On the Applications of Optimal Control Theory and Dynamic Programming in Ship Routing

S. J. BIJLSMA
Den Helder, The Netherlands

Received February 2002; Revised July 2002

ABSTRACT: *In this paper, the maximum principle of optimal control theory and the method of dynamic programming are discussed in relation to the minimization of fuel consumption in ship routing. The connection between the two methods is indicated for the case in which ship routing is treated as a continuous process, meaning that the sailing paths are not restricted to arcs of a grid as in the discrete dynamic programming method, but can vary continuously in the navigation area. Practical aspects are also discussed, such as the discrete approach of dynamic programming, as well as the finite version of the continuous approach and the limited predictability of the weather. Results are presented showing least-time routes, which are obtained with computational methods based on the maximum principle and the corresponding continuous type of dynamic programming.*

INTRODUCTION

In ship weather routing, one is concerned with the optimization of quantities such as the transit time of a ship as it travels over the ocean between two given points, its fuel consumption, or certain criteria for the benefit of passengers or cargo. To achieve such optimization, the sea state must be known for the complete passage, as well as the response of a ship to waves and its performance under specific weather conditions.

Several methods have been proposed for the solution of these optimization problems. The first was a manual method for minimal-time routing (see, for instance, [1]), using time fronts similar to wave fronts in geometrical optics. The most obvious next step was to program the manual method for the computer. This has been done in [2–4] for the computation of a least-time track and in [5] for the computation of a minimum fuel route. An alternative approach includes applications of optimal control theory or calculus of variations (see, for example, [6–9]), as well as applications of the theory of multistage decision processes (dynamic programming) (see, for instance, [10–13]).

Optimal control theory is an extension of the classical calculus of variations in that the spatial and control variables may belong to closed sets instead of being restricted to open sets, as is required in the calculus of variations; however, these terms are used indifferently. For simplicity, it is assumed here that the spatial and control variables lie in open sets.

In a previous paper [9], a numerical method is presented for the computation of an optimal route that minimizes fuel consumption. This method is based on Pontryagin's maximum principle of optimal control theory [14], in the calculus of variations corresponding with the Weierstrass condition completed with the Euler-Lagrange equations ([9], p. 146). Optimal route calculations using dynamic programming are based on Bellman's principle of optimality [15].

In this paper, the equivalence of the two methods is demonstrated. This is done by showing that the maximum principle can be derived from the principle of optimality if ship routing is considered as a continuous optimization problem, meaning that the sailing paths are not restricted to arcs of a grid as in the discrete dynamic programming method, but can vary continuously in the navigation area. Relevant here is a paper [16] that considers the time-optimal control problem as an optimal wave propagation problem. The computerized manual method described in [3] is based on the proposition ([16], p. 9) that the points of a minimal-time route fall within the set of ultimately attainable points (time fronts) for all times. Of course this result is valid only for infinitesimal time steps and not necessarily for the time steps of 12 h used in this method. It certainly holds, however, in fine weather regions where the optimal route is expected to be found. Successive time fronts are constructed using the principle of dynamic programming, in this case expressed

NAVIGATION: *Journal of The Institute of Navigation*
Vol. 49, No. 2, Summer 2002
Printed in the U.S.A.

by equation (10), which implies that the course of a ship along an optimal route must be chosen so that the velocity attains a maximum in the direction of the normal to a time front. Therefore, the method in [3] can be considered the finite version of the continuous dynamic programming method as well.

With the discrete version of the method of dynamic programming, the optimal route calculation usually begins with the introduction of a guess route and the construction of (one-dimensional) search grids at a distinct number of points perpendicular to this route. Then an approximate optimal route connecting points of successive search grids is calculated by applying the principle of optimality [10, 11]. Considering this approximate optimal route as a new guess route, a better approximation can be obtained by refining the search grids and the distance between them along the new guess route and repeating the optimal route calculation until further refinement yields no substantial improvement [12, 13]. For the sake of completeness, it is noted that the method of [12] has been applied to a ship routing problem that is studied in [7].

The application of optimal control theory as presented in [8, 9] has a number of advantages relative to the application of the discrete version of dynamic programming in ship weather routing. Here we note its ability to reduce the number of control variables, its capacity to provide insight into the nature of the problem, and its clarity of visual presentation by showing all possible extremals emanating from the point of departure on their way to the end point.

The aim of this paper is to elucidate the relation between dynamic programming and optimal control theory for those who use these methods in ship routing and to counter the arguments made against the application of optimal control theory. For instance, it has been suggested that the application of optimal control theory, which treats ship routing as a continuous optimization problem subject to a number of necessary conditions for a local minimum, could lead to problems associated with convergence to a global minimum [4], as well as to the presence of spatial derivatives that cannot be approximated with sufficient accuracy because of errors in the forecast data [4, 5], especially in the face of severe weather conditions.

With respect to the first point, it is noted that, apart from the iterative methods used in optimal control theory (see [6]), it is the iterative variant of the discrete method of dynamic programming in particular that could involve convergence problems. This is the case because, as noted earlier, a proper optimal route calculation using the discrete method starts on a coarse grid and is then continued iteratively on more and more refined grids. This is so in turn because refinement of the initial coarse grid to make it suitable for a sufficiently accurate optimal route calculation is unrealistic given the required computation time and memory space.

With respect to the second point, the first-order spatial derivatives occurring in the equations governing the course of an optimal route in optimal control theory can be approximated extremely well. This fact is demonstrated by the experiments in this paper, in which the methods of [3] and [9] are compared for the case of minimal-time ship routing.

The application of the maximum principle to ship routing is described in the following section. Next the connection with dynamic programming is discussed. It was noted above that to solve the optimization problem, the sea state must be known for the complete passage. This will in general not be the case. The limited predictability of the weather plays an important role in the optimal route calculation. Therefore, attention is paid to the limited predictability of the weather and to the practical aspects of the two optimization methods. Results of these methods are provided for the case of minimal-time ship routing. Conclusions are presented in the last section.

THE MAXIMUM PRINCIPLE IN SHIP ROUTING

In this section, the maximum principle from [14] is applied to the problem of ship routing. The problem of minimizing fuel consumption during an ocean crossing is considered. For the sake of completeness, it is noted that the resulting equations also apply in the case of other cost or penalty functions. The speed V and the course p of the ship are used as control variables. The equations of motion of a ship in a Cartesian coordinate system with coordinates x_1 and x_2 are given by

$$\dot{x}_1 = V \cos p + S_1(t, x_1, x_2) \qquad (1)$$

$$\dot{x}_2 = V \sin p + S_2(t, x_1, x_2) \qquad (2)$$

where the dot denotes differentiation to the time t. The functions $S_1(t, x_1, x_2)$ and $S_2(t, x_1, x_2)$ denote the velocity components of the ocean current. It is assumed that the rate of decrease of fuel can be described by the equation

$$\dot{x}_0 = f_0(t, x_1, x_2, V, p)$$

where the function $x_0(t)$ denotes fuel consumption. The path of a ship is completely determined by the initial values of x_1 and x_2 and by the values of V(t) and p(t) over the voyage, so that the problem under consideration is now to determine the functions

$$V(t), p(t) \qquad 0 \leq t \leq t_1$$

which will minimize the integral

$$\int_0^{t_1} f_0(t, x_1, x_2, V, p)\, dt \qquad (3)$$

among all paths with prescribed conditions on initial and final values

$$x_i(0) = x_{i0}, x_i(t_1) = x_{i1} \qquad (i = 1,2)$$

For the sake of simplicity, ocean current is omitted. Of course, ocean current may be included, but its inclusion would complicate the discussion unnecessarily at this point (see also [9], p. 147). In the results presented in this paper, ocean current is included. The maximum principle can now be formulated as follows (cf. [14], p. 59).

Let V(t) and p(t), $0 \le t \le t_1$ be continuous control functions such that the corresponding trajectory $x(t) = (x_1(t), x_2(t))$, satisfying equations (1) and (2) and issuing at instant $t = 0$ from point (x_{10}, x_{20}), passes at time $t = t_1$ through point (x_{11}, x_{21}). The necessary condition for the control functions V(t) and p(t) and the trajectory x(t) to be optimal (i.e., to minimize equation (3)) is that there exist a nonzero continuous differentiable vector function $\lambda(t) = (\lambda_0(t), \lambda_1(t), \lambda_2(t))$ and a function

$$H(t, x, V, p, \lambda) = \lambda_0 f_0(t, x_1, x_2, V, p) + \lambda_1 V \cos p + \lambda_2 V \sin p$$

connected by the differential equations

$$\dot{\lambda}_i = -H_{x_i} \qquad (i = 0, 1, 2) \tag{4}$$

where variables as subscripts denote partial differentiation, such that:

- For all t, $0 \le t \le t_1$, the function $H(t, x(t), V, p, \lambda(t))$ of the variables V and p attains a maximum at the point $V = V(t)$, $p = p(t)$:

  $$H(t, x(t), V(t), p(t), \lambda(t)) = M(t, x(t), \lambda(t)) \tag{5}$$

  where

  $$M(t, x(t), \lambda(t)) = \sup_{V, p} H(t, x(t), V, p, \lambda(t))$$

  for fixed t, x(t) and $\lambda(t)$.
- At the final instant $t = t_1$, the conditions

  $$\lambda_0(t_1) = \text{constant} \le 0, \; M(t_1, x(t_1), \lambda(t_1)) = 0$$

  are satisfied.

If the strict upper bound of the values of the function H is attained at some point of the control domain, as is the case in this paper, $M(t, x, \lambda)$ is the maximum of the values of the function H for fixed t, x, and λ. Therefore, this necessary condition for optimality is called the maximum principle. We return to this maximum principle in discussing methods of solution for the ship routing problem. The similarity of the maximum principle and the principle of optimality is indicated in the following section, with ship routing being treated as a continuous optimization problem.

CONNECTION WITH DYNAMIC PROGRAMMING

Referring to the preceding section, another method for the optimization of similar processes is dynamic programming. The functional equations governing the process of dynamic programming are obtained from application of the principle of optimality, using the concept of an optimal policy. If a policy is any rule for making decisions (the effect of a decision being a transformation of the state variables) that yields an allowable sequence of decisions, then an optimal policy is one that maximizes a preassigned function of the final state variables. We are now in a position to formulate the principle of optimality ([15], p. 83): An optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

The focus here is on the connection between the method of dynamic programming and the maximum principle for the case of ship routing. Let the fuel consumption along an optimal trajectory from point $x(t) = (x_1(t), x_2(t))$ $(0 \le t \le t_1, \; x_i(0) = x_{i0}, \; i = 1, 2)$, belonging to the set of points Ω from which an optimal passage to the end point $x_i(t_1) = x_{i1} (i = 1,2)$ is possible, to $x(t_1)$ be denoted by

$$J(x(t)) = \int_t^{t_1} f_0(t, x_1(t), x_2(t), V(t), p(t))dt$$

using the fact that every piece of an optimal trajectory is itself an optimal trajectory. For convenience (cf. [14], p. 68), we define

$$I(x(t)) = -\int_t^{t_1} f_0(t, x_1(t), x_2(t), V(t), p(t))dt = I(x(0)) + \int_0^t f_0(t, x_1(t), x_2(t), V(t), p(t))dt \tag{6}$$

From equation (6), it follows that

$$\dot{I}(x(t)) = \sum_{i=1}^{2} I_{x_i}(x(t))\dot{x}_i(t) = f_0(t, x_1(t), x_2(t), V(t), p(t)) \tag{7}$$

Let us consider the same problem for arbitrary control functions V and p, starting from point x(t) on the optimal trajectory. After a time interval dt, the new position will be x(t) + dx, with $dx = (dx_1, dx_2) = (V \cos p, V \sin p)dt$. Now the fuel consumption along a trajectory from x(t) along x(t) + dx to $x(t_1)$ (moving from x(t) + dx optimally to the end point $x(t_1)$) is given by

$$J(x(t) + dx) + f_0(t, x(t), V, p)dt$$

so that

$$J(x(t) + dx) + f_0(t, x(t), V, p)dt \ge J(x(t))$$

or

$$I(x(t) + dx) - I(x(t)) \le f_0(t, x(t), V, p)dt$$

This last expression can be written as

$$(I_{x_1}(x(t))V \cos p + I_{x_2}(x(t))V \sin p - f_0(t, x(t), V, p))dt \le 0 \tag{8}$$

From equations (7) and (8) it follows that

$$\max_{V,p} (I_{x_1}(x(t))V \cos p + I_{x_2}(x(t))V \sin p - f_0(t, x(t), V, p)) = 0$$

where the maximum is attained for $V = V(t)$, $p = p(t)$. We may define the equation

$$\max_{V,p} (I_{x_1}(x)V \cos p + I_{x_2}(x)V \sin p - f_0(t, x, V, p)) = 0 \qquad (9)$$

on the set of points Ω, where the maximum is attained for the values of the optimal control functions at the instant of emanation from point x. Analogous to equation (83) from [14], equation (9), expressing the principle of dynamic programming for the present problem, can be called Bellman's equation. It says that the control variables V and p must be chosen such that the scalar product of the gradient (I_{x_1}, I_{x_2}) and the velocity (V cos p, V sin p), diminished with the function $f_0(t, x_1, x_2, V, p)$, attains a maximum value (equal to zero). From this equation, the maximum principle can be deduced. It will not be surprising that the derivatives I_{x_1} and I_{x_2} can be identified with the Lagrange multipliers λ_1 and λ_2 from the foregoing section. Indeed, from equations (7) and (9) it follows that for optimal choices $V = V(t)$, $p = p(t)$, the expression

$$I_{x_1}(x)V(t) \cos p(t) + I_{x_2}(x)V(t) \sin p(t) - f_0 (t, x, V(t), p(t))$$

as a function of the variable x, defined on the set Ω, attains a maximum value at the point $x = x(t)$. Thus

$$I_{x_1x_j}(x(t))V(t) \cos p(t) + I_{x_2x_j}(x(t))V(t) \sin p(t) - f_{ox_j}(t, x(t), V(t), p(t)) = 0 \qquad (j = 1, 2)$$

is satisfied along an optimal trajectory or with equations (1), (2), and (7):

$$\dot{I}_{x_j}(x(t)) = f_{ox_j}(t, x(t), V(t), p(t)) \qquad (j = 1, 2)$$

which is in accordance with equation (4) with the choice $\lambda_0 = -1$. In the time-optimal case, the cost function $f_0(t, x, V, p) = 1$, and the speed is a known function of t, x, and p. In this case, equation (9) reads

$$\max_{p} (I_{x_1}(x)V \cos p + I_{x_2}(x)V \sin p) = 1 \qquad (10)$$

For an optimal choice $p = p(t)$, the expression

$$I_{x_1}(x)V \cos p(t) + I_{x_2}(x)V \sin p(t)$$

as a function of the variable x, defined on the set Ω, has a maximum at the point $x = x(t)$. Thus

$$\sum_{i=1}^{2} I_{x_ix_j}(x(t))\dot{x}_i(t) + I_{x_1}(x(t))V_{x_j} \cos p(t) + I_{x_2}(x(t))V_{x_j} \sin p(t) = 0$$

or

$$\dot{I}_{x_j}(x(t)) = -I_{x_1}(x(t))V_{x_j} \cos p(t) - I_{x_2}(x(t))V_{x_j} \sin p(t) \qquad (j = 1, 2)$$

In the following section, results of the method in [3], using relation (10), in which (I_{x_1}, I_{x_2}) could be identified as the normal to a time front, are compared with those of the method in [9] for the case of minimal-time routing.

PRACTICAL CONSIDERATIONS

In this section, practical aspects of computational methods based on both the maximum principle and the method of dynamic programming are considered. In these methods, it is assumed that the weather forecast is valid for the entire crossing. An overview of the essential features of these methods is presented.

The trajectory characterized by the functions x(t), V(t), and p(t) satisfying the necessary conditions of the maximum principle is supposed to be normal. As a consequence of that normality, the equality sign for the multiplier λ_0 is excluded. The abnormal case is highly singular and is not discussed. In fact, if the optimal trajectory is abnormal, it may be the only trajectory satisfying the conditions of the present problem (see [17]). As mentioned previously, it is supposed that the functions x(t), V(t), and p(t) lie in open sets. Cases in which these variables are limited to a closed set, meaning for practical ship routing that the navigation area is bounded or that certain courses are forbidden, can be found elsewhere [8].

Methods Based on the Maximum Principle

Most numerical methods used in ship routing that employ calculus of variations or optimal control theory are iterative methods. Essentially, these methods start with a guess route as a first approximation of an optimal trajectory that is to be determined by means of an iterative procedure using variational relations between fuel consumption and the position and control variables. The Euler-Lagrange equations ([9], p. 146)

$$H_v(t, x(t), V(t), p(t), \lambda(t)) = 0, \quad H_p(t, x(t), V(t), p(t), \lambda(t)) = 0$$

are also involved. These equations can be derived from equation (5), since V(t) and p(t) belong to open sets. A comprehensive survey of the iterative methods can be found in [6]. Since the appearance of that paper, little appears to have changed.

A disadvantage of these iterative methods is that they may lead to relative minima or may exhibit problems associated with convergence. Therefore, a method is discussed that does not have these disadvantages.

Reformulating equation (5) of the maximum principle, we seek to find functions V(t) and p(t) that satisfy

$$\max_{V,p} (\lambda_1(t) \frac{V}{f_0(t, x(t),V,p)} \cos p + \lambda_2(t) \frac{V}{f_0(t, x(t),V,p)} \sin p) = -\lambda_o \qquad (11)$$

for fixed t, x(t), and $\lambda(t)$. Since λ_0 can be chosen arbitrarily, and the parameters λ_0, λ_1, and λ_2 have a common factor of proportionality, the solutions of equation (4), where the control functions V and p are determined by equation (11), do not change if the parameters λ_1 and λ_2 are multiplied by an arbitrary constant. To include all extremals emanating from the point of departure, we write $\lambda_1(0) = \cos a$ and $\lambda_2(0) = \sin a$ for any choice of λ_0. For instance, $\lambda_0 = -1$ can be chosen. All extremals emanating from the point of departure can then be found by varying the parameter a. Equation (11) expresses that the speed V and the course p along an extremal must be chosen such that the 'velocity'

$$\left(\frac{V}{f_0(t, x(t), V, p)} \cos p, \frac{V}{f_0(t, x(t), V, p)} \sin p\right)$$

attains a maximum value in the direction ($\lambda_1(t)$, $\lambda_2(t)$). It can be shown that, under certain conditions, $x_1(t, a)$, $x_2(t, a)$, $\lambda_1(t, a)$, and $\lambda_2(t, a)$ as solutions of equations (1), (2), and (4), where V(t) and p(t) are implicitly defined by equation (11), depend continuously on the parameter a for a fixed value of t. This property serves as a starting point for a computational method that is described extensively in [9]. The essence of the method is illustrated in Figure 1 for the case in which the sailing time is minimized.

Methods Based on the Method of Dynamic Programming

The method described in [3] can be considered a finite version of the continuous dynamic programming approach to ship routing. The method is based on the fundamental property ([16], p. 9) that each point of a time-optimal trajectory at a specific time belongs to the boundary of the set of attainable points at that time. In fact, this property also holds if fuel consumption is minimized (see [18]). In that case, the boundary of the set of attainable points can be defined as the set of points that can ultimately be attained within a specific time with minimal fuel consumption. These boundaries are called time fronts. The optimal control problem is then reduced to the study of these time fronts.

Actually, this property holds for infinitesimal time steps and not necessarily for 12 h time steps, as is clear from Figure 2, which shows a situation in which the necessary condition of Jacobi ([9], p. 149) is violated. This condition expresses that a nonsingular extremal between two points cannot be a minimal curve if it contains a conjugate point between those points.

Successive time fronts are constructed using the principle of dynamic programming expressed by equation (10). The weakness of this method is associated with the geometrical construction of the gradients (normals) to these time fronts, a method that is more appropriate for a manual application. For that reason, the method could be considered a computerized manual method as well. The method has proven very useful as a frame of reference for the approximation of spatial derivatives used in the method described in [8] and [9]. Results of both methods are compared in Figures 3 through 6 for different ships' performance data for the case of minimal-time ship routing. The navigation area is mapped conformally onto a plane by means of stereographic projection (see [8], p. 14). The method of [3]

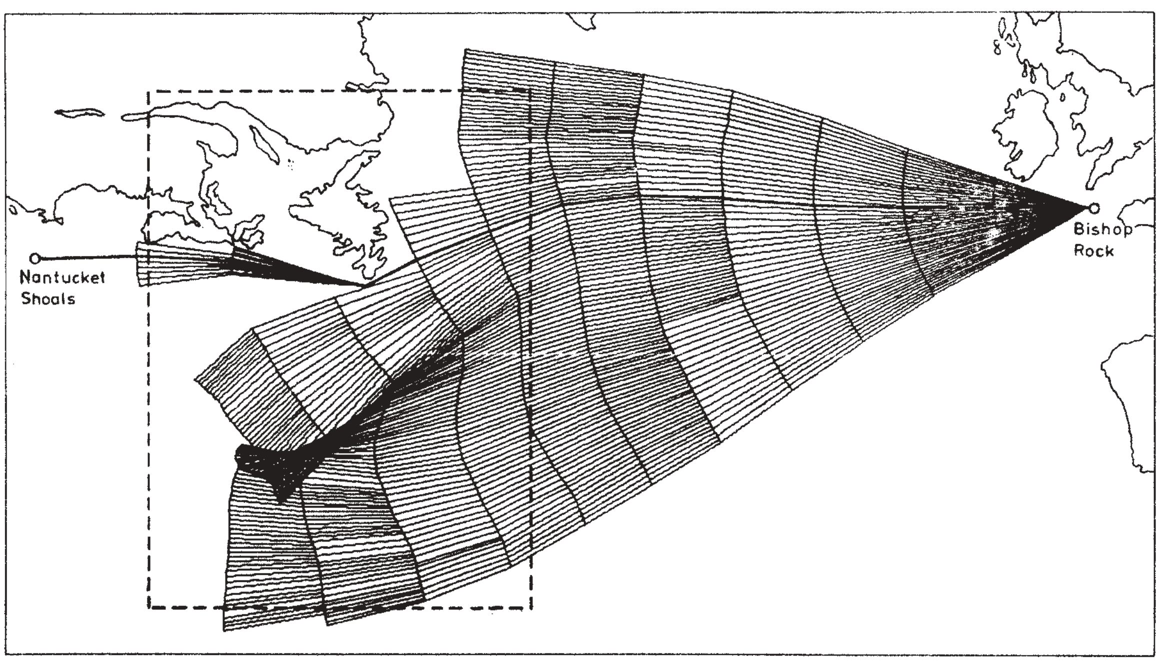

Fig. 1–Example of an Ocean Crossing Using a Computational Method Based on the Maximum Principle (The least-time track is indicated by the solid line.)

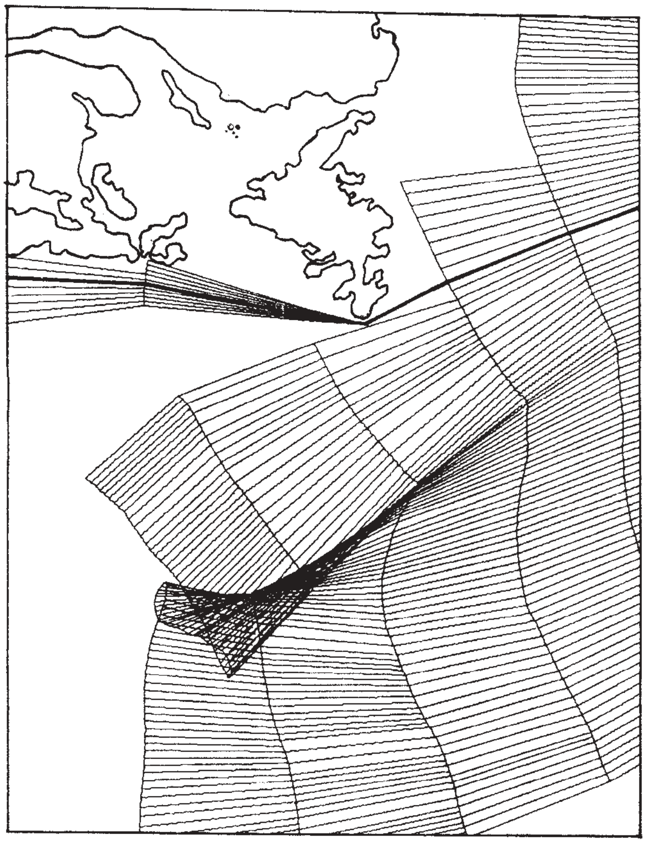

Fig. 2–A Detail from Figure 1 (Explanation is given in the text.)

is indicated by the dotted-dashed line, the method of [9] by the solid line, and the great circle by the dashed line. The service speeds of the chosen general cargo ships were 13 and 21 kn, respectively. The polar velocity diagrams used in the computations are of elliptic form and are constructed with the aid of performance data in the case of following, beam, and head waves.

These ship's performance data were corrected by recomputing the recommended routes of the ships, constructed manually by the former Ship Routing Office of the Royal Netherlands Meteorological Institute (KNMI), with analyzed wave charts. Ocean current was included. The crossing times along the least-time routes computed with the methods of [3] and [9] and the great circle in Figure 3 were, respectively, 6 d, 1.0 h; 6 d, 3.0 h; and 6 d, 17.2 h. The detours of the least-time routes of the methods of [3] and [9] with respect to the great circle were 107 and 101 mi, respectively. For Figures 4 through 6, these numbers were:

- Figure 4—6 d, 8.2 h; 6 d, 8.2 h; 6 d, 20.8 h; 104 mi; 89 mi
- Figure 5—10 d, 23.5 h; 11 d, 1.3 h; 14 d, 8.2 h; 242 mi; 215 mi
- Figure 6—11 d, 13.0 h; 11 d, 13.2 h; 13 d, 14.1 h; 363 mi; 350 mi

Of course, these numbers should be considered indications of the sailing times along the least-time tracks, rather than being interpreted in absolute terms. For the sake of completeness, the positions of the manually constructed optimal routes are indicated by the dotted lines. (For more information on this subject, the reader is referred to Scientific Reports WR 72-1, 72-2, 72-11, 73-2, and 73-5 [in Dutch] of the KNMI.)

As noted earlier, dynamic programming methods of the discrete type usually start an optimal route calculation by introducing a guess route (usually the great circle or a climatological route) and constructing (one-dimensional) initially rather coarse search grids at a distinct number of points perpendicular to this route, as shown in Figure 7. Then an optimal route connecting starting and end points is computed by applying the principle of optimality to the points of

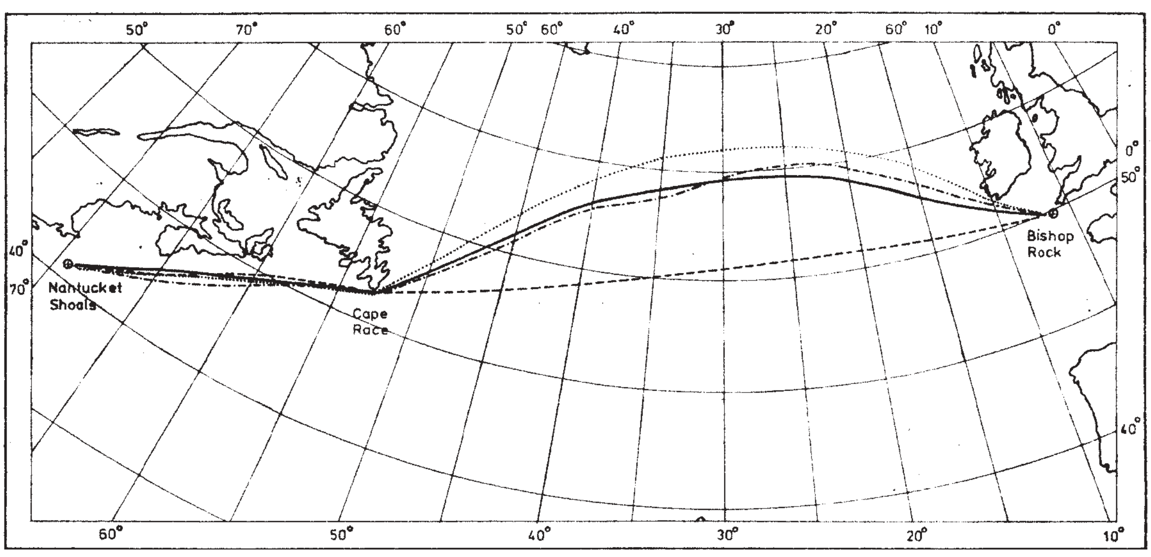

Fig. 3–Results of Both Methods (The ship departed from Bishop Rock on January 20, 1970, at 00.00 GMT. The service speed was 21 kn.)

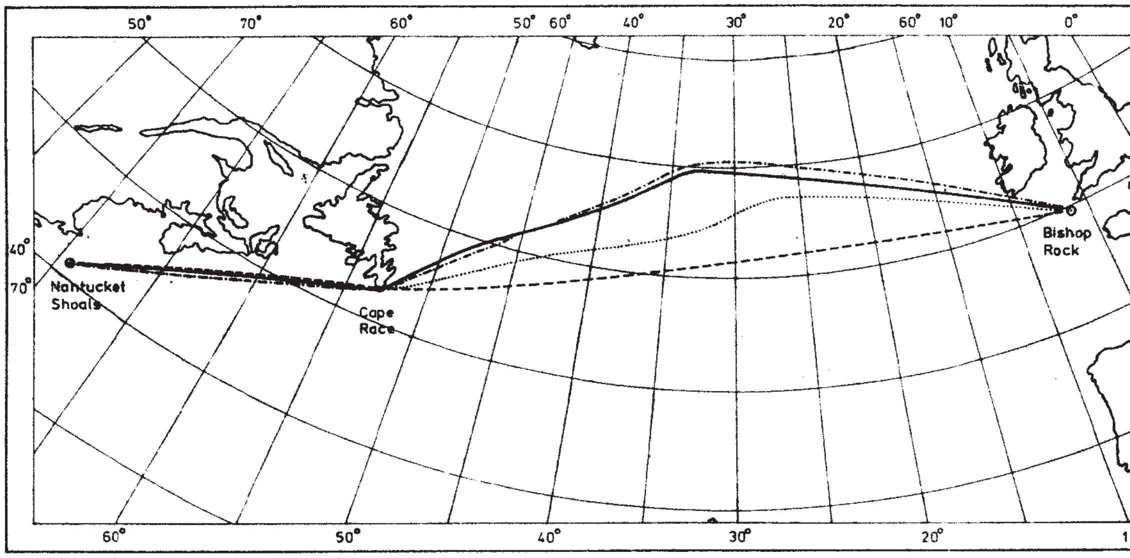

Fig. 4–Results of Both Methods (The ship departed from Bishop Rock on January 22, 1970, at 00.00 GMT. The service speed was 21 kn.)

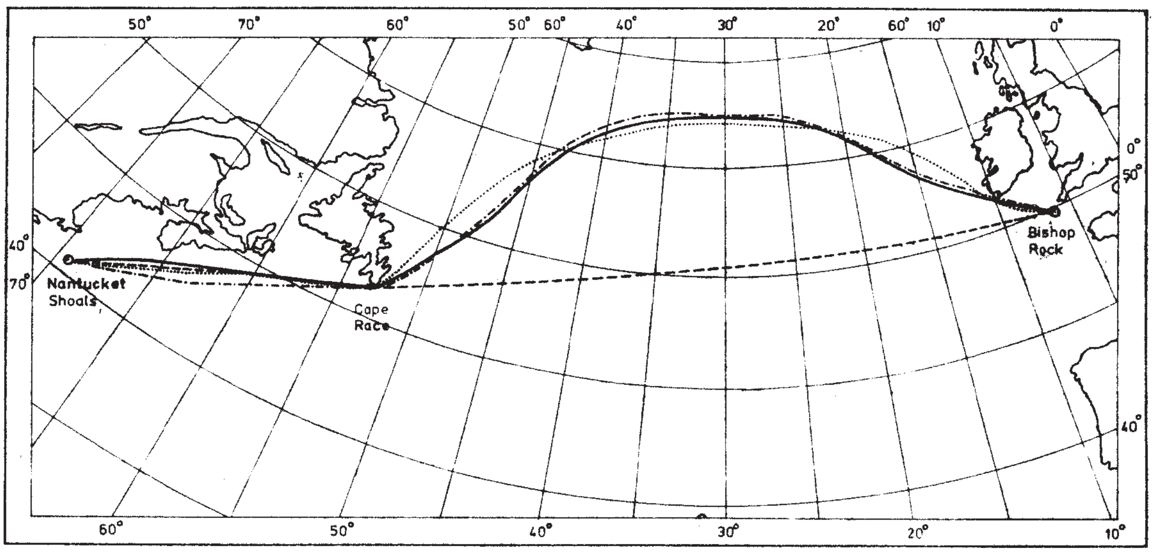

Fig. 5–Results of Both Methods (The ship departed from Bishop Rock on January 18, 1970, at 00.00 GMT. The service speed was 13 kn.)

successive search grids. This may be done by exhaustion or by using a shortest-path algorithm [12].

Two different options are then available. In the first, the inaccuracy in the optimal route calculation due to the coarse grid is taken for granted. In the second, the optimal route calculation is continued with the previous optimal route as a guess route, and the search grids perpendicular to it as well as the distance between them along this route are refined, leading to a better approximation of the optimal route. This procedure, which may be repeated a number of times until further refinement yields no substantial improvement [12, 13], emphasizes the continuous nature of the routing problem ([12], p. 20). A disadvantage of this iterative approach is that it could lead to convergence problems and to search grids that could overlap because of corners in the guess route ([13], p. 4). If the weather forecast is not valid during the complete passage, it might be completed with climatological data, and the above routines could then be applied.

Limited Predictability

The quality and range of the weather prediction have an important influence on the route calculation. Because of its importance, we pay particular attention to the predictability of large-scale atmospheric motions. The predictability associated with large-scale numerical weather prediction models can be measured in terms of the growth rate of small errors in the initial values of the model variables. In fact, it is the sensitivity to initial conditions that makes the system unpredictable.

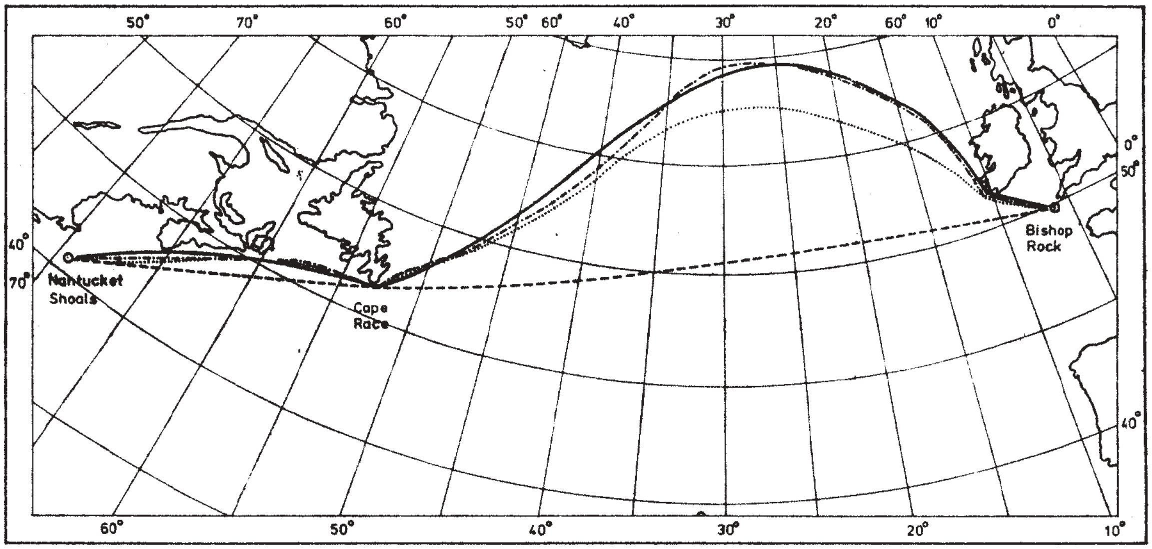

Fig. 6–Results of Both Methods (The ship departed from Bishop Rock on January 21, 1970, at 00.00 GMT. The service speed was 13 kn.)

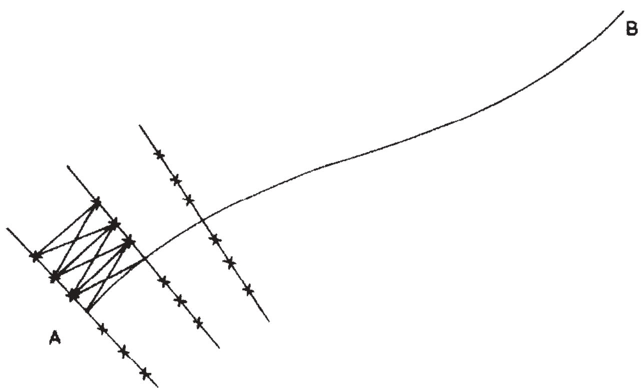

Fig. 7–First-Guess Route Between Points A and B and Part of Its Adjunctive Spatial Search Grids

The phenomena that determine the weather and therefore are extremely important for ship routing are formed by the coherent structures that are generated to adjust temperature differences in the north–south direction to maintain large-scale climatological circulation. These coherent structures (depressions), which have dimensions on the order of 1000 km, have a lifetime of approximately 5 days. The predictability of atmospheric circulation is closely related to the lifetime of these depressions. Once a depression has been created, its further development can be predicted rather accurately. Therefore, the range of prediction of the weather is about 5 days. At any moment, however, disturbances can arise that can negate the prediction of the course of a depression, depending on the sensitivity of the system to perturbations.

At the end of the last century, the weather forecasts provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) for 2 days ahead had a reliability of 95 percent, for 4 days ahead had a reliability of 90 percent, and for 5 days ahead had a reliability of 80 percent. The reliability of the 5-day forecast corresponds with that of the 2-day forecast in 1972.

The formation of a depression can be initiated by a coherent structure of much smaller dimensions that was formed to adjust differences in temperature on a smaller scale but is capable of creating motions on a much larger scale. Coherent structures of smaller dimensions have the capacity to adjust temperature differences much more rapidly than is the case for structures of larger dimensions. Therefore, they have a shorter lifetime. Prediction of these structures will result in a relatively small contribution to the range of predictability. For every resolution of the observation grid, however, there always remain sufficient small disturbances (coherent structures) that cannot be detected. Therefore, even if the larger scales could be observed perfectly, uncertainties at the smaller scales would after a day or so introduce errors at the larger scales comparable with the present larger-scale initial errors in the observed data [19].

Earlier estimations of the limits on the predictability of large-scale motion range from 10 days to more than a month, where the limit to predictability is considered to be the time required for the initial error to grow so that the model no longer gives useful weather information. Today it is not expected that the prediction range can be expanded substantially by refining the observation mesh and increasing the computer capabilities. Atmospheric predictability experiments with a large-scale numerical model [19] suggest that "predictions at least ten days ahead as skillful as predictions now made seven days ahead appear to be possible."

The sensitivity of the model to small variations in the initial conditions can be considered a measure of the reliability of the weather forecast. The ECMWF provides combined forecasts consisting of a control forecast and a collection of 50 forecasts with slightly different initial conditions [20]. These forecasts may be used in ship routing by producing a collection of optimal routes [13] and giving information on the most probable optimal route, a method that deserves further consideration.

CONCLUSIONS

In this paper, applications of optimal control theory and dynamic programming to optimal control problems in ship routing have been discussed. The connection between the two methods has been indicated for the case in which ship routing is treated as a continuous process, meaning that the sailing paths are not restricted to arcs of a grid as in the discrete dynamic programming method, but can vary in the navigation area. For this case, it is shown that the maximum principle of the optimal control theory can be derived from the principle of optimality, the application of which will yield the functional equations used in this continuous type of dynamic programming.

The method described in this paper (and more extensively in [8] and [9]), using optimal control theory, compares successfully for the case of minimal-time routing with a closely related computerized manual method [3], which can be considered a finite version of the continuous type of dynamic programming. A disadvantage of the latter method is related to the geometrical construction of normals to a time front, needed for the optimization procedure, a method that is more appropriate for manual use.

Optimal route calculation with the commonly used discrete approach of dynamic programming starts with the introduction of a guess route and the construction of a number of one-dimensional search grids perpendicular to that route. An optimal route connecting starting and end points is then computed by applying the principle of optimality to the points of the successive search grids. Clearly these grids are not attuned to the distances that can be covered by a ship in 6 or 12 h, usually the times at which weather information is provided and a new course is determined. Actually, a proper optimal route calculation using the discrete type of dynamic programming is an iterative method in which calculation starts on a coarse grid around a first-guess route and continues on increasingly refined grids around guess routes that are better and better approximations of the optimal route, emphasizing the continuous nature of ship routing ([12], p. 20). A disadvantage of this method is associated with convergence problems and overlapping search grids due to a corner in a guess route ([13], p. 4).

Summarizing, we may conclude that the application of optimal control theory as presented in this paper distinguishes itself from other methods by its clarity of visual presentation in showing all possible extremals emanating from the point of departure on their way to the point of arrival. This method is also more attuned to optimal control problems resulting from the meteorological navigation of ships than are the current applications of the commonly used discrete method of dynamic programming.

REFERENCES

1. Hansen, G. L. and R. W. James, *Optimum Ship Routing*, The Journal of Navigation, Vol. 13, No. 3, 1960, pp. 253–272.
2. Nagle, F. W., *Ship Routing by Numerical Means*, Publication NWRF 32-0361-042, U.S. Navy Weather Research Facility, Norfolk, VA, 1961.
3. de Wit, C., *Mathematical Treatment of Optimal Ocean Ship Routeing*, Ph.D. thesis, Delft University of Technology, The Netherlands, 1968.
4. Hagiwara, H., *Weather Routing of (Sail-Assisted) Motor Vessels*, Ph.D. thesis, Delft University of Technology, The Netherlands, 1989.
5. Klompstra, M. B., G. J. Olsder, and P. K. van Brunschot, *The Isopone Method in Optimal Control*, Dynamics and Control, Vol. 2, No. 3, 1992, pp. 281–301.
6. Faulkner, F. D., *Numerical Methods for Determining Optimum Ship Routes*, NAVIGATION, Journal of The Institute of Navigation, Vol. 10, No. 4, Winter 1963, pp. 351–367.
7. Bleick, W. E. and F. D. Faulkner, *Minimal-Time Ship Routing*, Journal of Applied Meteorology, Vol. 4, No. 2, 1965, pp. 217–221.
8. Bijlsma, S. J., *On Minimal-Time Ship Routing*, Mededelingen en Verhandelingen No. 94, Royal Netherlands Meteorological Institute, De Bilt (also Ph.D. thesis, Delft University of Technology), The Netherlands, 1975.
9. Bijlsma, S. J., *A Computational Method for the Solution of Optimal Control Problems in Ship Routing*, NAVIGATION, Journal of The Institute of Navigation, Vol. 48, No. 3, Fall 2001, pp. 145–154.
10. Zoppoli, R., *Minimum-Time Routing as an N-stage Decision Process*, Journal of Applied Meteorology, Vol. 11, No. 3, 1972, pp. 429–435.
11. Motte, R. H. and S. Calvert, *Operational Considerations and Constraints in Ship-Based Weather Routing Procedures*, The Journal of Navigation, Vol. 41, No. 3, 1988, pp. 417–433.
12. Braddock, R. D., *Optimal Problems in Physical Oceanography, Part I. Theory, Part III. Applications to Meteorological Navigation*, Research Paper No. 19, 21, Horace Lamb Centre for Oceanographical Research, Flinders University of South Australia, Bedford Park, South Australia, July 1968.
13. Hoffschildt, M., J-R. Bidlot, B. Hansen, and P. A. Janssen, *Potential Benefit of Ensemble Forecasts for Ship Routing*, Technical Memorandum No. 287, Research Department, European Centre for

Medium-Range Weather Forecasts, Reading, United Kingdom, August 1999.

14. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Pergamon Press, New York, 1964.
15. Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.
16. Halkin, H., *On the Necessary Condition for Optimal Control of Nonlinear Systems*, Journal d'Analyse Mathematique, Vol. XII Jérusalem, 1964.
17. Hestenes, M. R., *Calculus of Variations and Optimal Control Theory*, J. Wiley, Inc., New York, 1966.
18. Halkin, H., *The Principle of Optimal Evolution*, in Nonlinear Differential Equations and Nonlinear Mechanics, J. P. LaSalle and S. Lefschetz, eds., Academic Press, New York, 1963, pp. 284 – 302.
19. Lorenz, E. N., *Atmospheric Predictability Experiments with a Large Numerical Model*, Tellus, Vol. 34, No. 6, 1982, pp. 505 – 513.
20. Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, *The ECMWF Ensemble Prediction System: Methodology and Validation*, Quarterly Journal of the Royal Meteorological Society, Vol. 122, No. 529, 1996, pp. 73 – 120.