JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 14, 493-498 (1966)

The Shortest Route Through a Network with Time-Dependent Internodal Transit Times

KENNETH L. COOKE AND ERIC HALSEY

Pomona College, Claremont, California

Submitted by Richard Bellman

1. INTRODUCTION

Consider a set of N cities with every two linked by a road; if the cities are numbered in an arbitrary manner we can let the destination be N. Given the matrix function $G(t) = (g_{ij}(t))$, not necessarily symmetrical, where $g_{ij}(t)$, the time required to travel from city i to city j, varies as a function of the starting time t at i, we wish to find the path which minimizes the travel time from i to N when we begin our entire journey at $t = t_0$. If we wish, we can allow that two cities i and j are not connected at all or only by, say, a one-way road; in these cases we set the appropriate travel time $g_{ij}(t)$ equal to a very large number, or to ∞. Furthermore the times $g_{ij}(t)$ can vary with time in any way required.

In the case where each $g_{ij}(t)$ is a constant function t_{ij}, Bellman has established [1] the quantities f_i, where the f_i are the lengths of the optimal paths from i to N for $i = 1, \cdots, N-1$ and $f_N = 0$. He proved the existence and uniqueness of the f_i and using the principle of optimality [2] he showed that they satisfy the equations

$$f_i = \min_{j \neq i} (t_{ij} + f_j), \qquad i = 1, 2, \cdots, N-1$$

$$f_N = 0.$$

These were to be solved by the iteration scheme

$$f_i^{(k)} = \min_{j \neq i} (t_{ij} + f_j^{(k-1)}), \qquad i = 1, 2, \cdots, N-1$$

$$f_N^{(k)} = 0,$$

where the $f_i^{(k)}$ would converge after at most $N-1$ iterations provided that the initial guesses $f_i^{(0)}(t)$ were made correctly.

In order to simplify questions of existence and to provide a convenient basis for computation, we shall assume a discrete time scale t_0, $t_0 + 1$,

$t_0 + 2, \cdots$. Accordingly, we shall assume that all the $g_{ij}(t)$ are defined and have positive integer values for $t \in S = \{t_0, t_0 + 1, t_0 + 2, \cdots\}$ and $i \neq j$ (indeed we could as well assume that all $g_{ij}(t)$ are multiples of some unit Δ and use a time scale $\{t_0, t_0 + \Delta, t_0 + 2\Delta, \cdots\}$.) We now let $E_i(t)$ denote the set of all paths[1] which leave i at time $t \in S$ and reach N in a finite time after a finite number of steps. Since the time for each path in $E_i(t)$ is a positive integer, there is a minimum among these times. Now for $t \in S$ we define $f_i(t) =$ minimum time for a path in $E_i(t)$, $i = 1, 2, \cdots, N - 1$, $f_N(t) = 0$, where the minimum time $f_i(t)$ is achieved by at least one path from i to N with a finite number of links.

By the principle of optimality [2] we establish that for $t \in S$,

$$\begin{aligned} f_i(t) &= \min_{j \neq i} (g_{ij}(t) + f_j(t + g_{ij}(t))), \qquad i = 1, 2, \cdots, N - 1 \\ f_N(t) &= 0. \end{aligned} \tag{1}$$

2. Iteration

We now wish to define a sequence of iterates which will converge to the number $f_i(t_0)$ for each i, $1 \leqslant i \leqslant N$. For purposes of convenient calculation, we shall do this in such a way that we need values of the iterates only on a fixed finite set of time points $S_T = \{t_0, t_0 + 1, t_0 + 2, \cdots t_0 + T\}$, where the integer T is to be chosen so that all the quantities appearing in our iteration are well-defined for $t \in S_T$. As we shall see, one suitable choice is $T = M$, where the inequality

$$1 \leqslant g_{iN}(t_0) \leqslant M, \qquad i = 1, 2, \cdots, N \tag{2}$$

is satisfied by the integer M, since it is certainly possible to start at i at time t_0 and to reach N at or before time $t_0 + M$.

Let us proceed to the definition of the iterates. We begin by choosing $T = M$ to satisfy (2) and then by modifying the numbers $g_{ij}(t)$ as follows: for $t \in S$, we let

$$\tilde{g}_{ij}(t) = \begin{cases} g_{ij}(t), & \text{if} \quad t + g_{ij}(t) \leqslant t_0 + T, \\ \infty, & \text{if} \quad t + g_{ij}(t) > t_0 + T, \end{cases} \qquad i, j = 1, 2, \cdots, N. \tag{3}$$

In other words we eliminate from consideration any path-step for which the arrival time at the end of that path-step is not in S_T. Now we define for $t \in S$

$$\begin{aligned} f_N^{(0)}(t) &= 0 \\ f_i^{(0)}(t) &= \tilde{g}_{iN}(t), \qquad i = 1, 2, \cdots, N - 1. \end{aligned} \tag{4}$$

[1] By the assumptions above, $E_i(t) \neq \phi$.

Thus $f_i^{(0)}(t)$ is the time for the one-link path from i to N starting at time t provided that N is reached at or before time $t_0 + T$, and otherwise is ∞. In particular $f_i^{(0)}(t) = \infty$ for $t > t_0 + T$ and $f_i^{(0)}(t) = g_{iN}(t)$ for $t = t_0$.

Now let $E_i^{(2)}(t)$ denote the set of all paths (if any) consisting of one or two links, leaving i at time t, and reaching N at or before time $t_0 + T$. Then define for $t \in S$

$$f_N^{(1)}(t) = 0$$

$$f_i^{(1)}(t) = \begin{cases} \text{minimum time for a path in } E_i^{(2)}(t),\ E_i^{(2)}(t) \neq \phi, \\ \infty,\ E_i^{(2)}(t) = \phi, \qquad i = 1, 2, \cdots, N-1. \end{cases}$$

Then the principle of optimality implies that

$$f_N^{(1)}(t) = 0$$

$$f_i^{(1)}(t) = \min_{j \neq i} (\tilde{g}_{ij}(t) + f_j^{(0)}[t + \tilde{g}_{ij}(t)]), \qquad i = 1, 2, \cdots, N-1. \tag{5}$$

To prove (5), we observe that if $E_i^{(2)}(t)$ is empty, the one-link path from i to N and starting at i at time t must reach N after time $t_0 + T$, so that $\tilde{g}_{iN}(t) = \infty$. Also, no two-link path from i to j to N can reach N by time $t_0 + T$, so that no one-link path leaving a j at time $t + \tilde{g}_{ij}(t)$ can reach N by time $t_0 + T$. Therefore either $\tilde{g}_{ij}(t) = \infty$ or $f_j^{(0)}(t + \tilde{g}_{ij}(t)) = \infty$, and the right member in (5) is ∞, which is the value assigned to $f_i^{(1)}(t)$ by definition. If, on the other hand, $E_i^{(2)}(t)$ is nonempty, then there exists at least one path of one or two links leaving i at time t and reaching N by time $t_0 + T$, and there is a minimal such path. This minimal path proceeds from i to some j, and then follows the one-link path from j to N which takes the least time starting at j at time $t + \tilde{g}_{ij}(t) = t + g_{ij}(t)$ (unless the minimal path occurs for $j = N$, in which case the validity of (5) is clear).

In general, we let $E_i^{(k)}(t)$ denote the set of all paths of at most k links leaving i at time t and reaching N at or before time $t_0 + T$ and we define for $t \in S$

$$f_N^{(k)}(t) = 0$$

$$f_i^{(k)}(t) = \begin{cases} \text{minimum time for a path in } E_i^{(k+1)}(t), \qquad \text{if} \qquad E_i^{(k+1)}(t) \neq \phi, \\ \infty, \qquad \text{if} \qquad E_i^{(k+1)}(t) = \phi, \end{cases}$$

$$i = 1, 2, \cdots, N-1,$$

$$k = 1, 2, 3, \cdots.$$

We assert that for $t \in S$

$$f_N^{(k)}(t) = 0$$

$$f_i^{(k)}(t) = \min_{j \neq i} (\tilde{g}_{ij}(t) + f_j^{(k-1)}[t + \tilde{g}_{ij}(t)]), \qquad i = 1, 2, \cdots, N-1. \tag{6}$$

The validity of Eq. (6) is established by arguments similar to those which justified Eq. (5).

Now, suppose that T has been chosen to satisfy (2). Since the optimal path from i to N starting at time t_0 reaches N at time $t_0 + f_i(t_0)$, which is $\leqslant t_0 + T$, and since it has a certain finite number of steps $k_0 + 1$, it is a path in $E_i^{(k_0+1)}(t_0)$. It follows that we must obtain convergence in a finite number of iterations, that is

$$f_i^{(k)}(t_0) = f_i(t_0) \qquad \text{for} \qquad k \geqslant k_0 .$$

In fact, whenever $t + f_i(t) \leqslant t_0 + T$, there is a minimal path from i reaching N by $t_0 + T$, and having a finite number of steps. Therefore there exists an integer k_0 depending on i and t such that

$$f_i^{(k)}(t) = f_i(t) \qquad \text{for} \qquad k \geqslant k_0 \qquad \text{iff} \qquad t + f_i(t) \leqslant t_0 + T.$$

(By the definitions of the $f_i^{(k)}(t)$, we ascertain that

$$f_i^{(k)}(t) \leqslant f_i^{(k-1)}(t) \leqslant \cdots \leqslant f_i^{(0)}(t), \qquad 1 \leqslant i \leqslant N, \qquad t \in S.$$

In particular, then,

$$f_i^{(k)}(t_0) \leqslant f_i^{(0)}(t_0) = g_{iN}(t_0) \leqslant M = T,$$

so we are indeed sure that $t_0 + f_i^{(k)}(t_0) \leqslant t_0 + T$, and finally that $t_0 + f_i(t_0) \leqslant t_0 + T$.)

3. Computation

Like other iterative schemes, whether they are for the shortest route problem or not, this one is well suited to the abilities of a digital computer. In practice, if we wish to compute $f_i(t_0)$ for $1 \leqslant i \leqslant N$ we begin by selecting a number T large enough to satisfy (2); or, if we wish to compute simultaneously the values of $f_i(t)$ for $t = t_0, t_0 + 1, \cdots, t_0 + \tau$, we can choose

$$T \geqslant \max_{i,t} (g_{iN}(t)), \qquad \begin{array}{l} i = 1, 2, \cdots, N, \\ t = t_0, t_0 + 1, \cdots, t_0 + \tau. \end{array} \tag{7}$$

In some cases either of these definitions of T may yield $T = \infty$ (or "T equals a very large number"). Since this will be seen to be undesirable in actual computation, a valid alternate definition of T is obtained by replacing $g_{iN}(t_0)$ in (2) (or $g_{iN}(t)$ in (7)) by "the path starting at i at time t_0 (or "at time t" in (7)), containing the smallest possible number of links, yet still reaching N in an acceptably short time." We find such a path by first examining all two-link paths from i to N, then all three-link paths, etc., until we find an acceptable k_i'-link path. "Acceptably small" may be defined by the maximum storage capacity of our digital computer, by limitations on computing time, or *quoi que ce soit.*[2] Now if (for instance) (2) were used to choose T, then we could only be sure that $t_0 + f_i^{(k)}(t_0) \leqslant t_0 + T$ if $k \geqslant k' - 1$, where

$$k' = \max_i (k_i'), \qquad 1 \leqslant i \leqslant N.$$

Now the $g_{ij}(t)$ are replaced by the $\tilde{g}_{ij}(t)$ defined in (3) and an "initial guess" matrix $F_i^{(0)}(t)$, which contains values of $f_i^{(0)}(t)$ for $1 \leqslant i \leqslant N$, $t \in S_T$, is calculated, using Eq. (4). (It is not necessary to carry the values of the $f_i^{(0)}(t)$ for $t \notin S_T$.) Then, using (5), we calculate the matrix $F_i^{(1)}(t) = (f_{i,t})$, where $f_{i,t} = f_i^{(1)}(t)$: If $\tilde{g}_{ij}(t) = \infty$ for all j, then $f_i^{(1)}(t) = \infty$. If $t + \tilde{g}_{ij}(t) \notin S_T$ for any j then by (4) and (3) $f_j^{(0)}(t + \tilde{g}_{ij}(t)) = \infty$, and by (5) $f_i^{(1)}(t) = \infty$. When both $\tilde{g}_{ij}(t)$ and $f_j^{(0)}(t + \tilde{g}_{ij}(t))$ are finite for some j, we compute the former quantity directly, extract the latter from $F_i^{(0)}(t)$, and compute the minimum over j of their sum to find $f_i^{(1)}(t)$. In the same way, using (6), we successively compute the matrices $F_i^{(2)}(t), \cdots, F_i^{(k)}(t), \cdots$, for $1 \leqslant i \leqslant N$, $t \in S_T$. Eventually the vectors

$$< f_i^{(0)}(t_0) >, < f_i^{(1)}(t_0) >, \cdots < f_i^{(k)}(t_0) >, \cdots$$

necessarily converge to the shortest route vector $\langle f_i(t_0) \rangle$. The shortest route itself is obtained by keeping track of the indices for which the various minima occur.

If greater accuracy than integer accuracy is required in the functions $g_{ij}(t)$, then the units Δ mentioned above may be used instead of integers.

---

[2] Our method requires storage of the values of $\tilde{g}_{ij}(t)$ and $f_i(t)$ for $1 \leqslant i, j \leqslant N$, $t \in S_T$, a total of $(N^2 + N)(T + 1)$ values.

If the minimum route from i to N starting at t_0 lies in $E_i^{(k_0+1)}(t_0)$, but not in $E_i^{(k)}(t_0)$ for $1 \leqslant k \leqslant k_0$, then $f_i(t_0) \geqslant k_0 + 1$, since $g_{ij}(t) \geqslant 1$ for $t \in S$. But we know that $f_i(t_0) \leqslant T$. Therefore $k_0 \leqslant T - 1$ and convergence occurs in at most $k_0 \leqslant T - 1$ iterations.

Various methods to reduce the storage requirement and to speed convergence may be divised, but they are not treated here.

Summary

The shortest route problems so far analyzed fall seriously short of reality in that they assume that the time required between any two vertices (nodes) is constant, an assumption which is certainly not true in many physical and biological applications. This analysis demonstrates how to perform a shortest route iteration in the more realistic case where internodal time requirements are time-variable.

In the paper a modified form of Bellman's iteration scheme [1] for finding the shortest route between any two vertices in a network is developed for application to our generalized case. It converges to the shortest path between any two vertices in a finite number of iterations, and for any initial starting time, provided that certain initial conditions are satisfied. One of the main points of this treatment has been to arrange the work so that when computing a new iterate one does not have to recompute previous iterates at more advanced time points.

References

1. R. Bellman. On a routing problem. *Quart. Appl. Math.* **16** (1958), 87-90.
2. R. Bellman. "Dynamic Programming," Princeton University Press, Princeton, New Jersey, 1957.