sum_{(i,j)\,\varepsilon A^0} \sum r_{ij} x_{ij} \\ &\sum_{i\varepsilon n^0} (x_{ij} - x_{ji}) = 0 \qquad j \,\varepsilon\, n^0 \\ &x_{ij} \geq 0 \qquad\qquad (i, j)\,\varepsilon\, A^0 \end{aligned} \tag{28}$$

with new feeder arc unit profits:

$$r_{i0} = -\frac{\lambda}{w_i^-}\,;\; r_{0i} = -\frac{\lambda}{w_i^+}. \qquad i \,\varepsilon\, n \tag{29}$$

This corresponds to a restatement of the original problem as:

Find feasible $\{v_i\}$ such that

$$\begin{aligned} &v_j - v_i \geq r_{ij} \qquad (i, j)\,\varepsilon\, A^0 \\ &\frac{-\lambda}{w_i^-} \leq v_i \leq \frac{\lambda}{w_i^+}. \qquad i \,\varepsilon\, n \end{aligned} \tag{30}$$

Here λ is a Lagrange Multiplier which is increased slowly from zero to the first value λ^* at which (30) is feasible, at which point $\lambda^* = e^*$, and $F^* \rightarrow 0$.

In this formulation, (28) is an ordinary optimal profit flow problem, except that there are *parametric costs* (29) on the feeder arcs, which increase from zero, until $F^* = 0$. (Thus, these dual flows are not those of (27).) The most convenient way to solve the problem is then:

(i) Add, temporarily, the constraints (20), and solve the corresponding absolute norm problem for $\{v^0\}$ and $\{x^0\}$;
(ii) Decrease the unit profits on the feeder arcs according to (29) by increasing λ, and resolving (28), "extinguishing" the flows in the various blocks until x^0 just reaches zero.

There are network flow algorithms available to do this parametric variation directly, or it may be carried out through man-machine interaction with any computer code. A computational advantage is that one need not keep separate track of all the weighted errors, and that the final $\{v_i^*\}$ will be automatically adjusted to their final values, in the sense described above of the best absolute error particular solution within the optimal Chebyshev solution. The last (s, t) flow path to be extinguished gives the limiting pair of maximal weighted errors.

7. Isotonic Regression Problem

The symmetric least-squares norm (1) has been extensively studied in the statistical literature; see in particular Reference [2] and the bibliography therein.

We describe first the elegant algorithm for the *simple isotonic regression* problem (4) *over a complete order*, with no side constraints, due to Reid-Brunk-Grenander [2]:

(a) Compute cumulative weighted ideals and cumulative weights:

$$F_j = \sum_{i=1}^{j} w_i f_i; \quad W_j = \sum_{i=1}^{j} w_i \quad (j = 1, 2, \ldots, n). \tag{31}$$

(b) Plot the points $P_j = (W_j; F_j)$ in the Cartesian plane, connected by straight lines, and stretch a string, attached at $P_0 = (0, 0)$ and P_n, from below until it is taut. (This is the Newton-Puiseux polygon, or greatest convex minorant of the graph $\{P_j\}$.)

(c) If, for any j, the string is below P_j, then $y_j^* = y_{j+1}^*$, and so on, until a point $j + 1$, or $j + 2$, or ... k is reached where the string reaches a support P_k.

(d) This partitions the m classes into $1 \leq r \leq m$ *solution blocks*, $\{B_1; B_2 \ldots B_r\}$ over which the optimal solution is constant. By direct calculation:

$$y_j^* = \left(\sum_{i\varepsilon\beta_k} w_i f_i \Big/ \sum_{i\varepsilon\beta_k} w_i\right) \qquad j \,\varepsilon\, B_k \tag{32}$$

which is also the slope of the taut string from $j - 1$ to j.

(e) If the string follows the segment (P_{j-1}, P_j), then P_j is in a solution block of one member, $y_j^* = f_j$, and $y_j^* > y_{j-1}^*$.

Other "pooling" algorithms for this case are given in Reference [2].

This approach is easily extended to the case where the graph G of the partial order is an arborescence. The algorithms proposed for an arbitrary G are more complicated to understand as are the modifications imposed by constraints (6) and (7). (See the algorithms of Thompson, Kruskal, Alexander, and van Eeden in Chapters 1 and 2 of [2].) We shall follow instead the remark of Veinott [24] (in a more general context), that general isotonic regression problem "is a separable quadratic network flow problem for which special algorithms are available."

Suppose that the constraints are expressed as in (18), but with the unsymmetric quadratic norm:

$$\text{Min } E = 1/2 \sum_{i\varepsilon n} w_i^+ (u_{i0})^2 + w_i^- (u_{0i})^2. \tag{33}$$

Then the Dorn dual [3] becomes:

$$\begin{aligned} \text{Max } D = & \sum_{(i,j)} \sum_{\varepsilon A} r_{ij} x_{ij} - 1/2 \sum_{i\varepsilon n} (u_{i0} x_{i0} + u_{0i} x_{0i}) \\ & \sum_{i\varepsilon n^0} (x_{ij} - x_{ji}) = 0 \qquad j \,\varepsilon\, n^0 \\ & \qquad x_{ij} \geq 0 \qquad (i, j) \,\varepsilon A^0 \\ & \qquad x_{i0} \leq w_i^+ u_{i0} = m_{i0}(u_{i0}) \\ & \qquad x_{0i} \leq w_i^- u_{0i} = m_{0i}(u_{0i}) \end{aligned} \tag{34}$$

Notice the presence of the primal variables in the dual functional and in the (varying) capacities of the feeder arcs. Using complementary slackness and the fact that the weights are positive, we infer that in the optimal solution:

$$
\begin{aligned}
x_{i0}^* &= w_i^+ u_{i0}^* = w_i^+[\max v_i^*, 0] \\
x_{0i}^* &= w_i^- u_{0i}^* = w_i^- \max[-v_i^*, 0]
\end{aligned} \tag{35}
$$

and re-express the dual as:

$$
\begin{aligned}
\text{Max } D &= \sum_{(i,j)}\sum_{\varepsilon A} r_{ij}x_{ij} - 1/2 \sum_i \left[\frac{(x_{i0})^2}{w_i^+} + \frac{(x_{0i})^2}{w_i^-}\right] \\
&\sum_{i\varepsilon n^0} (x_{ij} - x_{ji}) = 0 \qquad j \varepsilon n^0 \\
&x_{ij} \geq 0 \qquad (i, j) \varepsilon A^0.
\end{aligned} \tag{36}
$$

which is essentially the Dennis dual [8]. In this form, we have an *optimal flow problem with quadratic costs in the feeder arcs.*

The structure of the optimal flow solution is as discussed earlier: the nodes are separated into solution blocks, which, together with their linking arcs, form *shrubs* of 1, 2, . . ., or up to n nodes, over which *tight* primal relationships (5) or (6) obtain. Thus, the *relative* values of the v_i^* and y_i^* within a solution block B_k are completely defined by a set of simple equality relations.

If one finds a certain relative solution $\{v_i^0; i \varepsilon B_k\}$ for a block $_k$, then by (35), the exogenous flows $\{x_{i0}; x_{0i}\}$ into and out of the block are completely defined; if these flows balance over B_k, then the current solution is optimal. (An isolated node k can then only have $v_k^* = x_{0k}^* = x_{k0}^* = 0$.) Contrarywise, an excess [deficit] of flow in B_k indicates that potentials v_i^0 should be raised [lowered] by an amount θ. The adjustment of exogenous flows at node i will be $\pm [\mp]\, w_i^{\pm}\theta$ by (35), and thus a choice of the correct θ to achieve Kirchoff's conservative law will also determine the correct *absolute value* of the tariffs in that solution block. The result will be the appropriate generalization of (32).

Thus, the key problem is that of separating the nodes into solution blocks, rather than finding the best regression within the block. For small problems, a manual solution can proceed as follows:

(1) Establish a tentative partition of n into blocks $\{B_1, B_2, \ldots, B_r\}$, by linking together nodes for which r_{ij} or $r_{ji} > 0$ with arcs of correct orientation. $v_i^0 = 0$ is thus initially feasible for inter-bloc arcs.

(2) Select a block B_k not previously examined, and pick a set of relative potentials satisfying $v_j^0 - v_i^0 = r_{ij}$ $(i, j \varepsilon B_k)$ and calculate exogenous flows from (35):

   (a) If the flows balance, the current solution is (locally) optimal; pick another block;
   (b) If there is excess [deficit] flow, all $v_i^0 (i \varepsilon B_k)$ must be increased [decreased] by an amount θ until the flows from (35) balance, or until;

      (i) If a v_i^0 changes sign, the computation should stop, and recommence with new asymmetric weight;
      (ii) If an arc flow x_{ij}^0 decreases to zero, the computation stops, and the block is split between i and j;
      (iii) If the final solution violates $v_j^0 - v_i^0 \geq r_{ij}$ for one or more arcs between blocks, then merge the current block with that block for which the discrepancy is greatest, and repeat (2) for the enlarged block j.

(3) The above process of merging and dissolving blocks, and floating potentials is repeated until:

   (a) Flows balance within blocks;
   (b) $v_j^0 - v_i^0 = r_{ij}$ within blocks;
   (c) $v_j^0 - v_i^0 \geq r_{ij}$ and $x_{ij}^0 = 0$ across blocks.

   These values are then optimal.

As a final check, the weighted mean-square error E^* should be computed and checked with $R^* = \Sigma \Sigma r_{ij} x_{ij}^*$; the result should be $E^* = 1/2 R^*$. This is related to well-known theorem about power dissipation in electrical engineering [8]!

If bounds (7) are added, this places flow capacities of

$$m_{i0} = w_i^+ (B_i - f_i); \quad m_{0i} = w_i^- (f_i - A_i) \tag{36}$$

on the feeder arcs, and complicates the procedure somewhat.

Unfortunately, there do not seem to be any quadratic network computer codes available for large problems. One possibility is to use a general quadratic linear programming code (see, e.g. [12], [15], [16]). However, because of the efficiency of ordinary linear network algorithms, my personal suggestion for a large problem would be to approximate the quadratic (or other convex) norms by a simple piecewise linear convex function, in a manner similar to Figures (3) and (4), over some reasonable ranges. Then, when the initial solution values $\{u^0_{i0}; u^0_{0i}\}$ were known, one could redefine current feeder arc parameters as:

$$\begin{aligned} m^0_{i0} &= w_i^+ u^0_{i0}; \quad r^0_{i0} = -1/2\, u^0_{i0};\\ m^0_{0i} &= w_i^- u^0_{0i}; \quad r^0_{0i} = -1/2\, u^0_{0i}. \end{aligned} \tag{37}$$

(thus effectively having only one feeder arc), and then re-iterate. In most problems, the solution should stabilize quickly into solution blocks; side calculations could then replace the within-block convergence to v_i^*. An out-of-kilter code [21], [11] would be ideal for this interaction.

8. Solution Examples

To illustrate these ideas, the isotonic ordering of Figure 1b and the ideal values and weights shown in Table I, were used to find optimal solutions for the three different assymetric norms, referred to below as $L_1^\pm$, $L_2^\pm$, and $L_\infty^\pm$. (A missing weight indicates it is not influential in the final solution.)

Figure 5 shows the ideal values f_i on each node, and the derived r_{ij}; the example is probably atypical, in that six out of ten isotonic relationships are violated. The optimal solution to the $L_1^\pm$ norm is shown in Figure 6a; v_i^* are shown on the nodes, $x_{ij}^* > 0$ (solution blocks) on solid arcs, and x_{0i}^* or $x_{i0}^* > 0$ as numbers on feeder arrows. Figure 6b show the $L_\infty^\pm$ with the v_i^* which also make $L_1^\pm$ as small as possible; more general ranges for y_4^* and y_7^* are possible, as shown in Table I. The last flow to be extinguished is that from node 1 to node 10 over the solid arcs; the final value of the LaGrange multiplier in (30) and the weighted errors at nodes 1 and 10 is $\lambda^* = e^* = 18.75$. The final solution for the $L_2^\pm$ norm, as shown in Figure 6b, also has two solution blocks, node 4 as a singlet and all others linked together; notice the complexity of the flow, and the uniqueness of v_4^* and v_7^*. The x_{ij}^* are not unique, however, since the r_{ij} were chosen to be conservative around the two paths from node 2 to node 10; this means that arc (5, 7) (shown dotted) is actually priced in, and up to 5 1/3 units of dual flow could be rerouted via the lower path.

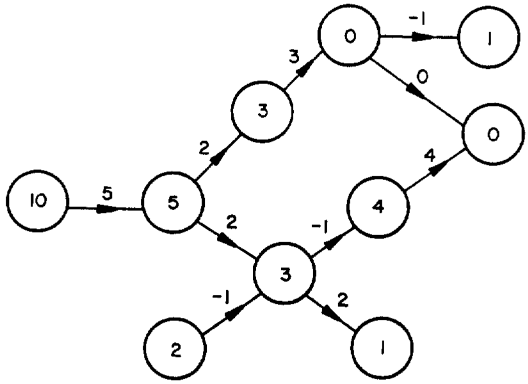
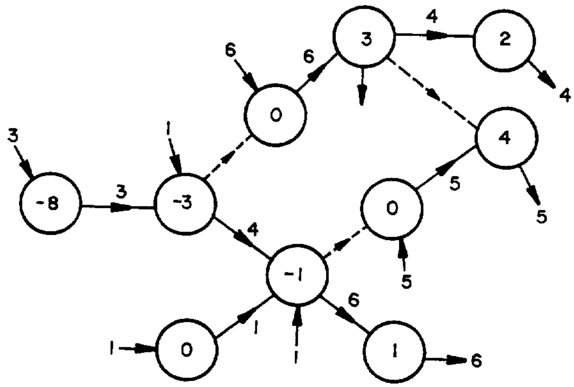
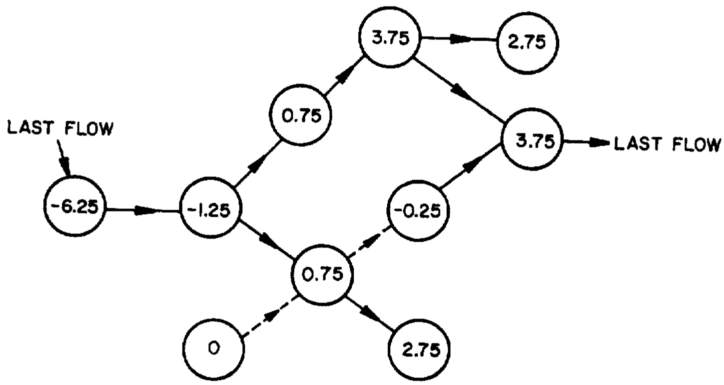


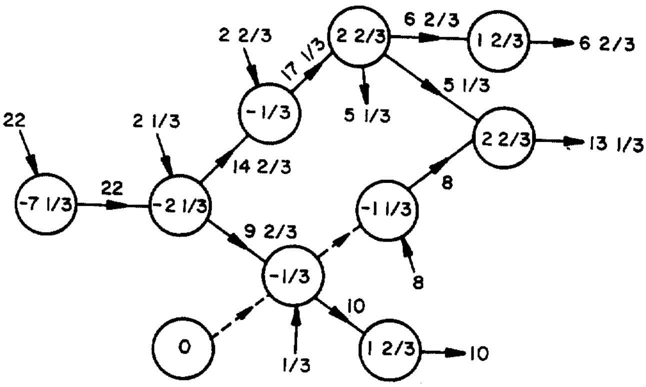
Fig. 5. Ideal Values f_i and r_{ij} for Example.



(a) $L_1^{\pm}$ NORM.



(b) $L_\infty^{\pm}$ NORM.



(c) $L_2^{\pm}$ NORM.

TABLE 1
Data and Solution for Example.

		1	2	3	4	5	6	7	8	9	10
Class i		1	2	3	4	5	6	7	8	9	10
Weights	w_i^+	–	–	8	5	4	2	–	6	4	5
	w_i^-	3	1	8	3	1	–	6	–	–	–
Ideal Values f_i		10	5	3	2	3	0	4	1	1	0
Optimal Values y_i^*	$L_1^{\pm}$ Norm	2	2	3	2	2	3	4	2	3	4
	$L_\infty^{\pm}$ Norm	←	3.75	→	(−4.25, 3.75)	←	3.75	→	(3.75, 4.125)	(3.75, 5.6875)	3.75
	$L_2^{\pm}$ Norm	←	2 2/3	→	2	←			2 2/3		→

TABLE 2
Relative Value of All Norms for Different Optimal Solutions

Solution Norm and Optimal Value		Relative Value of All Norms: E for $L_1^{\pm}$	E for $L_2^{\pm}$	E for $L_3^{\pm}$
$L_1^{\pm}$	$E^* = 68$	1.000	1.239	1.039
$L_2^{\pm}$	$E^* = 18\ 3/4$	1.280	1.000	1.173
$L_3^{\pm}$	$E^* = 128$	1.258	1.172	1.000

Although the different optimal norms are not, strictly speaking, comparable, Table II may be of interest in showing the relative trade-offs. (For the $L^{\pm}_{\infty}$ norm, the best $L^{\pm}_{1}$ solution was assumed.)

9. Nature of Solutions

Let us summarize briefly the nature of the optimal solutions for the three norms applied to the simple isotonic problem (4) with arbitrary G^0. In general (although not necessarily for a single problem):

(i) The $L^{\pm}_{\infty}$ norm gives the greatest freedom in the $\{y^*_i\}$, followed by the $L^{\pm}_{1}$ and $L^{\pm}_{2}$ solutions, which are almost always unique;

(ii) $L^{\pm}_{2}$ tends to have fewer and larger solution blocks than $L^{\pm}_{1}$ for the same problem;

(iii) In the constrained block, the $L^{\pm}_{\infty}$ solution is usually related to a harmonic mean (26) of the weights of two classes, and the values in other blocks can be conveniently adjusted to minimize the local value of $L^{\pm}_{1}$;

(iv) In the $L^{\pm}_{1}$ optimal solution, the "weight of evidence" within a block sets all y^*_i to the same reference ideal f_s or f_t;

(v) In the $L^{\pm}_{2}$ optimal solution, a complicated weighted consensus (32) is used for the values of $y^{\pm}_i$ within the same block.

Further characterizations are possible for special G^0, as in the complete ordering [2].

10. Models for Norms and Ideal Values

Let us now consider various models which might be appropriate in an insurance context for picking a norm and ideal tariffs for the different risk classes.

In the classical statistical formulation, one assumes a set of n_i independent measurements $\{x_{it},\ t = 1, 2, \ldots, n_i\}$ for each tariff class i, and minimizes the sum of squared errors of each measurement, weighted equally. This is equivalent to using the sample mean as ideal, $f_i = \bar{x}_i = (\sum_t x_{it}/n_i)$, and L_2 symmetric weights proportional to observations, $w_i = n_i$. Or, one may start with a normality assumption and known variances, σ^2_i, for each class, and obtain a maximum likelihood estimate over all observations, with $f_i = \bar{x}_i$ and $w_i = n_i/\sigma_i^2$.

Statistical arguments for the absolute norm [7] usually emphasize the difficulty with outliers, and the relative "freedom" of a norm $E = \Sigma\, n_i \mid y_i - f_i \mid$ (with f_i as the sample median or sample average), even though it has no direct statistical interpretation.

In my opinion, greater emphasis should be placed upon economic, rather than statistical models of rate-making. To illustrate this, let us first suppose that the *premium volume* in class i is N_i, and that it is relatively insensitive to the premium level y_i over the range of interest. Then the *net profit in class i* is:

$$P_i(y_i) = N_i(y_i - k_i) \tag{38}$$

where k_i is the (known) *expense-loaded fair premium.* Clearly $A_i = k_i$ in (7) if the company is unwilling to lose money on any tariff class; or, it could be set at $A_i = 0.95\, k_i$, etc. To maximize total profit as an objective, one would then set $f_i = B_i$ and use an $L_1^{\pm}$ norm, with only one weight $w_i^- = N_i$; clearly the result depends upon how greedy the company is (for one class within each solution block) in this perfectly inelastic market! Perhaps the B_i are fixed by the insurance commissioner, or some rate-of-return rule, or can be determined as a level at which the (high) tariffs become too visible to customers and competitors.

A more realistic model would assume, for instance, that the market was slightly elastic with price, say:

$$N_i(y_i) = N_i^k - \alpha_i(y_i - k_i), \tag{39}$$

where N_i^k is a reference volume at the breakeven level, and α_i a known elasticity coefficient. Then (38) above leads to a quadratic profit function, in which penalties will occur for variation about the point of maximal profit, $y_i = f_i$. In terms of the above:

$$f_i = k_i + (N_i^k/2\alpha_i) \tag{40}$$

and the maximal profit is:

$$P_i(f_i) = (N_i^k)^2/4\alpha_i. \tag{41}$$

The symmetric L_2 norm would be used with f_i given by (40) and

$$w_i = 2\alpha_i. \tag{42}$$

Of course, a linear approximation could be made to the above norm, as discussed previously, especially if f_i turned out to be outside the range (A_i, B_i) allowed by competitive or other factors. The piecewise linear norm is, strictly speaking, only exact in the above model when there are discontinuous steps in $N_i(y_i)$, as when we automatically lose a certain fraction of business when we go one cent over a competitor's rates.

It seems quite difficult to justify the $L^{\pm}_{\infty}$ norm in the insurance context, unless one assumes that competitors, customers, or the insurance commissioner are looking for rates or profits that are excessively "out of line". Perhaps the norm would be appropriate if the classes were poorly defined or monitored, or the insuree were self-rated, and there was a certain moral hazard of shifting to nearby categories. This would seem, however, to be better handled by using (6) to avoid large neighboring class discontinuities.

Finally, the complete rate-making process involves a complex series of interactions between different parties of interest, and it may be that the actuary-operations researcher will prefer one norm over another solely in terms of the nature of the solutions it gives, its ease of computation, or its defensability to management or regulatory agencies. Even the specification of the desired structure between tariffs involves a certain element of judgement, and, in the real world, would involve continued iterations between solution and formulation.

11. LIMITATIONS

To conclude, we consider the ways in which this model could be extended, but which would lead outside the network flow formulation.

First, we must realize that our model is essentially an *approximation theory* for a *single* function for each class. For example, if we were trying to approximate a given function $\{g_i\}$ as closely as possible by another function $\{h_i + \beta\}$, β unknown, we could force this problem into our model by setting $f_i = g_i - h_i$, and then force $y_1 = y_2 = \dots y_n = \beta$ by using a complete order G with $R_{ij} = S_{ij} = 0$. Alternatively, if we had a free multiplication parameter choice, α, in an approximand $\{\alpha h_i\}$, we could use our model with $f_i = g_i/h_i$, and new weights $w_i^{\pm}/h_i$ for the $L_1^{\pm}$, $L_{\infty}^{\pm}$ norms, or $w_i^{\pm}/(h_i)^2$

for the $L_2^{\pm}$ norm, again using a complete order to force $y_1 = y_1 = \ldots y_n = \alpha$.

But, a two-dimensional approximand $\{\alpha h_i + \beta\}$ is already outside the realm of our model, since by setting $y_i = \alpha h_i + \beta$ or $y_i = \alpha + (\beta/h_i)$ one can get rid of either α or β, but not both, through the difference $y_j - y_i$. For this and higher-dimensional approximation theory, with or without order restrictions on the coefficients, one needs more general methods of linear or quadratic programming [8], [18], [12], [22], [16], [19], [5], [4].

12. Rate Relativities

Interestingly, a double-classification, additive problem in *rate relativities* [13], [1], can be formulated as a network model, with arbitrary norm, and solved by the methods here. Let $-y_i$ ($i = 1, 2, \ldots, p$) be the relativities for the first classification factor, $+y_j$ ($j = p + 1, p + 2, \ldots, p + q$) the relativities for the second, and f_{ij} the observed risk variable for joint class (i, j). The problem can be stated as the problem of determining the best-fit relativities $\{y = y_1, y_2, \ldots, y_{p+q}\}$ in the sense of minimizing the norm $\| y_j - y_i = f_{ij} \|$ over all joint classes.

When the details are carried through, one finds that the network G is of the *transportation* type [21], with p "plants" and q "customers", and unit profits $r_{ij} = f_{ij}$ on every arc (i, j) from plant to customer.

The flow dual depends upon the norm chosen; in the $L_1^{\pm}$ case, the exogenous "supplies" and "demands" are zero, but the flows can be negative, $-w_{ij}^+ \leq x_{ij} \leq x_{ij}^-$. In the $L_2^{\pm}$ case, the capacities are $-w_{ij}^+ u_{ij}^+ \leq x_{ij} \leq w_{ij}^- u_{ij}^-$, and so on.

Supplementary partial ordering among the y_i can be forced by adding additional arcs between appropriate nodes, while absolute bounds can be imposed by feeder arcs connecting a reference node 0, etc. Further details are left to the reader.

References

[1] Bailey, R. and L. Simon, "Two Studies in Automobile Insurance Ratemaking," *ASTIN Bulletin*, Vol. I, Part IV, pp. 192-217, (December 1960).

[2] Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, Statistical Inference under Order Restrictions, Wiley and Sons, New York, 388 pp., (1972).

[3] BARLOW, R. E. and V. A. UBHAYA, "Isotonic Approximation," in Proceedings of the Symposium on Optimizing Methods in Statistics, Academic Press, New York.
[4] BARROWDALE, I., "On Computing Best L_1 Approximations," in Approximation Theory, (A. Talbot, ed.), Academic Press, 356 pp., (1970).
[5] BARROWDALE, I. and A. YOUNG, "Algorithms for Best L_1 and L_∞ Linear Approximations as a Discrete Set," *Num. Math.*, Vol. 8, pp. 295-306, (1966).
[6] BOOT, J. C. G., Quadratic Programming, North-Holland, 210 pp., (1964).
[7] CRYER, J. D., T. ROBERTSON, F. T. WRIGHT and R. J. CASODY, "Monotone Median Regression," Dept. of Statistics, University of Iowa, (August 1972).
[8] DENNIS, J. B., Mathematical Programming and Electrical Networks, J. Wiley and Sons, 186 pp. (1959).
[9] DORN, W. S., "Duality in Quadratic Programming," *Quant. Appl. Math.*, Vol. 18, pp. 155-162, (1960).
[10] DUFFIN, R. J. and L. A. KARTOVITZ, "Formulation of Linear Programs in Analysis in Approximation Theory," *SIAM J. Appl. Math.*, Vol. 16, No. 4, pp. 662-675, (July 1968).
[11] FORD, L. R., JR. and D. R. FULKERSON, Flows in Networks, Princeton University Press, 1974 pp. (1962).
[12] GOLUB, G. H. and M. A. SAUNDERS, "Linear Least Squares and Quadratic Programming," Chapter 10 in Integer and Nonlinear Programming, (J. Abadie, ed.), North-Holland, 544 pp., (1970).
[13] JUNG, J., "On Automobile Insurance Ratemaking," *ASTIN Bulletin*, Vol. V. Part I, pp. 41-48, (May 1968).
[14] KELLEY, J. E., JR., "An Application of Linear Programming to Curve Fitting," *J. Soc. Indust. Appl. Math.*, Vol. 6, No. 1, pp. 15-22, (March 1958).
[15] KÜNZI, H. P., H. G. TZSCHACH, and C. A. ZEHNDER, Numerical Methods of Mathematical Optimization, Academic Press, 219 pp., (1971).
[16] LEWIS, J. T., "Approximation with Convex Constraints," *SIAM Review*, Vol. 15, No. 1, pp. 193-217, (January 1973).
[17] RABENOWITZ, P., "Mathematical Programming and Approximation," in Approximation Theory (A. Talbot, ed.), Academic Press, 356 pp., (1970).
[18] RABENOWITZ, P., "Applications of Linear Programming to Numerical Analysis," *SIAM Review*, Vol. 10, No. 2, pp. 121-159, (April 1968).
[19] RICE, J. R., The Approximation of Functions, Vol. I: Linear Theory, Addison-Wesley, 203 pp., (1964).
[20] SCHMITTER, H. and E. STRAUB, "Quadratic Programming in Insurance," *ASTIN Bulletin*, Vol. 8, No. 3, pp. 311-322, (March 1974).
[21] SIMMONARD, M., Linear Programming, Prentice-Hall, 430 pp., (1966).
[22] STIEFEL, E., "Note on Jordan Elimination, Linear Programming and Tchebycheff Approximation," *Num. Math.*, Vol. 2, pp. 1-17, (1960).
[23] UBHAYA, V. A., "Isotonic and Convex Optimization," *ORC* Report 71-13, Operations Research Center, University of California, Berkeley, (July 1971).
[24] VEINOTT, A. F., JR., "Least d-Majorized Network Flows with Inventory

and Statistical Applications," *Management Science*, Vol. 17, No. 9, pp. 547-567, (May 1971).

[25] WAGNER, H. M., "Linear Programming Techniques for Regression Analysis," *Journal Amer. Stat. Assoc.*, Vol. 54, pp. 206-212, (1959).

[26] WAGNER, H. M., "Non-Linear Regression with Minimal Assumptions," *Journal Amer. Stat. Assoc.*, Vol. 57, No. 299, pp. 573-578, (September 1962).