OPTIMAL TRACK SELECTION AND 3-DIMENSIONAL FLIGHT PLANNING

Theory and practice of the optimization problem in air navigation under space-time varying meteorological conditions

KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

MEDEDELINGEN EN VERHANDELINGEN

No. 93

H. M. DE JONG

OPTIMAL TRACK SELECTION AND 3-DIMENSIONAL FLIGHT PLANNING

Theory and practice of the optimization problem in air navigation under space-time varying meteorological conditions

1974

STAATSUITGEVERIJ/'S-GRAVENHAGE

PUBLIKATIENUMMER K.N.M.I. 102-93

U.D.C.: 551.509.322.7

PREFACE

In the series of scientific reports: 'Mededelingen en Verhandelingen' issued by the Royal Netherlands Meteorological Institute the author in 1956 presented a study on the problem of optimal track selection in air navigation, entitled: 'Theoretical aspects of Aeronavigation and its application in Aviation Meteorology' (nr. 64). The optimization problem was solved there using the classical tools of the calculus of variations. The study also presented an exposition of a simple prototype model algorithm for solving this problem. The corresponding graphical technique for the construction of the 'minimum flight path' found wide acceptance in operational practice.

With the advent of high speed electronic computers it was recognized that long distance flight planning was particularly suited to be automated. But this procedural change would demand for the analysis and synthesis of some computational algorithms, which would meet the special requirements of numerical mathematics in computer science. In the meantime it occurred that the computational methods used in the calculus of variation were generalized resulting in the foundation of the theory of optimal control processes.

It was interesting to know whether this new approach might contribute to a better insight and understanding of some intricate questions which had arisen in the course of time by experience with the manual dispatch technique of optimal track selection.

The intent of this report is to present in a unified manner the generalization of all the solutions known to date and some new formulations for solving the track selection problem considered as an optimal control problem under general space-time varying meteorological conditions and cruise performance.

Part I deals with the solution of the time optimal control problem in two dimensions (Least Time Track). Here the maximum principle of Pontryagin was a leading principle.

A special section is devoted to a modification of the governing control equations to be valid in a conformal map projection.

In Part II the theory is extended to cover the 3-dimensional case of track selection including the identification of the optimal vertical (stepped) altitude profile. Here the optimality criterium is examined in terms of different options: time-, fuel- and cost-saving. This 3-dimensional flight planning problem is solved by using some principles of modern graph theory.

Throughout this memorandum attention is paid to the numerical aspects of the various approaches within the scope of possible computer applications.

This work is the outcome of a co-operative project organized by the Royal Netherlands Meteorological Institute (K.N.M.I.) and the Royal Dutch Airlines (KLM) with the objective to study the potentialities of computerized flight planning for trans-atlantic crossings.

The Director in Chief of the

Royal Netherlands Meteorological Institute

M. W. F. SCHREGARDUS

VOORWOORD

In de serie 'Mededelingen en Verhandelingen', uitgegeven door het K.N.M.I., wijdde de schrijver in 1956 een studie aan het probleem van de selectie van de optimale vliegroute in de luchtvaartnavigatie, onder de titel: 'Theoretical aspects of Aeronavigation and its application in Aviation Meteorology' (nr. 64). Daarin werd voor het betreffende probleem een oplossing beschreven waarbij gebruik gemaakt werd van de klassieke methoden van de variatierekening.

In deze studie werd een uiteenzetting gegeven van een eenvoudige rekenmethode op basis waarvan een grafische techniek werd ontwikkeld om in de praktijk de optimale vliegroute te construeren.

Met de komst van snelle elektronische rekenmachines werd ingezien dat de vluchtvoorbereiding ten behoeve van het lange-afstandverkeer zich bijzonder goed leende voor automatisering. Maar deze verandering van werkmethode maakte de analyse en synthese van rekenprocessen noodzakelijk die moeten voldoen aan de specifieke eisen van elektronische informatieverwerking.

Intussen werden ook de grondbeginselen van de variatierekening herzien, hetgeen resulteerde in de ontwikkeling van de theorie van optimale controle processen.

In dit verband was het interessant te onderzoeken of deze nieuwe ontwikkeling zou kunnen bijdragen tot een beter inzicht in enkele moeilijke vraagstukken die uit de ervaringen met de grafische methode voor de dag waren gekomen.

Het doel van deze verhandeling is een samenvatting te geven van gegeneraliseerde versies van alle oplossingsvergelijkingen die tot op heden bekend zijn. Voorts worden enkele nieuwe formuleringen gepresenteerd die gevonden worden als men het probleem beschouwt als een besturingsprobleem, met inachtneming van de naar tijd en en plaats variërende meteorologische condities en vliegtuig-prestaties.

In deel I wordt de oplossing behandeld van het probleem in twee dimensies gebaseerd op de kortste vliegtijd. Een afzonderlijk hoofdstuk is gewijd aan het opstellen van het stelsel basisvergelijkingen dat geldt in een conforme kaartprojectie.

In deel II is de theorie uitgebreid voor het probleem van de optimale route selectie in de ruimte. De oplossing omvat o.a. de identificatie van het optimale hoogteprofiel. Daarbij wordt het criterium van een optimum beschouwd in verband met vliegtijd, brandstofverbruik en kosten. Dit probleem werd opgelost door gebruik te maken van enkele beginselen van de 'graph' theorie.

In de verhandeling is speciaal aandacht besteed aan de numerieke aspecten van de verschillende methodieken met het oog op computer-toepassingen.

De studie is het resultaat van een project dat tot stand is gekomen door samenwerking van het Koninklijk Nederlands Meteorologisch Instituut en de Koninklijke Luchtvaart Maatschappij (KLM). Dit project had tot doel de mogelijkheden na te gaan van de automatisering van de vluchtplanning voor het trans-atlantische luchtverkeer.

De Hoofddirecteur van het

Koninklijk Nederlands Meteorologisch Instituut

M. W. F. SCHREGARDUS

CONTENTS

I. INTRODUCTION

In aviation much effort is put into the development of graphical and numerical techniques for finding the most adventageous route, the adjective 'adventageous' in the sense of some specified optimization criterium for example minimum operating time.

Interest in this problem was raised shortly after world war II when regular trans-oceanic flights came into operation. In those days the long distance flights with piston-engined aircraft were often payload-critical: strong headwinds had an adverse effect on the economy of the flight. It was soon realized that the scale of the atmospheric circulation was such that one could allow for track diversions in the hope that the lengthening of the route would be overcompensated by favourable wind drifts. Those aircraft commanders challenging nature's forces mostly ventured to make excursions of 500 miles or more to circumnavigate whole depressions. The principle of 'pressure pattern flying' was born. As more reliable forecasts of the upper airflow became available this principle became a basic tool in practices and procedures of pre-flight planning. In the beginning the selection of the 'best' route was mere guess-work. The 'minimum flight path' was identified qualitatively and intuitively by visual inspection of the prevailing upper air circulation as depicted from the isobaric charts. The necessary aids to construct the desired track were still lacking at the time although some distinguished mathematicians had laid the foundation of the theory already in the thirties. [1], [2], [3], [4].

The problem of numerical evaluation of the optimum path remained to be settled. After some less successful attempts [5], [6] the break-through came in 1953 with the introduction of the time-front method as explained in the KLM-report. [7]. Since then most companies changed over from trial and error methods to the manual construction of the optimal route.

In the last decennium we observed the manual dispatch procedures gradually being superseded by numerical techniques using high speed electronic computers.

By chance this new strategy had its beginning almost simultaneously with the advent of regular turbo-jet operations. Computerized flight planning and track selection paved the way to rationalizing and centralizing existing practices, but above all to attacking problems which in the pre-computer era had been impracticably to solve by manual handling, for example '3-dimensional flight planning'.

This new trend stimulated the optimization problem to be reviewed in a wider perspective. It promoted also the analysis and synthesis of new algorithms which were manageable by computer within tolerable limits of running time, central core storage and costs.

The necessity to review current methods of optimal track selection and performance may be demonstrated by a brief discussion on the existing graphical construction method as described in [7]. See Fig. 1.

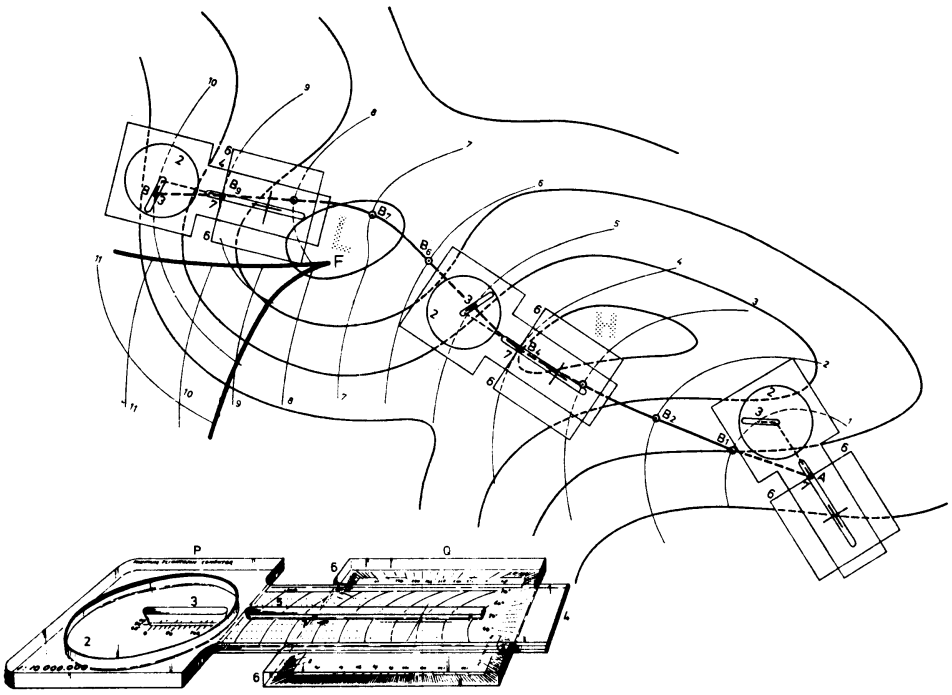

Fig. 1 Manual construction of the Least Time Track or 'minimum flight path' by means of the time front method. Note the phenomenon of the focussing effect near point F (focus). Insert: perspective drawing of the template used.

In this figure the fully drawn lines represent streamlines, which display the airflow in an isobaric surface. The stated problem is, given the aircraft speed (True Air Speed), to find the trajectory along which the time of flight between the point of departure A and destination B is an (absolute) minimum with respect to all admissible trajectories between A and B. The solution of this problem is analogous to Huyghen's principle of wave propagation in geometrical optics. The contruction where use is made of a special template, drawn in perspective in the insert of Fig. 1, evolves in two stages. From [7] we cite:

'a Time fronts.
Starting from point of departure A, ensure that in any given position of the template point 7 coincides with A. Point 7 is the setting of the average true air speed for the first hour, climb included. Rotate wind disk 2 so that groove 3 coincides with the mean wind direction for the first hour and plot in this groove the average wind speed (point 3).
Wind direction and wind speed are read from isohypses or streamlines and isotachs or spot winds in the forecast upper air chart in which the time fronts are to be constructed. Point 3 is a point of the first time front. In the same way mark a number of other points and draw through these points a fluent curve: the first time front 1. To find the next time front, move the device so that edge 6 is a tangent in point 7 to the time front in any given point; in other words: so that the true

heading is always perpendicular to the time front (point 7 is now set to the true air speed for cruise flight). In this position plot the wind vector with disk 2 in the same manner as before. Repeat in order to find other points of time front 2. According to the foregoing build up a system of time fronts until the system covers the point of destimation B.

b Minimum Flight Path.
In order to find the minimum flight path between A en B work in reversed order from point B, in a similar way as the time fronts were built up. From B, in this case located at the last time front, rotate wind disk 2 so that the average wind vector for the last hour can be plotted in opposite direction. Mark the endpoint of the opposite wind vector with a pencil and keep the pencil in this position. Now move the template so that the zero point of groove 3 coincides with the pencil point. Then rotate the template around this zero point (with pencil still in the same position) until edge 6 is tangent to time front 9 (to be achieved by moving part Q of the template along slide 4). This tangent point B_9, is a point of the minimum flight path from A to B. In the same manner, starting from point B_9, the other points B_8, B_7 etc. are constructed until point of departure A is reached.
If terminal B is not exactly located on the last time front, but between two time fronts, interpolation must be applied. From B plot a fraction of the wind vector, which is defined by the flight time between B and the last time front before B. If for instance B is located exactly halfway between the two time fronts, plot from B a distance equal to half the wind velocity and apply the construction accordingly.'

Note:
Part P of the template is adapted to the scale and projection of the weather chart on which the minimum flight path is constructed. Part Q is a protractor to be used for compiling the corresponding flight plan.

The synthesis of an algorithm which, for computer use, replaces the above described graphical construction principle, requires the reformulation of all steps in mathematical terms. As the method is a simple geometric one its elaboration looks straightforward. In the present dynamical system, however, there is a phenomenon which hinders the development of a suitable routine considerably. This concerns the occurrence of a *focussing effect*: the overcrossing of time fronts and the loss of uniqueness of the solution. (See time fronts 8, 9, ... 11 in Fig. 1). The occurrence of multiple solutions (relative extremes) is a reality in practice. Thus is it inevitable that the procedure should involve a subroutine to account for this focussing effect. Unfortunately, this is not easily to be solved numerically.

Obviously there are ample reasons to look for alternative solutions of the present problem statement and for more efficient and convenient process algorithms.

The present unconstrained optimization problem is a typical specimen of a problem which may be attacked using classical tools of variational calculus. The reference to Huyghen's principle suggests that the time front method be the result of the Hamiltonian point of view and is described by Hamilton's partial differential equation for this problem. [8]. Fundamental is also the differential equation of Euler Lagrange for this problem elaborated by Zermelo. [1]. The 'steering equation' of Zermelo lends itself extremely well for the application of an iterative procedure. It is also advisable to consider the feasibility of a numerical method based on Von Mises'

refraction law, which is equivalent with the law of refraction due to Snellius in optics. [3].

The navigational aspects may also be considered in the context of modern control theory. Here research is engaged with the dynamics of physical processes, which are influenced by some external controlling forces.

The variational methods used in these systems to solve optimal control conditions are of a non-classical nature, when the control domain is a closed set.

It is well known that this theory is centralized about the maximum principle of Pontryagin [9], [10]. Its strength lies in the elegant synthesis of different viewpoints, its applicability to non-linear dynamic systems (e.g. anisotropic inhomogeneous conditions of velocity control, time variability of the hydrodynamic flow etc.), and the power to solve the problem with a minimum of computational effort.

In aviation the aircraft's heading is a specific controlling entity, which indeed suggests that the optimization problem may be reviewed using the technique and methods of optimal control theory.

In this treatise (Part I) the theory is surveyed in a general way. (Sections 1, 2). This approach leads not only to a recapitulation and generalization of the existing methodology, but also to some new interpretations of the governing system equations.

A point of secondary importance, yet of vital interest in practice, is that the theory has been extended to cover the effect of the earth's curvature. (Section 4). It turns out that the system equations should be expanded to include corrective terms involving the scale of the (conformal) chart projection used.

Section 5 is devoted to the numerical aspects of some computation schemes developed to solve the optimal track selection problem by electronic means. Special attention is paid to the merits and drawbacks of the various approaches.

So far the problem was considered under the assumption of idealized conditions of airspace utilization. It was assumed that the aircraft were not deprived from operating along the most advantageous route due to various limitations imposed by ATC-centres of Providing States. These centres are responsible for an adequate organization of the traffic flow. There are but a few flight regions now, where the assignment of the most adventageous flight path is assured by these centres. The limitations force the theory of optimal track selection and performance to branch off into a direction of *constrained* optimization in a discrete system of route patterns, check points and flight levels.

In flight planning, where the ultimate goal is to achieve peak efficiency and to produce the optimum plan for each flight which will fulfil the company's policy of carrying the offered payload in the most economical manner under the safest possible conditions, this constrained optimization will result in a sub-optimum path selection only.

To attack discrete problems of this sort within the limitations of navigational capability, performance characteristics, flight control and separation criteria, one can

have recourse to the concepts of graph theory [11], [12], [13], in particular to the category of transport problems. (Section 3).

As the performance depends also on the flight profile in the vertical, the problem statement is wider in its scope than a quarter of a century ago.

Present techniques [14], [15] solve this problem by means of a two-stage processing, i.e. a separate track selection in the horizontal followed by a separate selection of the best altitude profile (stepped climb) in the vertical. In Part II we have made an attempt to tackle the problem in its most primitive form, namely in space-time using a special graph structure. In the last section some process algorithms are described, which may actually be programmed to produce the desired path- and performance information. The usefulness of these techniques is proved by experimentation.

PART I

THEORY AND PRACTICAL EVALUATION OF THE LEAST TIME TRACK

1. Optimal control problem. Review of Pontryagin's approach

1.1 Non-autonomous control processes

In this section we shall review Pontryagin's approach to the solution of optimal control problems for non-autonomous systems. Due to the complexity of the theory this review must be very descriptive and scanty. For more details, in particular for the proof of the wellknown maximum principle, the reader is referred to modern treatises on this subject. [9], [10], [17], [18].

We consider a non-linear dynamical system governed by the equations of motion of its state vector $\mathbf{x}$ in the Euclidian n-dimensional space R^n:

$$\frac{dx^i}{dt} = f^i(x^i, u^j, t), \quad i = 1, \ldots, n; j = 1, \ldots, r \tag{1.1}$$

or in vector form:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t). \tag{1.2}$$

Note:
Vectors are denoted by lower case bold-faced Latin letters, components by superscripts, matrices by capital letters.

Here $\mathbf{x}(x^1, \ldots, x^n)$ is the state vector in R^n. As usual $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt}$.

$\mathbf{u}(u^1, \ldots, u^r)$ is the *control vector* restricted to remain in the closed and bounded set U of dimension r. $\mathbf{u}(t)$ is supposed to be piecewise continuous. Any control function $\mathbf{u}(t)$ satisfying the above conditions is said to be admissible.

The constraints on the set U of admissible control variables $u^1, \ldots, u^r$ could be weakened, but they are adequate for the optimization problem in navigation to be reviewed in the next sections.

The vector valued function $\mathbf{f}(\mathbf{x}, \mathbf{u}, t)$ is assumed to have continuous first partial derivatives at the points of $R^n \times U \times I$, where I denotes the real time axis.

It is desired to find an admissible control law $\mathbf{u}(t)$ such that the system in R^n is carried from an initial state $\mathbf{x}(t_i) = \mathbf{x}_0$ to a final state $\mathbf{x}(t_f) = \mathbf{x}_1$ for which the functional

$$\mathcal{J} = \int_{t_i}^{t_f} f^0(\mathbf{x}(t'), \mathbf{u}(t'), t')\, dt' \tag{1.3}$$

taken along the desired trajectory is a minimum. f^0 is a specified objective function.

The functional $\mathcal{J}$ is associated with an optimality criterium for the specific problem under consideration e.g. minimum costs, minimum fuel expenditure, minimum time of transfer etc.

The time t_i is assumed to be given, t_f is the sought time at which the trajectory passes through the point $\mathbf{x}_1$ in R^n.

Next, we shall restate the problem by introducing an auxiliary state variable x^0 according to the relations:

$$\frac{dx^0}{dt} = f^0(\mathbf{x}, \mathbf{u}, t)$$

$$x^0(t_i) = 0. \tag{1.4}$$

Then the system of equations (1.2) may be replaced by the system in the augmented space R^{n+1}:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t), \quad i = 0, 1, \ldots, n \tag{1.5}$$

Here the new state vector $\mathbf{x}(x^0, x^1, \ldots, x^n)$ is the state vector in the augmented space R^{n+1}.

f is assumed to be independent of x^0.

Note:
In a more generalized form f may depend on all state variables $x^0, x^1, \ldots, x^n$. For instance, if the optimality criterium is based upon the fuel flow, the evolution of the flight depends on the fuel consumed at every intermediate instant of time.

The problem to be solved may now be formulated as follows:

In the augmented space R^{n+1} the point $\mathbf{x}(t_i) = (0, \mathbf{x}_0)$ is given and the line S parallel to the x^0-axis and passing through the point $(0, \mathbf{x}_1)$. Among all admissible controls $\mathbf{u} = \mathbf{u}(t)$ having the property that the corresponding solution $\mathbf{x}(t)$ of the system

$$\frac{dx^i}{dt} = f^i(\mathbf{x}, \mathbf{u}, t), \quad i = 0, 1, \ldots, n$$

with initial condition $\mathbf{x}(t_i) = (0, \mathbf{x}_0)$ intersects the line S, find one whose intersection with S has the smallest coordinate x^0.

In the context of Pontryagin's theory the solution to this problem is given by the maximum principle and a transversality condition.

To formulate the related theorems it is required to introduce a set of auxiliary variables $\lambda^0, \lambda^1, \ldots \lambda^n$ which satisfy the system of equations:

$$\frac{d\lambda^i}{dt} = -\sum_{j=0}^{n} \frac{\partial f^j}{\partial x^i}\lambda^j, \qquad j = 0, 1, \ldots, n \tag{1.6}$$

To the vector $\boldsymbol{\lambda}(\lambda^0, \lambda^1, \ldots, \lambda^n)$ are associated different names: adjoint vector, momentum vector, costate vector, Lagrange multipliers etc.

The set of equations (1.6) is known as the set of *adjoint variational equations.*

Next, we define the function $\mathscr{H}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\lambda})$ by the relation:

$$\mathscr{H}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\lambda}) = \boldsymbol{\lambda} . \mathbf{f}(\mathbf{x}, \mathbf{u}, t) = \sum_{0}^{n} \lambda^i f^i(\mathbf{x}, \mathbf{u}, t). \quad * \tag{1.7}$$

This function, known as *Pontryagin's quasi-Hamiltonian,* becomes a function of the vector parameter $\mathbf{u}$ for fixed values of $\mathbf{x}$, t and $\boldsymbol{\lambda}$. Let us denote the least upper bound of the values of this function by $\mathscr{M}(\mathbf{x}, t, \boldsymbol{\lambda})$:

$$\mathscr{M}(\mathbf{x}, t, \boldsymbol{\lambda}) = \sup_{\mathbf{u} \in U} \mathscr{H}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\lambda}). \tag{1.8}$$

Then we obtain the following theorem (maximum principle for non-autonomous systems).

Let $\mathbf{u}(t)$, $t_i \leqslant t \leqslant t_f$ be an admissible control along the corresponding trajectory $\mathbf{x}(t)$ of system (1.5). In order that $\mathbf{u}(t)$ and $\mathbf{x}(t)$ be optimal it is required that there exist a non-zero continuous solution $\boldsymbol{\lambda}(t)$ of the adjoint variational equations

$$\frac{d\lambda^i}{dt} = -\sum_{j=0}^{n} \frac{\partial f^j}{\partial x^i}\lambda^j, \qquad i = 0, 1, \ldots, n$$

such that

a for all t the function $\mathscr{H}(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\lambda}(t))$ of the variable $\mathbf{u} \in U$ attains its maximum at the point $\mathbf{u} = \mathbf{u}(t)$:

$$\mathscr{H}(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\lambda}(t)) = \mathscr{M}(\mathbf{x}(t), t, \boldsymbol{\lambda}(t)).$$

b for $t = t_f$ in the terminal point

* By convention the notation $\mathbf{a} . \mathbf{b}$ indicates the scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$.

$$\mathcal{M}(\mathbf{x}(t_f), t_f, \boldsymbol{\lambda}(t_f)) = 0. \tag{1.9}$$

The condition b is related to the transversality condition at the open end-point, where t_f is not known in advance.

The adjoint vector obtains a physical meaning if the control problem is conceived as a topological (geometric) problem. In fact, the optimization problem may also be solved by analogy with the *wave propagation in optics* (generalized Huyghen's principle).

The propagation analogy has been studied thoroughly by Halkin [16] applying the 'principle of optimal evolution', which states that: 'every event of an optimal process belongs to the boundary of the set of possible events.'

By defining the set $V(t)$ as the set of all states in R^{n+1} that are reachable at the time t, i.e.

$$V(t) = \{\mathbf{x}(t\,;\mathbf{u}), \mathbf{u} \in U\} \tag{1.10}$$

the principle of optimal evolution states that if $\mathbf{u}_0(t)$ is an *optimal* control function, then for every $t \in [t_i, t_f]$ the state $\mathbf{x}(t\,;\mathbf{u}_0)$ belongs to the boundary $\partial V(t)$ of the set $V(t)$. Then the adjoint vector $\boldsymbol{\lambda}(t)$ is identical to the outward normal to the 'wave front' $\partial V(t)$ at the point $\mathbf{x}(t\,;\mathbf{u}_0)$. The length of this vector has not yet been defined up to a multiplicative factor, but the knowledge of the length of the vector for some $t \in [t_i, t_f]$ will be sufficient to derive the length of $\boldsymbol{\lambda}(t)$ for all t in the interval.

As in geometric optics we can define the velocity of the wave front $\partial V(t)$ at the point $\mathbf{x}(t\,;\mathbf{u}_0)$. This wave front velocity is parallel to the vector $\boldsymbol{\lambda}(t)$. Denoting the speed of the wave front by $s(t)$ we have

$$\frac{\mathbf{f}(\mathbf{x}(t\,;\mathbf{u}_0), \mathbf{u}_0(t), t) \,.\, \boldsymbol{\lambda}(t)}{|\,\boldsymbol{\lambda}(t)\,|} = s(t).$$

Solving the control problem in analogy with the wave propagation in optics it has been shown that every trajectory belonging to the boundary of the set of possible events satisfies the maximum principle.

The procedure to tackle a practical problem is first to determine the optimal control $\mathbf{u}_0$ as a function of the state- and adjoint variables from the maximum principle.

Thereafter $\mathbf{u}_0$ is entered into Pontryagin's quasi-Hamiltonian $\mathcal{H}$ giving the *classical* Hamiltonian $\mathcal{H}^0$:

$$\mathcal{H}^0(\mathbf{x}, t, \boldsymbol{\lambda}) = \mathcal{H}(\mathbf{x}, \mathbf{u}_0(\mathbf{x}, t, \boldsymbol{\lambda}), t, \boldsymbol{\lambda}). \tag{1.11}$$

Then the system equations (1.5) and (1.6) must be solved for the dynamics associ-

ated with the optimal trajectory. With the aid of $\mathscr{H}^0$ these equations may be written in the form

$$\frac{dx^i}{dt} = \frac{\partial \mathscr{H}^0}{\partial \lambda^i},$$

$$\frac{d\lambda^i}{dt} = -\frac{\partial \mathscr{H}^0}{\partial x^i}, \quad i = 0, 1, \ldots, n \tag{1.12}$$

This set of equations constitutes the wellknown *canonical system of Hamilton.*

Given the appropriate end conditions this set will describe how the state vector and adjoint vector vary along the desired optimal track.

From a point of view of engineering applications considerable progress has been made to develop suitable computing techniques for solving the optimization problems, but it must be admitted that the solution of specific boundary value problems so far are still way behind the manipulations of these formulas. This aspect will be discussed further, when we shall study the synthesis of optimal processes in meteorological navigation.

For the sake of completeness it is worth while to call attention to the formulation of a partial differential equation for the objective function V, cf. (1.10), known as the *dartial differential equation of Hamilton-Jacobi.*

Halkin [16] has shown that for the class of problems of classical calculus of variations which can be restated as optimal control problems, the Hamiltonian $\mathscr{H}^0$ may be used to actually find this Hamiltonian-Jacobi equation.

Let

$$V(\mathbf{x}, t) = 0 \tag{1.13}$$

be the equation defining the boundary of reachable events $V(t)$, which is a hypersurface in R^{n+1}.

Then the equation of Hamilton-Jacobi is obtained by putting

$$\frac{\partial V(\mathbf{x}, t)}{\partial t} + \mathscr{H}^0(\mathbf{x}, t, \boldsymbol{\lambda})_{\lambda^i = \frac{\partial V(\mathbf{x},t)}{\partial x^i}} = 0. \tag{1.14}$$

The Hamiltonian point of view and its interpretation in operational practice have proved to be of particular value in regard to the graphical evaluation of the least time track in air navigation. [7], [8].

Its importance lies in the synthesis of the problem as distinct from an analysis of the control to be maintained along an individual optimal trajectory.

1.2 Time optimal control

From the theorem for general non-autonomous optimal control processes we shall derive the necessary conditions for time optimality.

This special case will be the subject of the time optimal navigation problem to be discussed in later sections.

Let us start with the equation defining an isochronal hypersurface:

$$V(\mathbf{x}, t) = 0,\ \mathbf{x} = \mathbf{x}(x^0, x^1, \ldots, x^n).$$

For problems in variation calculus restated as optimal control processes, $V(\mathbf{x}, t) = 0$ can always be written in the form:

$$\tau(\mathbf{x}, t) - x^0 = 0,\ \mathbf{x} = \mathbf{x}(x^1, \ldots, x^n). \tag{1.15}$$

Here τ denotes the time of transfer.

Let λ be an outward normal to the isochronal hypersurface in R^{n+1}, then except for a multiplicative factor

$$\lambda = (-1, \frac{\partial \tau}{\partial x^1}, \ldots, \frac{\partial \tau}{\partial x^n}). \tag{1.16}$$

Since for time optimality

$$f^0(\mathbf{x}, \mathbf{u}, t) \equiv 1. \tag{1.17}$$

Pontryagin's quasi-Hamiltonian will take the form:

$$\mathscr{H} = -1 + \frac{\partial \tau}{\partial x^1} f^1(\mathbf{x}, \mathbf{u}, t) + \ldots \frac{\partial \tau}{\partial x^n} f^n(\mathbf{x}, \mathbf{u}, t). \tag{1.18}$$

The maximum in terms of the control $\mathbf{u}$ is

$$\mathscr{M} = -1 + \sup_{\mathbf{u} \in U} (\nabla \tau . \mathbf{f}).$$

The maximum principle may be given the following physical meaning: the control law $\mathbf{u} = \mathbf{u}(t)$ should be chosen such that at any point of the trajectory the component of the system state velocity in R^n along the gradient of the optimal isochrones is always a maximum. (Fig. 2.)

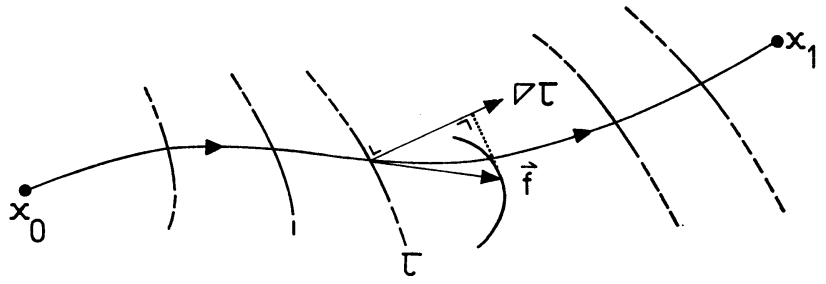

Fig. 2 Illustrating Pontryagin's Maximum Principle in case of time optimal control.

The condition that $\mathcal{M}$ should vanish at the terminal point is fulfilled as $(\nabla\tau \,.\, \mathbf{f})$ represents the rate at which time increases along the trajectory, hence should be equal to unity. The same holds for every $t \in [t_i, t_f]$.

The maximum principle delivers an optimal control $\mathbf{u}_0$ as a function of $\mathbf{x}$, t and $\frac{\partial\tau}{\partial x^i}$.

Substitution of the Hamiltonian $\mathcal{H}$ for $\mathbf{u}_0$ yields the classical Hamiltonian $\mathcal{H}^0$. The canonical system of Hamilton associated with $\mathcal{H}^0$ becomes:

$$\frac{dx^i}{dt} = f^i(\mathbf{x}, \mathbf{u}_0, t),$$

$$\frac{d}{dt}\left(\frac{\partial\tau}{\partial x^i}\right) = -\sum_{i=1}^{n} f^i(\mathbf{x}, \mathbf{u}_0, t)\,\frac{\partial\tau}{\partial x^i}, \qquad i = 1, \ldots, n. \tag{1.19}$$

The application, in particular of the second set of equations (1.19), to the navigation problem will result into the formulation of some control laws different from the well-known steering equation of Zermelo.

The partial differential equation of Hamilton-Jacobi takes the form, cf. (1.14)

$$\frac{\partial\tau}{\partial t} + \mathcal{H}^0(\mathbf{x}, \mathbf{u}_0(\mathbf{x}, t, \frac{\partial\tau}{\partial x^i}), t) = 0$$

or referring to (1.18):

$$\frac{\partial\tau}{\partial t} + \sum_{1}^{n} \frac{\partial\tau}{\partial x^i}\, f^i\,(\mathbf{x}, \mathbf{u}_0\,(\mathbf{x},\, t,\, \frac{\partial\tau}{\partial x_i}),\, t) = 1. \tag{1.19}$$

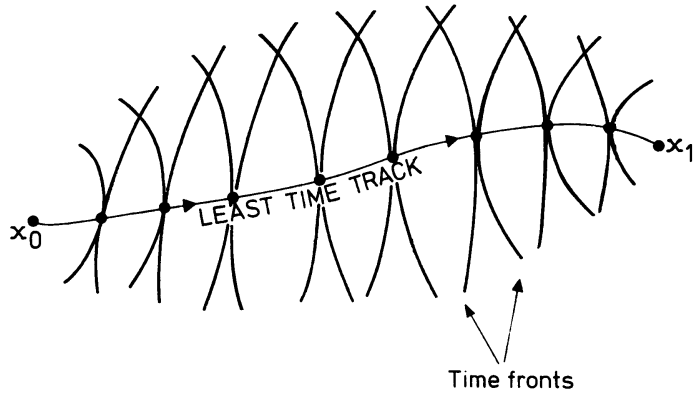

Fig. 3 Configuration of 'complementary' time fronts and Least Time Track.

The solutions of this Hamilton-Jacobi equation embrace for instance the set of isochronals or time fronts, which, given the time of departure, propagate from a given location and overtake the terminal point in the least possible time.

By the same token a solution may also define the set of isochronals which propagate towards the terminal point after having bypassed the departure point at the prescribed initial time.

It is interesting to note that the optimal trajectory between starting point and destination appears to be the locus of tangent points of isochrones appertaining to both these configurations. (Fig. 3.)

When an isochronal hyper-surface becomes non-differentiable at a point of an optimal track the application of the generalized Huyghen's principle may lead to the possibility of *branching* of optimal trajectories. Then often the phenomenon of generation of *caustica* may be observed. The structure of the optimal tracks will then depend on the particular problem considered.

This item will be the subject of a more detailed analysis when dealing with the synthesis of the time optimal control in air navigation, together with some examples of practical applications and computational techniques.

2. Time optimal control in air navigation

2.1 General

We shall now direct our attention to the navigation problem in aviation.

Consider a long range flight within a non-stationary atmospheric airflow to determine the flight track which is flown in the least possible time between the points of departure and destination.

It is assumed that the flight takes place on a constant pressure level with variable air speed.

The problem as stated is typically a time optimal control problem in 2-dimensional space.

The desired track is sometimes referred to as *Minimum Flight Path*, but the denomination *Least Time Track* is more in conformity with the character of the problem.

In mathematical terms the motion of the airplane is governed by the differential equation:

$$\dot{\mathbf{x}} = \mathbf{w} + \mathbf{c} \tag{2.1}$$

where $\mathbf{x}(x^1, x^2)$ is the position vector with respect to a Cartesian coordinate system OX_1X_2; $\mathbf{w}(w^1, w^2)$ represents the wind vector and $\mathbf{c}(c^1, c^2)$ the (controlling) airspeed vector.

Both the wind vector and airspeed vector may vary with position and time, so that the equation of motion (2.1) may be written:

$$\dot{x}^1 = w^1(x^1, x^2, t) + c\,(x^1, x^2, t)\cos\xi$$
$$\dot{x}^2 = w^2(x^1, x^2, t) + c\,(x^1, x^2, t)\sin\xi$$

where

$$c = \sqrt{(c^1)^2 + (c^2)^2} \tag{2.2}$$

ξ denotes the *heading* or steering angle, i.e. the angle subtended by the airspeed vector $\mathbf{c}$ and the x^1-direction. (Fig. 4.)

Whereas in navigation the heading is taken as increasing in a clockwise direction, we shall follow the usual convention in mathematics and consider ξ as increasing in anti-clockwise direction.

Note:

In some flight regions e.g. the North Atlantic Region where the Air Traffic Services have no other suitable means of ensuring the separation between successive aircraft a type of navigation has been introduced known as the *Mach Number Technique*. This term is used to describe a technique where-

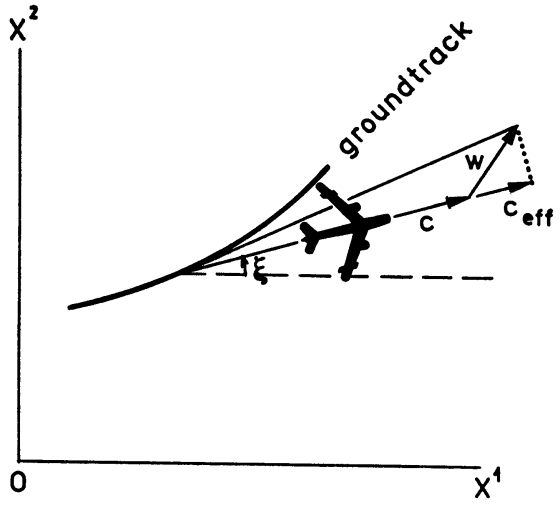

Fig. 4 Coordinate frame used and control of flight in a wind environment.

by turbo jet aircraft are cleared by ATC to maintain appropriate Mach numbers for the en-route phase of the flight. The aircraft are required to adhere to the prescribed Mach number within close tolerances in order to maintain longitudinal separation between successive aircraft on the same route. The application of this technique implies that the airspeed will vary with ambient temperature, as the Mach number by definition is the ratio of the airspeed to the local speed of sound and this speed of sound at a constant pressure altitude is known to be proportional to the square root of absolute temperature. Since the atmospheric temperature changes with time and position it follows that the utilization of the Mach Number Technique introduces a position and time variable airspeed $c(x^1, x^2, t)$.

The heading ξ functions as the controlling parameter. Decomposing $\mathbf{c}$ in polar coordinates c and ξ, the purpose of the time optimal control is to find an optimal navigation function for $\xi = \xi(t)$, which directs the aircraft along the Least Time Track. $\mathbf{c}$ therefore acts indirectly as a controlling entity. In the present navigation problem $\mathbf{c}$ is allowed at a given location to point in any direction so that the set U of the admissible control variable ξ is an *open set*. This implies that Pontryagin's approach merely replaces the computational tools of the classical variational calculus.

If not otherwise stated we assume that the airspeed is not surpassed by the windspeed in the whole region, so that the manoeuvrability determined by the sum vector $\mathbf{w} + \mathbf{c}$ be warranted locally in all directions.

There is a quantity which is time and again to return in the present analysis of the control problem. (Fig. 4.)

This quantity, the *effective airspeed* c_{eff}, is the algebraic sum of the airspeed and tail- or headwind:

$$c_{\text{eff}} = c + \frac{1}{c}(\mathbf{c}^{\mathbf{T}} . \mathbf{w}).* \tag{2.3}$$

Since $c > |\mathbf{w}|$ the effective airspeed is always positive. The vector

$$\mathbf{c}_{\text{eff}} = (1 + \frac{1}{c^2}(\mathbf{c}^{\mathbf{T}} . \mathbf{w}))\mathbf{c}, \tag{2.4}$$

which is directed along $\mathbf{c}$ with length c_{eff} will also play an important role.

To solve the 2-dimensional navigation problem we shall follow Pontryagin's approach as described in the previous section.

Here the equations of motion (1.5) read:

$$\begin{aligned} \dot{x}^1 &= f^1(x^1, x^2, t) = w^1(x^1, x^2, t) + c^1(x^1, x^2, t), \\ \dot{x}^2 &= f^2(x^1, x^2, t) = w^2(x^1, x^2, t) + c^2(x^1, x^2, t), \\ \dot{x}^0 &= f^0(x^1, x^2, t) = 1. \end{aligned}$$

subject to the end conditions:

$$\begin{aligned} x^1(t_i) = x_1^1,\ x^2(t_i) = x_1^2 \\ x^1(t_f) = x_2^1,\ x^2(t_f) = x_2^2 \\ x^0(t_1) = 0 \end{aligned} \tag{2.5}$$

The objective is to find a control function $\mathbf{c}(t)$ such that $x^0(t_f)$is a minimum.

Introducing the adjoint vector

$$\boldsymbol{\lambda} = (-1, \frac{\partial \tau}{\partial x^1}, \frac{\partial \tau}{\partial x^2}), \tag{2.6}$$

where $\tau(x^1, x^2)$ denotes the minimum time going from a point $\mathbf{x}_1(x^1(t_i), x^2(t_i))$ towards a point $\mathbf{x}(x^1(t), x^2(t))$ along an optimal trajectory, we can define Pontryagin's quasi-Hamiltonian $\mathscr{H}$:

$$\begin{aligned} \mathscr{H} = \sum_0^2 \lambda^i f^i &= -1 + \lambda^1 w^1 + \lambda^2 w^2 + \lambda^1 c^1 + \lambda^2 c^2 \\ &= -1 + \boldsymbol{\lambda}^{\mathbf{T}} . \mathbf{w} + \boldsymbol{\lambda}^{\mathbf{T}} . \mathbf{c}, \end{aligned} \tag{2.7}$$

* In matrix notation a vector is represented by a column matrix. *T* denotes transposition of a matrix. $\mathbf{c}^{\mathbf{T}}$ for instance, is a row matrix.

where it is understood that $\lambda(\lambda^1, \lambda^2)$ from now on is the outward normal to an isochrone or timefront in R^2.

According to the maximum principle a necessary condition for obtaining a speediest flight is that at every point of the Least Time Track the scalar product

$$\lambda^{\mathrm{T}} . \mathbf{c}$$

be a maximum.

In words: *at every point of the Least Time Track the airspeed vector should be directed along the outward normal to the isochrones.* (Fig. 5.)

In mathematical terms:

$$\mathbf{c}_0 = \frac{c}{\lambda}\lambda, \tag{2.8}$$

where

$$\lambda = |\lambda| = \sqrt{(\lambda^1)^2 + (\lambda^2)^2}.$$

Substitution of $\mathscr{H}$ (2.7) for the control condition (2.8) gives:

$$\mathscr{H}^0 = -1 + \lambda^1 w^1 + \lambda^2 w^2 + c\sqrt{(\lambda^1)^1 + (\lambda^2)^2}. \tag{2.9}$$

In case of time optimal control a second requirement is that $\mathscr{H}^0$ should vanish at the open endpoint $(x^1(t_f), x^2(t_f))$:

$$\lambda^{\mathrm{T}} . \mathbf{w} + \lambda^{\mathrm{T}} . \mathbf{c}_0 = 1.$$

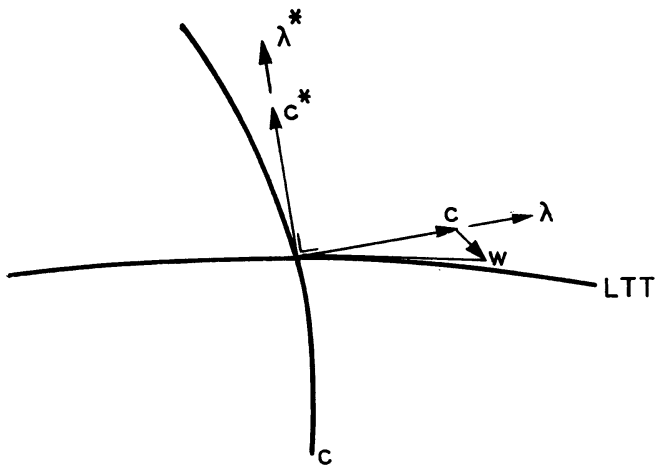

Fig. 5 Pontryagin's Maximum Principle along the Least Time Track.

Entering (2.8) into this condition yields:

$$\frac{\lambda}{c}\,\mathbf{c}_0^{\mathrm{T}}\,(\mathbf{w}+\mathbf{c}_0)=1.$$

Therefore, referring to (2.3):

$$\lambda=\frac{1}{c_{\mathrm{eff}}}. \tag{2.10}$$

As λ equals $\nabla\tau$, up to a multiplicative factor to be taken unity here, we have also

$$|\nabla\tau|=\frac{1}{c_{\mathrm{eff}}}. \tag{2.11}$$

which states that along the Least Time Track *the phase speed of isochronals is equal to the effective airspeed.*

These characteristic properties concerning an effective control along the time optimal trajectory within a given atmospheric wind regime are well-known. [8]. They could have been derived using classical methods of variational calculus. The advantage of Pontryagin's approach is, however, that these properties may be found without solving the entire problem.

2.2 The gradient equation

Although the geometric properties as derived in the previous section are adequate for a synthesis of the problem for instance by a 'wave front' construction analogous to Huyghen's principle, they are not suited to describe the evolution of an individual optimal flight.

For that purpose it is desired to establish the equations of state for the adjoint vector $\boldsymbol{\lambda}$, the gradient vector $\nabla\tau$ and the position vector $\mathbf{x}$. These equations may directly afford the required control function and resulting motion.

Consider the second set of the canonical system of Hamilton (1.12):

$$\dot{\boldsymbol{\lambda}}=-\frac{\partial\mathscr{H}^0}{\partial\mathbf{x}}.$$

From (2.9) we derive:

$$\begin{aligned}\dot{\lambda}^1&=-\lambda^1\frac{\partial w^1}{\partial x^1}-\lambda^2\frac{\partial w^2}{\partial x^1}-\lambda\frac{\partial c}{\partial x^1},\\ \dot{\lambda}^2&=-\lambda^1\frac{\partial w^1}{\partial x^2}-\lambda^2\frac{\partial w^2}{\partial x^2}-\lambda\frac{\partial c}{\partial x^2}.\end{aligned} \tag{2.12}$$

Let the matrix W:

$$W = \begin{bmatrix} \dfrac{\partial w^1}{\partial x^1} & \dfrac{\partial w^1}{\partial x^2} \\ \dfrac{\partial w^2}{\partial x^1} & \dfrac{\partial w^2}{\partial x^2} \end{bmatrix}$$

be the Jacobian of the wind vector, then the set of equations (2.12) may be written in vector notation:

$$\dot{\boldsymbol{\lambda}} = -W^T \boldsymbol{\lambda} - \boldsymbol{\lambda} \nabla c. \tag{2.13}$$

This Euler equation for the adjoint vector $\boldsymbol{\lambda}$ may be put in a more convenient form by introducing the vector

$$\mathbf{n} = \frac{1}{c} W^T \mathbf{c} + \nabla c. \tag{2.14}$$

Then, in view of (2.8) and (2.10)

$$\dot{\boldsymbol{\lambda}} = -\frac{1}{c_{\text{eff}}} \mathbf{n} \tag{2.15}$$

or

$$\dot{\nabla}\tau = -\frac{1}{c_{\text{eff}}} \mathbf{n}. \tag{2.16}$$

This vector differential equation for the gradient of isochrones, $\boldsymbol{\lambda} = \nabla\tau$, will be referred to as 'gradient equation'.

The synthesis of the navigation problem would be incomplete without an examination of the first set of the canonical system (1.12):

$$\dot{\mathbf{x}} = \frac{\partial \mathscr{H}^0}{\partial \boldsymbol{\lambda}}.$$

From (2.9) we obtain:

$$\dot{x}^1 = w^1 + c\frac{\lambda^1}{\lambda},$$

$$\dot{x}^2 = w^2 + c\frac{\lambda^2}{\lambda}.$$

Referring to (2.8) we get

$$\cos \xi^0 = \frac{\lambda^1}{\lambda}, \sin \xi^0 = \frac{\lambda^2}{\lambda},$$

where ξ^0 is the *optimal* steering angle, so that with

$$\mathbf{c}_0 = (c \cos \xi^0, c \sin \xi^0),$$

$$\dot{\mathbf{x}} = \mathbf{w} + \mathbf{c}_0 \tag{2.17}$$

Summarizing the results we may observe that the canonical system, in particular the gradient equation (2.16) and state equation (2.17), afford the solution to the time optimal navigation problem.

The canonical system offers at least in principle the necessary means to evaluate the desired parameters for the optimum flight path. It would indeed be very welcome if we could first find a solution of the gradient equation, and by substitution of the optimal control $\mathbf{c}_0(t)$ into the equations of motion, could finalize the 'construction' of the definite track. There is however an interdependency between the adjoint and state variables in the canonical system and the question of finding the optimal state of the adjoint vector for arbitrary systems under general conditions of initial and final state has never been completely answered. Nevertheless this structure of the canonical system does strongly suggest that some useful computation algorithms may be developed to explore appropriate searching techniques by means of iterative procedures.

The previous discussion has shown the significance of the adjoint system. Its controlling properties will become more clear when we attempt to reformulate the gradient equation with respect to the steering angle, phase speed, etc.

2.3 The steering equation

The application of efficient cruise control procedures in practical operations requires the formulation of an appropriate control law for the heading.

In the previous section it was shown that the time optimal control is governed by the set of adjoint equations (2.15)-(2.16).

In order to derive a vector equation for the control function $\xi = \xi(t)$ it is necessary to find another version of the adjoint equation.

Let the adjoint vector $\boldsymbol{\lambda}$ be decomposed in polar coordinates:

$$\boldsymbol{\lambda} = (\lambda, \xi).$$

Then we have

$$\dot{\lambda}^1 = \frac{d}{dt}(\lambda \cos \xi) = \dot{\lambda} \cos \xi - \lambda \sin \xi \, \dot{\xi},$$

$$\dot{\lambda}^2 = \frac{d}{dt}(\lambda \sin \xi) = \dot{\lambda} \sin \xi + \lambda \cos \xi \, \dot{\xi}. \tag{2.18}$$

Let $\boldsymbol{\lambda}^*(-\lambda^2, \lambda^1)$ and $\mathbf{c}^*(-c^2, c^1)$ represent the vector $\boldsymbol{\lambda}$ and $\mathbf{c}$ rotated over an angle $\pi/2$ in anti-clockwise direction (Fig. 5), (tangent to an isochrone).

Then (2.18) may be written:

$$\dot{\boldsymbol{\lambda}} = \frac{\dot{\lambda}}{\lambda}\boldsymbol{\lambda} + \dot{\xi}\boldsymbol{\lambda}^*.$$

The adjoint equation (2.13) takes the form:

$$\dot{\lambda}\boldsymbol{\lambda} + \lambda\dot{\xi}\boldsymbol{\lambda}^* = -\lambda(W^T\boldsymbol{\lambda} + \lambda\nabla c). \tag{2.19}$$

Next take the scalar product of (2.19) with $\boldsymbol{\lambda}^*$:

$$\dot{\lambda}\boldsymbol{\lambda}^{*\mathbf{T}}.\boldsymbol{\lambda} + \lambda\dot{\xi}\boldsymbol{\lambda}^{*\mathbf{T}}.\boldsymbol{\lambda} = -\lambda(\boldsymbol{\lambda}^{*\mathbf{T}}W^T\boldsymbol{\lambda} + \lambda\boldsymbol{\lambda}^{*\mathbf{T}}.\nabla c).$$

Noting that

$$\mathbf{c} = \frac{c}{\lambda}\boldsymbol{\lambda}, \quad \mathbf{c}^* = \frac{c}{\lambda}\boldsymbol{\lambda}^*$$

$$\boldsymbol{\lambda}^{\mathbf{T}}.\boldsymbol{\lambda} = \boldsymbol{\lambda}^{*\mathbf{T}}.\boldsymbol{\lambda}^* = \lambda^2, \; \boldsymbol{\lambda}^{*\mathbf{T}}.\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathbf{T}}.\boldsymbol{\lambda}^* = 0 \tag{2.20}$$

we obtain:

$$\dot{\xi} = -\frac{1}{c^2}\mathbf{c}^{*\mathbf{T}}(W^T\mathbf{c} + c\nabla c) \tag{2.21}$$

or, using (2.14):

$$\dot{\xi} = -\frac{1}{c}\mathbf{c}^{*\mathbf{T}}.\mathbf{n}. \tag{2.22}$$

A result which expresses the required rate of change of the steering angle in terms of the local structure of the windfield and airspeed.

It can easily be shown that the control equation turns out to be a generalized form of the well-known *steering equation* of *Zermelo*.

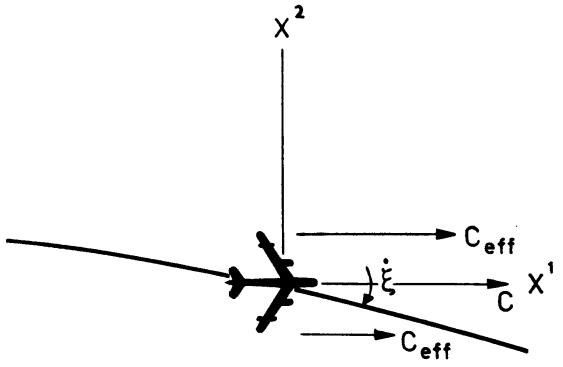

Fig. 6 Interpretation of Zermelo's steering equation.

Indeed, substituting (2.21) for **c** ($c \cos \xi$, $c \sin \xi$) and **c*** ($-c \sin \xi$, $c \cos \xi$) affords:

$$\dot{\xi} = -\frac{\partial w^1}{\partial x^2} \cos^2 \xi + \left(\frac{\partial w^1}{\partial x^1} - \frac{\partial w^2}{\partial x^2}\right) \sin \xi \cos \xi + \frac{\partial w^2}{\partial x^1} \sin^2 \xi + \frac{\partial c}{\partial x^1} \sin \xi - \frac{\partial c}{\partial x^2} \cos \xi. \tag{2.23}$$

Except for the contribution of the variable airspeed this Euler equation for the heading is equivalent to Zermelo's solution to the problem [8]. Zermelo announced this solution in a lecture delivered at Prague in 1930: 'Über die Navigation in der Luft als Problem der Variationsrechnung'.

A further exposition was given in a later paper [1]. Whereas Zermelo used highly original methods, Levi Cività [2] derived the same equation along more orthodox lines, analogous to those used in the theory of geometric optics.

To bring the steering equation in a more perceptible form, we may choose a rectangular coordinate system at a given point of the optimal trajectory such that the x^1-axis points into the direction of the heading. (Fig. 6.)

Then (2.23) reduces to

$$\frac{d\xi}{dt} = -\frac{\partial w^1}{\partial x^2} - \frac{\partial c}{\partial x^2}$$

or

$$\frac{d\xi}{dt} = -\frac{\partial c_{\text{eff}}}{\partial x^2}. \tag{2.24}$$

indicating that at every point on the optimal track the *rate of change of the heading*

is equal to the negative shear of the effective airspeed normal to the heading.

It should be borne in mind that the steering equation (2.22) holds for flights within non-stationary windfields and time and location variable airspeed.

2.4 Phase speed equation

We have seen that the steering equation is one of the component equations associated with the adjoint vector equation (2.13).

To obtain the second component equation we compute the scalar product of (2.19) with $\boldsymbol{\lambda}$:

$$\dot{\lambda}\boldsymbol{\lambda}^{\mathbf{T}}.\boldsymbol{\lambda} + \lambda\dot{\boldsymbol{\xi}}\boldsymbol{\lambda}^{\mathbf{T}}.\boldsymbol{\lambda}^{*} = -\lambda(\boldsymbol{\lambda}^{\mathbf{T}}W^{T}\boldsymbol{\lambda} + \lambda\boldsymbol{\lambda}^{\mathbf{T}}.\nabla c).$$

Using the relations (2.20):

$$\dot{\lambda} = -\frac{\lambda}{c^{2}}\mathbf{c}^{\mathbf{T}}.(W^{T}\mathbf{c} + c\nabla c). \tag{2.25}$$

This is the second component equation in terms of the quantity $\lambda = |\boldsymbol{\lambda}|$.

In view of the relation, cf. (2.10):

$$\lambda = \frac{1}{c_{\mathrm{eff}}},$$

the equation may more conveniently be expressed in terms of the effective airspeed.

As

$$\frac{\dot{\lambda}}{\lambda} = -\frac{\dot{c}_{\mathrm{eff}}}{c_{\mathrm{eff}}}$$

and

$$\mathbf{c}_{\mathrm{eff}} = \frac{c_{\mathrm{eff}}}{c}\mathbf{c}.$$

Eq. (2.25) may be written:

$$\dot{c}_{\mathrm{eff}} = \frac{1}{c}\mathbf{c}_{\mathrm{eff}}^{\mathbf{T}}(W^{T}\mathbf{c} + c\nabla c) \tag{2.26}$$

or, using (2.14), in a more condensed form:

$$\dot{c}_{\mathrm{eff}} = \mathbf{c}_{\mathrm{eff}}^{\mathbf{T}}.\mathbf{n} \tag{2.27}$$

The result is an Euler equation for the rate of change of the speed of propagation of isochronals along the optimal trajectory.

This new equation which will be called the '*phase speed equation*', describes the rate of change of the phase speed of isochrones along the Least Time Track.

Substitution of (2.26) for the components of **w** and **c** gives:

$$\dot{c}_{\text{eff}} = c_{\text{eff}}\left\{\frac{\partial w^1}{\partial x^1}\cos^2\xi + \left(\frac{\partial w^1}{\partial x^2} + \frac{\partial w^2}{\partial x^1}\right)\sin\xi\cos\xi + \frac{\partial w^1}{\partial x^2}\sin^2\xi \right.$$

$$\left. + \frac{\partial c}{\partial x^1}\cos\xi + \frac{\partial c}{\partial x^2}\sin\xi\right\}. \tag{2.28}$$

By taking a Cartesian coordinate system at a given point of the track such that the x^1-axis points into the direction of the heading, this equation reduces to

$$\dot{c}_{\text{eff}} = c_{\text{eff}}\left(\frac{\partial w^1}{\partial x^1} + \frac{\partial c}{\partial x^1}\right)$$

or

$$\dot{c}_{\text{eff}} = c_{\text{eff}}\frac{\partial c_{\text{eff}}}{\partial x^1}. \tag{2.29}$$

Along the Least Time Track the rate of change of the effective airspeed is proportional to the local increase or decrease of the square of the effective airspeed.

The steering equation (2.23) and phase speed equation (2.28) together with the equations of motion (2.17) constitute an alternative form for the canonical system of Hamilton for the navigation problem.

When we shall discuss the numerical integration of these system equations it will appear that the integration may in general be performed by using only one of these component equations associated with the adjoint vector equation in addition to the equations of motion.

2.5 Phase velocity equation

The steering equation and phase speed equation associated with the component equations of the adjoint vector equation (2.13) suggest that we are able to establish yet another (vector) equation, which synthesizes both to give an equation of state for the effective airspeed *vector*.

The resulting equation should describe the rate of change of the phase velocity vector pertaining to the propagation of isochrones along the controlled optimal flight path.

Consider the identities:

$$\dot{c}^1{}_{\text{eff}} \equiv \frac{d}{dt}(c_{\text{eff}} \cos \xi) = \cos \xi \, \dot{c}_{\text{eff}} - c_{\text{eff}} \sin \xi \, \dot{\xi},$$

$$\dot{c}^2{}_{\text{eff}} \equiv \frac{d}{dt}(c_{\text{eff}} \sin \xi) = \sin \xi \, \dot{c}_{\text{eff}} + c_{\text{eff}} \cos \xi \, \dot{\xi}. \tag{2.30}$$

Substitution of (2.30) for (2.22) and (2.27) affords:

$$\dot{c}^1{}_{\text{eff}} = \frac{c_{\text{eff}}}{c}(\sin \xi \, \mathbf{c}^{*\mathbf{T}} + \cos \xi \, \mathbf{c}^{\mathbf{T}}) . \mathbf{n},$$

$$\dot{c}^2{}_{\text{eff}} = \frac{c_{\text{eff}}}{c}(-\cos \xi \, \mathbf{c}^{*\mathbf{T}} + \sin \xi \, \mathbf{c}^{\mathbf{T}}) . \mathbf{n}. \tag{2.31}$$

Introduce the matrix C:

$$C = \begin{bmatrix} \cos 2\xi & \sin 2\xi \\ \sin 2\xi & -\cos 2\xi \end{bmatrix}.$$

Then the set of equations (2.31) may be put in the matrix form:

$$\dot{\mathbf{c}}_{\text{eff}} = c_{\text{eff}} C\mathbf{n}.$$

If we put

$$\mathbf{n}' = C\mathbf{n}, \tag{2.32}$$

we obtain

$$\dot{\mathbf{c}}_{\text{eff}} = c_{\text{eff}} \, \mathbf{n}'. \tag{2.33}$$

This new vector equation which we will call the *phase velocity equation* has the same controlling properties as the gradient equation. By a decomposition into the component equations they both yield the steering equation and phase speed equation.

In a certain sense one is the inverse of the other. Combined with the equations of motion they afford both a solution to the time optimal control problem.

The control equations are linked by the vector $\mathbf{n}$ and vector $\mathbf{n}'$, for which a special symmetry property holds: $\mathbf{n}$ and $\mathbf{n}'$ are mirror images with respect to the normal to the isochrones. (Fig. 7).

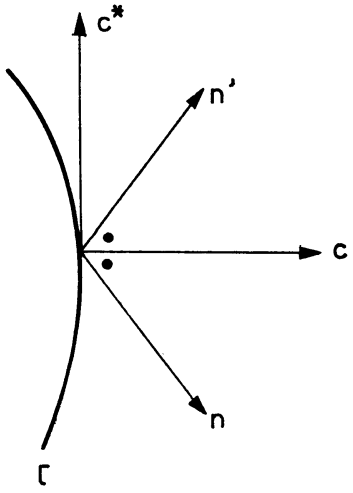

Fig. 7 See text.

Indeed, in view of the relations

$$C^T C = I,$$

I being the unit matrix,

$$\begin{aligned} \mathbf{c}^{\mathrm{T}} C &= \mathbf{c}^{\mathrm{T}}, \\ \mathbf{c}^{*\mathrm{T}} C &= -\mathbf{c}^{*\mathrm{T}}, \end{aligned} \tag{2.34}$$

we have

$$\begin{aligned} \mathbf{c}^{\mathrm{T}} . \mathbf{n}' &= \mathbf{c}^{\mathrm{T}} C \mathbf{n} = \mathbf{c}^{\mathrm{T}} . \mathbf{n}. \\ \mathbf{c}^{*\mathrm{T}} . \mathbf{n}' &= \mathbf{c}^{*\mathrm{T}} C \mathbf{n} = -\mathbf{c}^{*\mathrm{T}} . \mathbf{n}. \\ \mathbf{n}'^{\mathrm{T}} . \mathbf{n}' &= \mathbf{n}^{\mathrm{T}} C^T C \mathbf{n} = \mathbf{n}^{\mathrm{T}} . \mathbf{n}. \end{aligned} \tag{2.35}$$

which proves the symmetry of $\mathbf{n}$ and $\mathbf{n}'$ with respect to $\mathbf{c}$ normal to an isochrone.

As the phase velocity equation dictates the control of the flight similar to the gradient equation, the numerical integration of this new system will proceed along the same line as the integration of the system equations (2.16)-(2.17).

2.6 Refraction formulae

In view of the analogy of the time optimal control problem in navigation with problems in geometric optics in particular with Fermat's principle, it might be con-

jectured that in the presence of a discontinuity in the wind régime there will exist in air navigation a *refraction* formula similar to the law of refraction due to Snellius. The refraction formula in question is known as the refraction formula due to Von Mises. [3].

Given the interface between two wind régimes (media 1 and 2, see Fig. 8) with *uniform* wind velocity on either side of the interface, Von Mises' refraction formula reads:

$$\left(\frac{\sin i}{c_{\text{eff}}}\right)_1 = \left(\frac{\sin i}{c_{\text{eff}}}\right)_2. \tag{2.36}$$

i denotes the 'angle of incidence' or 'refraction angle' spanned by the airspeed vector **c** and the normal to the line of discontinuity.

When we consider the limiting case of an 'infinitesimal' discontinuity in the wind pattern the refraction formula tends to the differential form:

$$\frac{d}{dt}\left(\frac{\sin i}{c_{\text{eff}}}\right) = 0. \tag{2.37}$$

At first glance it may seem that for numerical integration of the navigation problem one could make use of this simple 'invariance' property. A more detailed investigation, however, reveals that this invariance property should be handled with the utmost care.

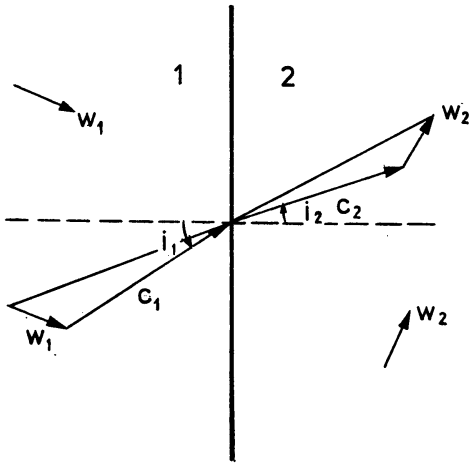

Fig. 8 Illustrating the law of refraction due to Von Mises.

First mode

Consider the gradient equation (2.15) in the form:

$$\dot{\boldsymbol{\lambda}} = -\frac{1}{c_{\text{eff}}}\mathbf{n}.$$

Let a local reference system be chosen such that the positive x^1-axis is directed along the vector $\mathbf{n}(n^1, n^2)$ and let i denote the heading with respect to the new x^1-axis (Fig. 9).

Then we have, in view of the properties (2.8) and (2.10):

$$\dot{\lambda}^2 = \frac{d}{dt}\left(\frac{\sin i}{c_{\text{eff}}}\right) = 0.$$

Here we immediately identify the differential form associated with Von Mises' refraction formula (2.36).

If the line l, normal to $\mathbf{n}$, is referred to as a fictitious *refraction line* we may formulate the result as follows:

In a comoving reference system along the optimal trajectory spanned by the refraction line l and its normal $\mathbf{n}$ the invariance property

$$\frac{d}{dt}\left(\frac{\sin i}{c_{\text{eff}}}\right) = 0 \tag{2.38}$$

holds.

This 'invariance property' is not to be considered in an absolute sense, as the property holds only with respect to a special comoving space along the optimal track!

(2.38) is the differential form for the refraction mode:

$$\left(\frac{\sin i}{c_{\text{eff}}}\right)_1 = \left(\frac{\sin i}{c_{\text{eff}}}\right)_2. \tag{2.39}$$

At every point of the Least Time Track there apparently exists a line of refraction and a *critical angle of incidence* for the heading, so that Von Mises' refraction formula is applicable.

In view of the definition of $\mathbf{n}$, see (2.14):

$$n_2 = \frac{\partial w^1}{\partial x^2}\cos i + \frac{\partial w^2}{\partial x^2}\sin i + \frac{\partial c}{\partial x^2} = 0.$$

For *fixed i*, given the definition of c_{eff}, cf. (2.3):

$$n_2 = \frac{\partial c_{\text{eff}}}{\partial x_2} = 0. \tag{2.40}$$

The refraction line is therefore characterized by the property that along this line for the critical angle of incidence the effective airspeed is constant.

By a transition to the limit we should be able to prove that the refraction formula will transfer into the steering equation. Indeed, writing our relation (2.38):

$$\cos i\, \dot{i} = \frac{1}{c_{\text{eff}}} \sin i\, \dot{c}_{\text{eff}}. \tag{2.41}$$

Substitution of (2.41) for (2.27) yields:

$$\dot{i} = \frac{\sin i}{\cos i} \frac{1}{c} \mathbf{c}^{\mathbf{T}} . \mathbf{n} = |\mathbf{n}| \sin i$$

or

$$\dot{i} = -\frac{1}{c} \mathbf{c}^{*\mathbf{T}} . \mathbf{n},$$

which proves the equivalence of the refraction formula with the steering equation (2.22).

This result states once more that quite different options might be given in terms of the optimal control. The option here lies in a proper choice of a reference system comoving along the trajectory such that the control equation obtains the structure of a refraction formula.

In the next section we shall see that we can easily derive another 'mode of refraction', which also governs the flight control in an optimal sense.

Second mode

The connection between the gradient equation and phase velocity equation suggests that we may establish still another invariance property.

Consider the phase velocity equation (2.33):

$$\dot{c}_{\text{eff}} = c_{\text{eff}} C\mathbf{n}.$$

Let now a local reference system be chosen such that the positive x^1-axis is directed along the vector $\mathbf{n}' = C\mathbf{n}$ and let $\hat{\imath}$ denote the heading with respect to the new x^1-axis. (Fig. 9.)

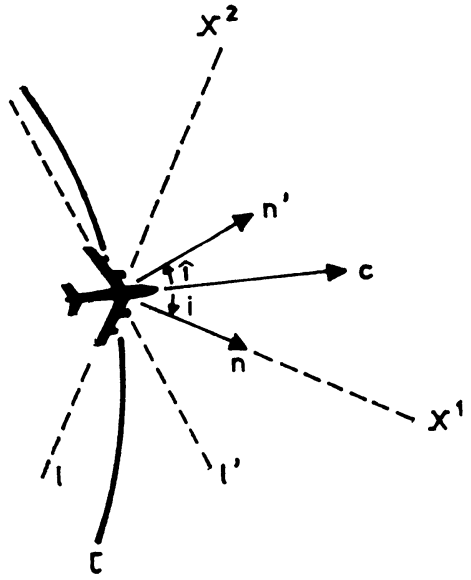

Fig. 9 Coordinate system with reference to the refraction normal and line of discontinuity.

Then we have

$$\dot{c}_{\text{eff}}^2 = \frac{d}{dt}(c_{\text{eff}} \sin \hat{\imath}) = 0. \tag{2.42}$$

This is a differential form for a mode of refraction, which might be expressed in the form:

$$(c_{\text{eff}} \sin \hat{\imath})_1 = (c_{\text{eff}} \sin \hat{\imath})_2. \tag{2.43}$$

Let l' denote the (fictitious) refraction line normal to $\mathbf{n}'$, then the invariance of $c_{\text{eff}} \sin \hat{\imath}$ is assured when adhering to the comoving reference system spanned by l' and $\mathbf{n}'$ along the trajectory.

Comparing (2.39) and (2.43) we have

$$\sin i_2 . \sin \hat{\imath}_2 = \sin i_1 . \sin \hat{\imath}_1$$

but

$$\sin i_1 = \sin \hat{\imath}_1,$$

whence

$$\sin i_2 . \sin \hat{\imath}_2 = \sin^2 i_1$$

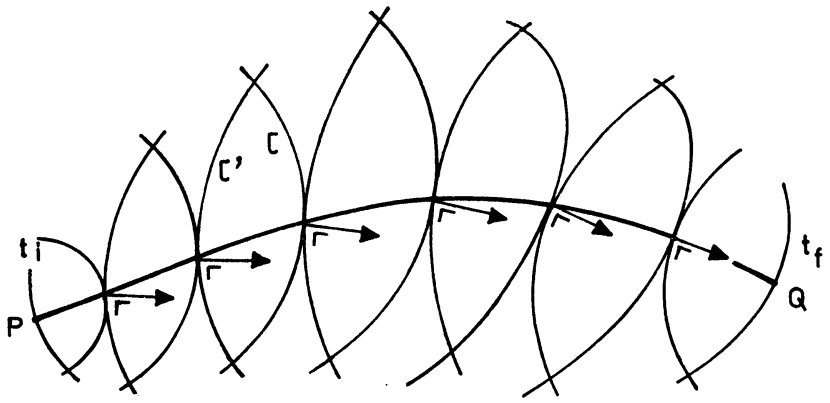

Fig. 10 Result of Hamilton's point of view.
Complete figure of time fronts and Least Time Track.

This equation is basic to describe the well-known method of the graphical construction of the Least Time Track by means of the complementary set of time fronts [8].

One of these sets $\tau = t - t_i = \text{const.}$ 'radiates' from the origin P at the initial time t_i. The complementary set $\tau' = t_f - t = \text{const.}$ 'radiates' towards the terminal point Q and reaches this point at the final time t_f (Fig. 10). The desired track is the *locus* of points of contact of corresponding isochrones for which

$$\tau + \tau' = t_f - t_i.$$

The sets of complementary isochrones together with the optimal trajectories are sometimes referred to as 'complete figure' due to Carathéodory.

The method bears much resemblance to that used in wave propagation studies in geometric optics (Huyghens' Principle).

Both the invariance properties can in principle be used to develop computational algorithms for the evaluation of the Least Time Track. At each integration step it will be required to determine the local refraction normal $\mathbf{n}$ or $\mathbf{n}'$. If correct, both refraction formulae should provide the same solution.

If we put

$$i_2 = i_1 + \varepsilon$$

this implies that we should have

$$\sin(i_1 + \varepsilon) . \sin(i_1 - \varepsilon) = \sin^2 i_1 .$$

This condition, however, is satisfied only if $\varepsilon = 0$ in the limiting case $dt \to 0$. As a consequence, it might be expected that in practice both numerical methods will produce slightly different tracks. For a further discussion we refer to section 5.

2.7 The Hamilton-Jacobi Equation

To obtain a partial differential equation for the optimal flight time we may follow a line of approach as indicated in section 1.2. We start with the classical Hamiltonian in the form

$$\mathscr{H}^0 = -1 + \frac{\partial \tau}{\partial x^1} w^1 + \frac{\partial \tau}{\partial x^2} w^2 + \frac{\partial \tau}{\partial x^1} c_0^1 + \frac{\partial \tau}{\partial x^2} c_0^2,$$

where $\mathbf{c}_0$ is the optimal control derived from Pontryagin's Maximum Principle, see (2.6) and (2.8).

$$c_0^1 = \frac{c}{\sqrt{\left(\frac{\partial \tau}{\partial x^1}\right)^2 + \left(\frac{\partial \tau}{\partial x^2}\right)^2}} \frac{\partial \tau}{\partial x^1}$$

$$c_0^2 = \frac{c}{\sqrt{\left(\frac{\partial \tau}{\partial x^1}\right)^2 + \left(\frac{\partial \tau}{\partial x^2}\right)^2}} \frac{\partial \tau}{\partial x^2}$$

Then the partial differential equation of Hamilton-Jacobi takes the form

$$\frac{\partial \tau}{\partial t} + \mathscr{H}^0(\mathbf{x}, \mathbf{c}_0(\mathbf{x}, t, \frac{\partial \tau}{\partial x^i}), t) = 0,$$

or

$$c^2 \left\{ \left(\frac{\partial \tau}{\partial x^1}\right)^2 + \left(\frac{\partial \tau}{\partial x^2}\right)^2 \right\} - \left(w^1 \frac{\partial \tau}{\partial x^1} + w^2 \frac{\partial \tau}{\partial x^2} + \frac{\partial \tau}{\partial t} - 1 \right)^2 = 0. \quad (2.44)$$

3. The graph method

In the foregoing analysis of the time optimal navigation problem it was tacitly assumed that there were no serious restrictions imposed on the utilization of the airspace, so that airlines can freely dispose of this space in order to make a profitable use of the prevailing atmospheric conditions.

In the last decennia, however, such flight regions have become sparser and sparser, mainly due to the increase of traffic density.

From a point of view of Air Traffic Control the airspace had, so to speak, to be 'atomized' and the routes to be flown had to be embedded in a more or less rigid traffic pattern.

In this situation the nature of the navigation problem has become entirely different from that studied previously.

The problem is getting involved in the class of *shortest path* problems pertaining to a traffic network which might be conceived as a *directed graph* of nodal points and directed lines or *arcs*. A specimen of such a graph is to be found in Fig. 37.
Within this graph the aim of optimal track selection is to find an ordering and grouping of the nodes and arcs such that the resulting path(s) be optimal in the sense of least flight time, costs or any other criterium.

In part II some graph theoretical concepts will form the central theme in a study of automation of '3-dimensional flight planning'. In this section we shall briefly touch upon the shortest path problem in a 2-dimensional graph which is not bound to be strictly identical with a preassigned route pattern.

Its purpose is to examine the time optimal problem in a *discrete* system as distinct from that in a *continuous* system, in particular to decide whether the corresponding computation algorithms might compete with the corresponding algorithms using one of the previously derived control equations in the continuous system.

The simplest way to tackle this problem is to replace the continuum by an arbitrary graph. This may be accomplished by 'sprinkling' – at random or not – a set of points over the area of interest and to define a set of arcs between all pairs of nodes. Some of these arcs may be blocked.

The nodes of the graph are junctions where one or more arcs meet, the arcs are directed flight segments between two junctions and associated with each arc is a number – its 'length'.

Then we may put the question: 'what is the shortest path between two given nodes?'

Denoting flight time as 'length' of an arc the problem reduces to that of finding the Least Time Track between two given nodes in the graph.

To evaluate the shortest path in a general directed graph or digraph a few algorithms exist, which are particularly adapted for machine computation. To mention a few: the methods of Dantzig-Mintz [19], Ford-Dantzig [19], Dijkstra [20], Bellman [13], [21].

Bellman's matrix algorithm is of particular interest in view of its relationship with the solution of the partial differential equation of Hamilton-Jacobi in a continuous system especially with the complete figures of isochrones and optimal trajectories. (See section 2.7.)

We shall give a brief account here of this method, especially within the framework of the navigation problem.

Let the nodes of the graph be $P_1, P_2, \ldots, P_n$. If nodes P_i and P_j are joined by an arc, let d_{ij} denote the distance from P_i to P_j. For those pair of nodes which are not directly linked by an arc we put $d_{ij} = \infty$ where '∞' denotes some suitable large number.

Naturally we put $d_{ii} = 0$. We now examine the problem of finding the 'length' of the shortest path between any two given nodes in terms of the matrix $D = d_{ij}$.

Consider the equation

$$d_{ij}^{(2)} = \min_k (d_{ik} + d_{kj}). \tag{3.1}$$

Here $d_{ik} + d_{kj}$ is the length of the path from P_i to P_j passing through just one intermediate node P_k. Since we allow $k = j$ the direct distance d_{ij} is included in the quantities on the left of (3.1).

Hence the matrix $D^{(2)} = d_{ij}^{(2)}$ is the distance matrix for the shortest path between P_i and P_j using at most two arcs.

The matrix $D^{(2)}$ might be considered as the 'square' of the matrix D under an odd sort of 'multiplication'.

If we square $D^{(2)}$ in exactly the same way a matrix $D^{(4)}$ is obtained whose (i,j) element is the length of the shortest path from P_i to P_j involving at most four arcs. After N similar 'squarings' we obtain a matrix whose (i,j) element is the length of the shortest path from P_i to P_j using at most 2^N arcs. If $2^N > n$ where n is the number of nodes of the graph, then the matrix given is what we want. *The matrix $D^{(N)}$ is the matrix of shortest distances between any pair of nodes.*

The algorithm is very convenient for calculation especially when applying a very elegant method described by Hu [21].

According to that method no more than two 'squaring operations' are needed to obtain the result.

Whenever we calculate an element $d_{ij}^{(2)}$ using (3.1) we immediately put it down in place of the element d_{ij} in the matrix $D^{(1)}$.

Subsequent calculations use these new values instead of the old. The order in which we compute the elements $d_{ij}^{(2)}$ in the first 'squaring' is from left to right and going downwards. When the squaring of the matrix is done the second time, we do exactly the opposite: right to left along the bottom and then upwards. The matrix resulting from this two-stage process is again the shortest distance matrix $D^{(N)}$.

An imperfection of the algorithm is that it gives no information about the optimal track itself corresponding to the shortest distance. To identify the optimal track between the given pair of nodes P_i and P_j we propose the following method.

This method is inspired by the property that in a continuous system the optimum track reveals itself as the locus of points of contact of isochrones in a complete figure (see Fig. 10).

Consider row i in the matrix D^N. The elements (i, k), $k = 1, \ldots, n$ represent the lengths of the shortest paths with P_i as a common node. This set of shortest paths constitutes a *shortest spanning tree with root* P_i.

In the same way the elements (k, j), $k = 1, \ldots, n$ in the column j of the matrix D^N are the lengths of the shortest paths with common node P_j — *shortest spanning tree with root* P_j. Both these shortest spanning trees have a shortest path in common: the shortest path between the nodes P_i and P_j. The nodes P_l on this common shortest path are specified by means of the property:

$$d_{il} + d_{lj} = \text{const.} = d_{ij}. \tag{3.2}$$

Therefore, to obtain the track information one simply determines the sequence of numbers

$$d_{ik} + d_{kj}, \; k = 1, \ldots, n$$

and sorts out those k-values for which

$$d_{ik} + d_{kj} = \text{const.} = d_{ij}$$

The procedure in itself is very conveniently to apply to a graph which corresponds to a given route network. Then the matrix D^N will afford all required information concerning the shortest path between any pair of airports depending on the notion of 'length' of an arc. The matrix algorithm is therefore particularly suited for determining such elements like the *geometric* shortest paths, the Least Time Tracks (based on a fixed time prognostic chart), time tables, etc.

Its usefulness for *practical* operations is, however, somewhat questionable for the following reasons.

The 'length' d_{ij} of an arc may be variable. For example, if the length is referred to flight time, the time of flight along a route segment may not be time-invariant due to the time variability of winds and temperature. The length d_{ij} may also be dependent on the past history. This may occur for instance when the length refers to fuel consumption: the fuel consumption depends on the fuel already consumed after take-off.

These examples explain why it is necessary to look out for alternative computational algorithms which are more suited to deal especially with boundary value problems, and which admit a computation proceeding stepwise starting from one of the end points.

This type of graph algorithm will be described in a special graph, which is often used in transport problems. Berge [11], [12].

Let the set of nodes in a graph be arranged in the sub-sets X_i $(i = 0, \ldots, n+1)$ having the following properties: each node in X_{i+1} is linked with one or more nodes in X_i. There are no links between any pair of nodes in X_i. X_0 is the entry point, X_{n+1} the exit point. The pictorial form of such a graph is indicated in Fig. 11.

Such a graph may be obtained by locating the nodes on (non-intersecting) curves, for instance on the gridlines of a meteorological grid (Fig. 31), the (standard) meridians in oceanic regions (Fig. 37), circles about the end points (Fig. 32), etc.

Let in the graph a node A be labelled by the indices (i, k), $k = k_1(i) \ldots k_2(i)$. The shortest distance between the entry point P and A be $d_{(i,\,k)}$.

To acquire the shortest path between entry point P and exit point Q we consider an optimization process which works as follows:

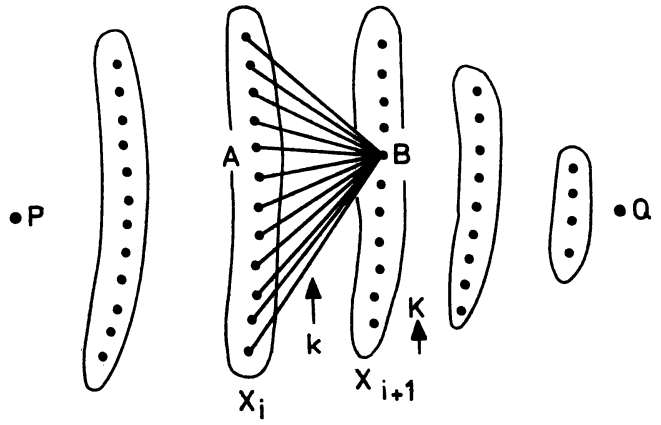

Fig. 11 Optimization in a special transport graph.

Fixate a node $B(i+1, K)$ in the set X_{i+1} and assume that for the nodes $A(i, k)$ in the set X_i the shortest distance from the entry point is known from previous optimization steps.

Then we consider:

$$d_{(i+1,K)} = \underset{k=k_1(i),\ldots,k_2(i)}{min.} \left(d_{(i,k)} + d^{(i,k)}_{(i+1,K)}\right) \tag{3.3}$$

The result is the shortest distance from the entry point to the fixated node B.

This optimization step is repeated by varying i and K in appropriate cycles until the exit point Q has been reached.

The optimization scheme may be put in the mathematical form:

$$d_{PQ} = \left[\left[\underset{k=k_1(i),\ldots,k_2(i)}{[min.}(d_{(i,k)} + d^{(i,k)}_{(i+1,K)})]\, K = K_1(i+1),\ldots,K_2(i+1)\right] i = 0,\ldots,n\right] \tag{3.4}$$

The information required to identify the shortest path is contained in a set of 'critical' indices $k_{crit}(i)$, which is recorded simultaneously with the serial steps:

$$[k_{crit}(i) \mid d_{(i+1,K)}]\; i = 0,\ldots,n \tag{3.5}$$

The algorithm works sequentially and is organized in such a manner that in a subsequent arc the 'length' may be adapted to the past history of the quantity or to actual time.

In addition, the upper and lower bounds of the cycles may be chosen at will, which makes the use of the algorithm very flexible. For instance, by equating the lower and upper bounds of k and K the scheme may be activated to present all flight data along a single *preselected* path in the graph.

The track information corresponding to the shortest distance between entry and exit point can be found by storing at each node $(i+1, K)$ the (i, k) coordinates by which a minimum is attained.

After reaching the exit point we obtain the nodes on the optimal track in the proper ordering by 'reading' the stored (i, k) coordinates starting with the exit point Q and going backwards up to the entry point P.

The graph method has some outstanding advantages in comparison with the methods using a control law. Most important is that the graph method avoids a lot of difficulties in regard to mathematical rigour (continuity, convergence, differentiability). This is a reason why for instance the graph method will always afford a unique solution between two given nodes, which guarantees an *absolute* minimum distance. This in

contrast with the methods based on the steering equation etc., where it may happen that between the starting point and the terminal a multiplicity of solutions is found, each giving a minimum with respect to admissible trajectories in its vicinity. To find an absolute minimum requires a special searching technique.

Another feature of the discretization using a graph is the possibility of defining a NEXT BEST route or, more generally, a k-th BEST route ($k \geqslant 1$). The order of preference of such routes is an important item in regard to the design of *organized track systems* in certain flight regions, e.g. the North Atlantic.

A disadvantage of the graph method is that the computer running time is in general not favourable compared with that of the iterative processes in a continuous system, at least in cases of a fast speed of convergence of the iteration.

4. The control equations for a conformal mapping of the earth's surface

The mathematical derivation to solve the time optimal control problem was valid for a plane surface. The control equations and equation of motion preserve their validity *locally* in a plane tangent to the two-dimensional sphere. Application of a step by step integration method in the tangent plane comoving along the trajectory will afford the correct solution on the earth's surface.

In meteorological practice as well as in flight- and air traffic control use must be made of a mapping of the earth's surface on a plane. This is done for obvious reasons, but a consequence is that the mapping will affect the solution on a *global* scale in the projection map.

The effect may be visualized when a computer programme is developed on the basis of one of the control equations together with the equation of motion, which is activated to generate the Least Time Track, for instance in a polar stereographic chart between, say, Amsterdam and New York in *zero wind* conditions (and constant temperature). Then the result will be a straight line on the chart. The desired solution, however, should have been the great circle path, which deviates a hundred miles from the straight line. (See Fig. 27.)

The spherical case should, therefore, be made to reduce to the plane case by a mapping of the sphere onto the plane. Every such mapping requires a compensation for the distortions caused by the scale of the mapping.

It is taken for granted that the mapping is conformal like the Lambert conformal cone projection, the polar stereographic projection and the cylindrical Mercator projection. In such mappings angles are preserved and the scale is a function of position only. In this section we shall derive the corrective terms to be applied in the previously described system equations. These equations corrected for the effect of the conformal mapping are fundamental for the remaining part of this treatise.

We might suppose that the sphere has been mapped conform to a plane surface. Establish on the plane map a system of rectangular coordinates x'_1 and x'_2. The inverse of the mapping is again a conformal mapping of the plane onto the surface.

Then, in view of the conformity of the mapping the system of rectangular coordinates x'_1, x'_2 will be mapped in a local system of rectangular coordinates x^1, x^2 on the sphere.

Let $S(x^1, x^2)$ be the scale of the projection of the sphere onto the plane at the point defined by (x^1, x^2) on the sphere.

For the Lambert conic projection of which the polar stereographic and Mercator

projections may be considered as limiting cases, S is a function of the latitude φ only [23]:

$$S = \frac{S_0}{(\cos \varphi)^{1-p}(1 + \sin \varphi)^p}.$$

S_0 is the map scale at standard parallel. For instance, in the polar stereographic projection with standard parallel at 60^0N: $S_0 = 1 + \sin \pi/3 = 1.86603$. ($p = 1$).
p is a parameter.

For the Mercator projection: $p = 0$, so that

$$S = \frac{S_0}{\cos \varphi}$$

For the polar stereographic projection: $p = 1$:

$$S = \frac{S_0}{1 + \sin \varphi}$$

For the Lambert projection: $0 < p < 1$ (4.1)

A displacement on the sphere corresponding to a small displacement dx_1', dx_2' on the map will have components

$$\begin{aligned} dx^1 &= S^{-1} dx_1', \\ dx^2 &= S^{-1} dx_2'. \end{aligned} \tag{4.2}$$

Similarly a vector such as the airspeed vector $\mathbf{c}$ or wind vector $\mathbf{w}$ on the sphere – being displacements per unit time – will correspond to the vector quantities in the plane according to

$$\begin{aligned} \mathbf{c} &= S^{-1}\mathbf{c}', \\ \mathbf{w} &= S^{-1}\mathbf{w}'. \end{aligned} \tag{4.3}$$

One of the advantages of the conformity of the mapping is that the steering angle is preserved by the mapping.

The form of the equation of motion (2.1) is not changed when expressed in the variables on the earth's surface.

Indeed, (2.1) transfers into

$$S\dot{\mathbf{x}} = S\mathbf{w} + S\mathbf{c}$$

or

$$\dot{\mathbf{x}} = \mathbf{w} + \mathbf{c},$$

as measured on the earth's surface.

If, however, a similar procedure is applied to one of the control equations like the gradient equation, the steering equation, etc. it turns out that some corrective terms are required to preserve the mathematical form of these equations.

For example, if we express the gradient equation (2.16) together with (2.14) in terms of the variables on the spherical surface of the earth, then we obtain:

$$\frac{d}{dt}\left(\frac{\nabla\tau}{S}\right) \equiv -\frac{1}{S^2}\frac{dS}{dt}\nabla\tau + \frac{1}{S}\frac{d\nabla\tau}{dt}$$

$$= -\frac{1}{Sc_{\text{eff}}}\left(\frac{1}{Sc}\frac{\partial S\mathbf{w}}{S\partial\mathbf{x}}.S\mathbf{c} + \frac{\partial Sc}{S\partial x}\right). \tag{4.4}$$

Since

$$\frac{\partial S}{\partial t} \equiv 0:$$

$$\frac{dS}{dt} = (\mathbf{c} + \mathbf{w})^T.\nabla S.$$

Referring to (2.6), (2.10) and (2.11):

$$\nabla\tau = \frac{1}{c_{\text{eff}}}\frac{\mathbf{c}}{c}.$$

In addition we have the identities

$$\frac{\partial Sc}{S\partial x} \equiv \frac{\partial c}{\partial x} + \frac{c}{S}\frac{\partial S}{\partial x} = \nabla c + \frac{c}{S}\nabla S,$$

$$\frac{\partial S\mathbf{w}}{S\partial\mathbf{x}} \equiv \frac{\partial \mathbf{w}}{\partial\mathbf{x}} + \frac{1}{S}\frac{\partial S}{\partial x}.\mathbf{w}^{\mathbf{T}} = W^T + \frac{1}{S}\nabla S.\mathbf{w}^T.$$

Furthermore, in view of the conformity of the mapping we may again introduce the effective airspeed

$$c_{\text{eff}} = c + \frac{1}{c}\mathbf{w}^{\mathbf{T}}.\mathbf{c},$$

as measured on the sphere.

By substitution of these relations into (4.4) we find:

$$\frac{d\nabla\tau}{dt} = -\frac{1}{c_{\text{eff}}}\left(\frac{1}{c}W^T.\mathbf{c} + \nabla c + c_{\text{eff}}\frac{\nabla S}{S} - \frac{(\mathbf{c}+\mathbf{w})^T.\nabla S}{S}.\frac{\mathbf{c}}{c}\right).$$

If we introduce a new 'refraction normal':

$$\mathbf{m} = \frac{1}{c}W^T.\mathbf{c} + \nabla c + c_{\text{eff}}\frac{\nabla S}{S} - \frac{(\mathbf{c}+\mathbf{w})^T.\nabla S}{S}\frac{\mathbf{c}}{c}, \tag{4.5}$$

the gradient equation may be written in the form

$$\dot{\nabla}\tau = -\frac{1}{c_{\text{eff}}}\mathbf{m}, \tag{4.6}$$

as measured on the sphere.

We see that the equations of state for the gradient of isochrones along the least time track with respect to a fixed Cartesian system in the image plane – with the variables measured on the sphere –, preserve the original form (2.16), given the new definition of the 'refraction normal' (4.5).

The corrective terms in **m** compensate for the distortion caused by the earth's curvature as expressed by the scale of the mapping.

By a similar reasoning it may be demonstrated that the vector equation for the phase velocity transfers into:

$$\dot{\mathbf{c}}_{\text{eff}} = c_{\text{eff}}C\mathbf{m}, \tag{4.7}$$

as measured on the sphere.

Its component equations become

steering equation:

$$\dot{\xi} = -\frac{\mathbf{c}^{*T}}{c}.\mathbf{m}, \tag{4.8}$$

phase speed equation:

$$\dot{c}_{\text{eff}} = c_{\text{eff}}\frac{\mathbf{c}^T}{c}.\mathbf{m}. \tag{4.9}$$

These control equations apply to any regular two-dimensional surface, since any

such surface can be mapped conformally onto a plane surface. In particular the equations will hold for a conformal mapping of the earth's *geoïde* onto a plane.

It should be noted that the mappings include the Lambert conformal conic projection and the cylindrical Mercator projection as the cone and cylinder may be unrolled to form the plane map.

In case of a plane surface the mapping may be regarded as a projection of the surface onto itself ($S \equiv 1$). Then the control equations obtain their original form valid for the plane surface.

The meaning of the corrective term may be understood if we consider the special case of a flight with constant airspeed within zero wind conditions.

Then:

$$\mathbf{w} = 0,\ c = \text{const.}$$

$$W = 0,\ \nabla c = 0$$

$$\mathbf{m} = c\frac{\nabla S}{S} - \frac{\mathbf{c}^T . \nabla S . \mathbf{c}}{cS}.$$

The component equations (4.8) and (4.9) in this case take the form:

$$\dot{\xi} = -\frac{1}{S}\mathbf{c}^{*T} . \nabla S,$$

$$\dot{c} = c\frac{\mathbf{c}^T . \nabla S}{S} - c\frac{\mathbf{c}^T . \nabla S}{S} = 0. \tag{4.10}$$

As a consequence of the optimization principle this set of equations should describe the *uniform motion along a great circle on the earth's surface and its conformal mapping onto a plane.*

The second equation is obvious, but to prove the first a rather lengthy calculation is required. In the Annex the proof is given in terms of the conformal mappings used in meteorological practice.

It may be expected that the conformal mapping will also affect the interpretation of the refraction principle.

To derive the desired invariance properties we may use here the same method as explained in section 2.6.

As a result it is found that the invariance properties (2.39) and (2.43) preserve their validity, provided that the (fictitious) refraction line in the image plane is adapted again to the vector $\mathbf{n}$, corrected for the effect of the earth's curvature.

In the first mode for which the invariance property:

$$\frac{d}{dt}\left(\frac{\sin i}{c_{\text{eff}}}\right) = 0 \tag{4.11}$$

holds, the vector **n** should be replaced by the vector **m**, cf. (4.5). The angle i denotes here the critical angle of incidence between **c** and **m**.

In the second mode for which the invariance property:

$$\frac{d}{dt}(c_{\text{eff}} \sin \hat{\imath}) = 0 \tag{4.12}$$

holds, the vector $\mathbf{n}' = C\mathbf{n}$ should be replaced by the vector $\mathbf{m}' = C\mathbf{m}$. Here $\hat{\imath}$ represents the critical angle of incidence between **c** and **m**′.

The vector **m**′ is identical to **m** reflected with respect to the airspeed vector **c**.

An interesting feature is that the mapping problem admits yet another formulation of the refraction principle.

Indeed, consider the relation (4.4).

Omitting total differentiation of the quantity $\frac{\nabla\tau}{S}$ the equation of state for this quantity may be written (4.5):

$$\frac{d}{dt}\left(\frac{\nabla\tau}{S}\right) = -\frac{1}{Sc_{\text{eff}}}\boldsymbol{\mu},$$

where

$$\boldsymbol{\mu} = \frac{1}{c}W^T \cdot \mathbf{c} + \nabla c + c_{\text{eff}}\frac{\nabla S}{S}. \tag{4.13}$$

Then, in a local reference system in the image plane for which the x^1-axis is directed along $\boldsymbol{\mu}$:

$$\frac{d}{dt}\left(\frac{\sin i}{Sc_{\text{eff}}}\right) = 0. \tag{4.14}$$

This is the differential form for the refraction equation

$$\left(\frac{\sin i}{Sc_{eff}}\right)_1 = \left(\frac{\sin i}{Sc_{eff}}\right)_2. \tag{4.15}$$

The subscripts 1 and 2 refer to the 'media' on either side of the refraction line normal to $\boldsymbol{\mu}$.

In a similar way the second mode of refraction tolerates an alternative formulation:

$$\frac{d}{dt}(Sc_{\text{eff}} \sin i) = 0 \tag{4.16}$$

which is the limiting form of the refraction formula:

$$(Sc_{\text{eff}} \sin i)_1 = (Sc_{\text{eff}} \sin i)_2 \tag{4.17}$$

Here the vector $\boldsymbol{\mu}$ should be replaced by the vector

$$\boldsymbol{\mu}' = C\boldsymbol{\mu}$$

which is the vector $\boldsymbol{\mu}$ reflected with respect to the airspeed vector $\mathbf{c}$.

In the case of the plane surface solution ($S \equiv 1$) both modes of refraction obtain the original meaning as

$$\boldsymbol{\mu} = \mathbf{m} = \mathbf{n},$$

$$\boldsymbol{\mu}' = \mathbf{m}' = \mathbf{n}'.$$

For checking purposes it is convenient to study the special case of a flight with constant airspeed in zero wind conditions. The refraction formulae should then produce the great circle arcs in the conformal mapping.

To that aim we put:

$$\mathbf{w} = 0, \; c = \text{const.}$$

$$W = 0, \; \nabla c = 0.$$

In the first mode using the version (4.14) the refraction normal (4.13) becomes:

$$\boldsymbol{\mu} = c\frac{\nabla S}{S}.$$

This vector is directed along the meridian in the projection map pointing to higher S-values. (Fig. 12.)

The corresponding refraction equation is

$$\frac{\sin i_1}{S_1} = \frac{\sin i_2}{S_2}, \tag{4.18}$$

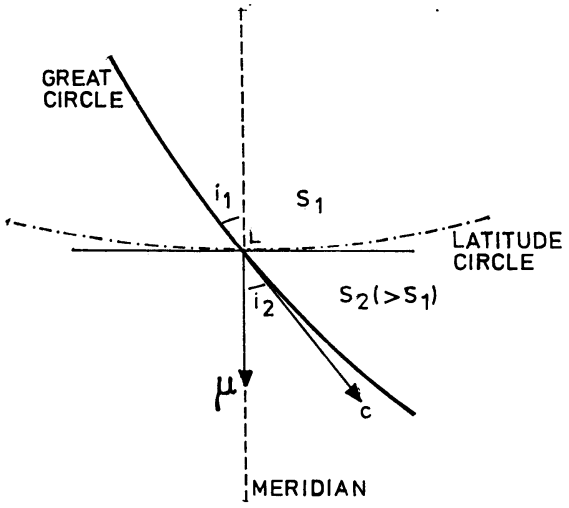

Fig. 12 Refraction in case of zero winds on a conformal mapping of the earth's surface. The result is a great circle arc.

which is equivalent to Snellius' law of refraction in geometric optics. The scale S plays the role of a refraction index.

Using the version (4.11) the refraction normal (4.5) becomes:

$$\mathbf{m} = c\frac{\nabla S}{S} - \frac{\mathbf{c} \cdot \nabla S}{S} \cdot \frac{\mathbf{c}}{c}.$$

$\mathbf{m}$ is the component of $\boldsymbol{\mu}$ normal to $\mathbf{c}$. (Fig. 13.)

In this case the refraction formula (4.11) reduces to:

$$\frac{d}{dt} \sin i = \cos i \frac{di}{dt} = 0.$$

The critical angle of incidence i, however, is equal to $\pi/2$ always, so that $\cos i \equiv 0$.

This implies that $\frac{di}{dt}$ is *indeterminate*, excluding the use of this version for the evaluation of the mapping of the great circle motion!

A similar form of degeneracy comes into play when considering the second mode of refraction.

Only in the form (4.17):

$$\frac{\sin i_1}{S_2} = \frac{\sin i_2}{S_1},$$

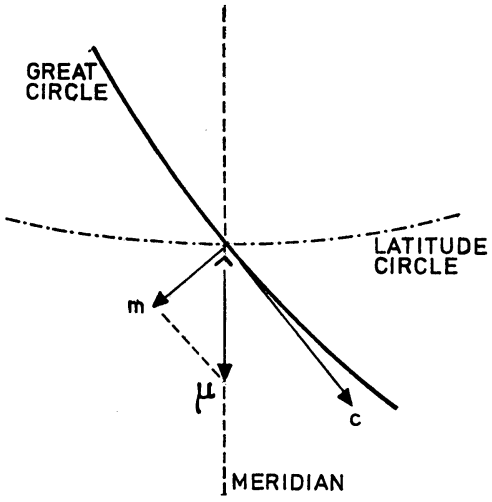

Fig. 13 Configuration of the refraction normal and geographical coordinate lines in case of zero winds on a conformal mapping of the earth's surface.

this mode may produce the desired track.

The generation of the great circle track by computer is an important detail for instance to check the correct functioning of the appropriate computer programmes.

Proceeding along the same lines we may investigate the effect of the conformal mapping on the Hamiltonian point of view as described in section 2.7.

It is easy to prove that the form of the equation of Hamilton-Jacobi (2.44) is unaffected by this mapping, for (2.44) transfers into:

$$S^2c^2\left\{\left(\frac{\partial\tau}{S\partial x^1}\right)^2+\left(\frac{\partial\tau}{S\partial x^2}\right)^2\right\}-\left(Sw^1\frac{\partial\tau}{S\partial x^1}+Sw^2\frac{\partial\tau}{S\partial x^2}+\frac{\partial\tau}{\partial t}-1\right)^2=0.$$

Working this out it is observed that the equation as expressed in terms of the variables on the spherical surface, preserves exactly the original form (2.44).

One of its consequences is that the mapping does not affect either the geometrical nature of the problem. The conformal mapping for instance does not disturb the intrinsic properties of the complete figure. The connection between the Hamiltonian point of view and the graph method implies that the mapping has also no effect at all on the solution of the shortest path problem in a graph, provided that the 'length' of an arc in the image plane is as measured on the sphere!

5. Practical evaluation of the Least Time Track

5.1 Introduction

Various studies are now being devoted to the development of computational algorithms suitable for a comprehensive programming [14], [15]. Most, if not all, of the proposed methods make use of some kind of optimal track selection within a graph associated with a prefixed route network.

In technical literature papers dealing with the optimization in a continuous system are very scanty. Apart from a report on the programming, using Von Mises' refraction formula, Ränicke [26], there are to the author's knowledge no accounts available in literature on the synthesis of the navigation problem using one of the control equations including Zermelo's steering equation.

We have already noticed that the problem concerned with the determination of the time optimal flight path falls in the domain of the calculus of variations and that this problem could be treated from the point of view of Pontryagin's approach to the theory of optimal control processes.

The study of the maximum principle and the canonical system of Hamilton has led to the formulation of the adjoint variational equation together with the equation of motion. The state of the adjoint vector corresponds directly with the gradient of isochrones. By the maximum principle this gradient is related to the controllable air speed vector, so that it may be stated that the optimal control is basically governed by the adjoint equation or gradient equation. The desired solutions of the optimization problem might be obtained by integration of the gradient equation together with the equations of motion.

The steering equation, phase speed equation etc. should merely be considered as substitutes for the controlling gradient equation but they might be easier to manipulate with in numerical computation.

In practice it is required to integrate the system equations so as to acquire a solution for a given initial and final state.

The solution of this two point boundary problem is possible only if the *optimal initial state* of the gradient of isochrones is known. For arbitrary systems and general conditions of end states the question of the optimal intial state of the adjoint vector however has never been completely answered. The problem can be solved by the idea of searching for the optimal initial state of the gradient in a systematic manner. This calls for the application of a suitable iterative procedure. It is recognized that such iterative procedures, if performed by manual handling, are very cumbersome and time-consuming. Electronic computers, however, form a powerful tool in mastering these iterative searching techniques.

5.2 Iterative scheme for the integration of the system equations

There are many paths open to the development of iterative searching techniques. The basic idea underlying this iterative scheme is the following:

First select an appropriate initial state of control by some predetermined means. The computer determines the response of the control system for some finite time interval into the future. The determination of a sequence of responses will lead to the generation of an optimal trajectory satisfying the basic equations of the system. The trajectory, however, will in general not satisfy the end state condition. The final state is bypassed at a certain 'distance' in phase space. The objective of the iteration scheme is to bring this distance or a function of it to zero.

After the completion of a cycle the computer determines if a new choice of the initial state of control is necessary. The criterium to re-start the computation depends on a certain tolerance limit in terms of the distance function, and the next estimate of the initial state is chosen such that a rapid convergence of the iteration process is obtained.

This iteration searching technique may be applied to all control equations, like the gradient equation, steering equation, as well as to one of the modes of refraction.

In the discussion to follow it is assumed that the optimization will be performed in a conformal polar stereographic projection with true scale at 60^0 N.

We first give a summary of the system equations and refraction formulae. By using the identities (2.34) which are unaffected by the mapping, the control equations are written in terms of the vector $\mathbf{m}$ or the vector $\mathbf{m}' = C\mathbf{m}$.

Reference is made to (2.3), (2.34), (4.1), (4.5) (4.6), (4.7), (4.8), (4.9), (4.11), (4.12), (4.13), (4.15) and (4.17).

Summary

Scale of stereographic projection:

$$S = \frac{1.86603}{1 + \sin\varphi} \tag{5.1}$$

Effective airspeed:

$$c_{eff} = c + \frac{1}{c}\,\mathbf{w}^{\mathbf{T}} \,.\, \mathbf{c}. \tag{5.2}$$

Matrix C:

$$C = \begin{bmatrix} \cos 2\xi & \sin 2\xi \\ \sin 2\xi & -\cos 2\xi \end{bmatrix}. \tag{5.3}$$

Refraction normal:

$$\mathbf{m} = \frac{1}{c} W^T . \mathbf{c} + \nabla c + c_{\text{eff}} \frac{\nabla S}{S} - (\mathbf{c} + \mathbf{w})^T . \frac{\nabla S}{S} \frac{\mathbf{c}}{c}, \tag{5.4}$$

$$\mathbf{m}' = C\mathbf{m}, \tag{5.5}$$

$$\boldsymbol{\mu} = \frac{1}{c} W^T \mathbf{c} + \nabla c + c_{\text{eff}} \frac{\nabla S}{S}, \tag{5.6}$$

$$\boldsymbol{\mu}' = C\boldsymbol{\mu}. \tag{5.7}$$

Equation of motion:

$$\dot{\mathbf{x}} = \mathbf{w} + \mathbf{c}. \tag{5.8}$$

Vector control equations:
a gradient equation:

$$\dot{\nabla}\tau = -\frac{1}{c_{\text{eff}}} \mathbf{m} = -\frac{1}{c_{\text{eff}}} C\mathbf{m}'. \tag{5.9}$$

b phase velocity equation:

$$\dot{\mathbf{c}}_{\text{eff}} = c_{\text{eff}} C\mathbf{m} = c_{\text{eff}} \mathbf{m}'. \tag{5.10}$$

Scalar control equations:
a steering equation:

$$\dot{\xi} = -\frac{1}{c} \mathbf{c}^{*\mathbf{T}} . \mathbf{m} = \frac{1}{c} \mathbf{c}^{*\mathbf{T}} . \mathbf{m}'. \tag{5.11}$$

b phase speed equation:

$$\dot{c}_{\text{eff}} = \mathbf{c}^{\mathbf{T}}{}_{\text{eff}} . \mathbf{m} = \mathbf{c}^{\mathbf{T}}{}_{\text{eff}} . \mathbf{m}'. \tag{5.12}$$

First mode of refraction:
a in terms of **m**:

$$\left(\frac{\sin i}{c_{\text{eff}}}\right)_1 = \left(\frac{\sin i}{c_{\text{eff}}}\right)_2. \tag{5.13}$$

b in terms of $\boldsymbol{\mu}$:

$$\left(\frac{\sin i}{Sc_{\text{eff}}}\right)_1 = \left(\frac{\sin i}{Sc_{\text{eff}}}\right)_2. \tag{5.14}$$

Second mode of refraction:
a in terms of $\mathbf{m}'$:

$$(c_{\text{eff}} \sin i)_1 = (c_{\text{eff}} \sin i)_2. \tag{5.15}$$

b in terms of $\boldsymbol{\mu}'$:

$$(Sc_{\text{eff}} \sin i)_1 = (Sc_{\text{eff}} \sin i)_2. \tag{5.16}$$

The system equations which should be integrated step by step in the iteration process consist here of the equation of motion (5.8) together with one of the control equations (5.9), (5.10), (5.11), (5.12) or one of the refraction formulae (5.13), (5.14), (5.15), (5.16).

These equations apply in regard to a fixed Cartesian coordinate system on the image plane. The state variables are evaluated as measured on the earth's surface.

The entire analysis is obviously centred about the notion of a refraction normal $(\mathbf{m}, \mathbf{m}', \boldsymbol{\mu}, \boldsymbol{\mu}')$. This vector quantity has been introduced in order to bring the results in a convenient mathematical form. The vector combines the effects of all system parameters on the control including the effect of the mapping operation. In case the constant Mach number technique is used the airspeed may be replaced by a function of temperature according to formula (5.21).

The step by step determination of the refraction normal will be crucial in the iteration process. Its computation requires the evaluation of some quantities to be derived from the local wind and temperature conditions and the scale of the mapping. These are obtained by application of finite difference techniques over a suitably defined grid.

5.3 Provision of basic operational data

The synthesis of the navigation problem requires the assimilation of a considerable amount of operational data. This data requirement comprises the meteorological conditions in the flight region, geographical and geometric data and some performance information.

Geographical data

The processing is developed with respect to a fixed orthogonal coordinate system in the image plane of a conformal mapping of the earth's surface.

Field data are depicted from the information points of a regular geographical *grid* or a Cartesian grid superposed on a conformal map. This requires a specification of the type and scale of the mapping, the scale of magnification and some parameters defining the grid geometry such as orientation and delineation of the grid and the grid spacing at the latitude of true scale.

A fixed coordinate system Oxy might therefore best be taken with reference to the grid system, and the x, y coordinates be expressed in grid units.

Example:

Consider a polar stereographic projection with true scale at 60° N. Let a rectangular grid be superposed on the map in which the pole is a grid point. One of the grid lines through the pole, parallel to the meridian λ_0, is taken as y-axis. (Fig. 14.) The origin $O(\lambda_0, \varphi_0)$ is taken in a grid point on this meridian.

A specimen of this type of grid is the octagonal grid used by the National Meteorological Center, Suitland, Md. The grid used in this report is part of this hemispheric grid (Fig. 26).

Then we frequently need a coordinate transformation which converts the geographical coordinates (λ, φ) of a point on earth into the Cartesian coordinates (x, y) and vice versa. Referring to Fig. 14 and Fig. 15 the mapping equations may easily be derived:

$$x = \frac{(1 + \cos \pi/6)\, a}{f\, M} \tan(\pi/4 - \varphi/2) \sin(\lambda_0 - \lambda),$$

$$y = \frac{(1 + \cos \pi/6)\, a}{f\, M} \{\tan(\pi/4 - \varphi_0/2) - \tan(\pi/4 - \varphi/2) \cos(\lambda_0 - \lambda)\} \qquad (5.17)$$

M is the scale magnification (e.g. $M = 30 \cdot 10^6$), f is a normalization factor to be used if the coordinates x and y are expressed in grid units. If measured in cm f is the grid spacing in cm.

For a we take the mean radius of the earth (international ellipsoid of reference):

$$a = 6371229.315 \text{ m}.$$

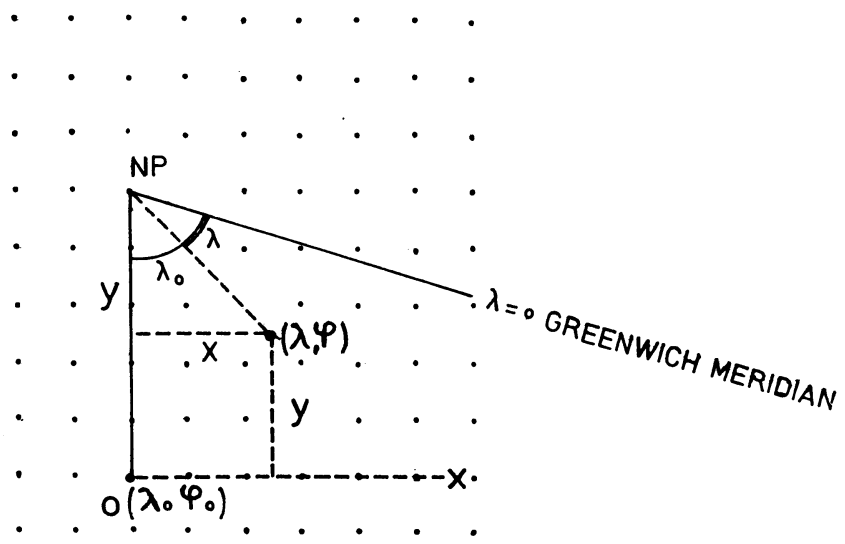

Fig. 14 Cartesian grid superposed on a polar stereographic projection map.

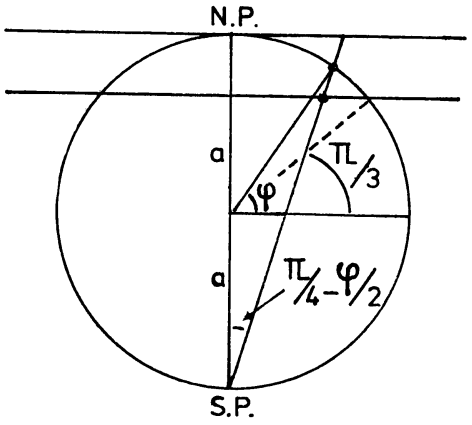

Fig. 15 Schematic illustration of the polar stereographic projection with true scale at 60° N.

The *inverse* mapping equations may be written as

$$\lambda = \lambda_0 - \arctan\,(x/(y - y_0)),$$

$$\varphi = \pi/2 - 2\arctan\{\frac{f\,M}{a\,(1+\cos\pi/6)}(x^2+(y-y_0)^2)^{\frac{1}{2}}\}, \tag{5.18}$$

where

$$y_0 = \frac{(1+\cos\pi/6)\,a}{f\,M}\tan(\pi/4 - \varphi_0/2)$$

is the chart distance of the origin to the pole.

We also need to convert map distances to greatcircle distances on earth.

For not too great distances in the order of the grid spacing use may be made of (4.2).

Let d_0 denote the grid spacing at latitude of true scale (60° N). In the octagonal grid $d_0 = 381$ km. Elsewhere the grid spacing d is then

$$d = \frac{d_0}{S}.$$

If a map distance s' is given in grid units the corresponding greatcircle distance s is then simply:

$$s = d\,s' = \frac{d_0 s'}{S}, \tag{5.19}$$

where S is computed in the midpoint of a segment.

For longer stretches, however, the distance between two points (λ_1, φ_1) and (λ_2, φ_2) should be taken from the goniometric relation

$$\cos(360s/2\pi a) = \sin\varphi_1 \sin\varphi_2 - \cos\varphi_1 \cos\varphi_2 \cos(\lambda_1 - \lambda_2). \tag{5.20}$$

5.4 Cruising system

The only basic parameter of control is the airspeed vector, or more specifically, the heading ξ. The airspeed is a function of position and time depending on the type of cruising system, for instance constant IAS, constant Mach Number.

The case of the constant Mach Number technique deserves special attention.

The Mach Number is by definition the ratio of the airspeed to the speed of sound c_s.

The speed of sound is related to the ambient temperature:

$$c_s = \left(\gamma \frac{R}{M} T\right)^{\frac{1}{2}}.$$

γ is the ratio of specific heat of air at constant pressure to that at constant volume;
R is the universal gas constant;
M is mean molecular weight of air;
T is temperature in K.

In the ISO Standard Atmosphere (formally called ICAO Standard Atmosphere [25]) we have for dry air:

$$\gamma = 1.40 \text{ (dimensionless)}$$

$$R = 8.31432\ 10^7 \text{ erg mol}^{-1}\ \text{K}^{-1}$$

$$M = 28.9644 \text{ (dimensionless)}$$

so that

$$c_s = 38.9826\, T^{\frac{1}{2}} \text{ kt.}$$

The airspeed c will therefore be given by

$$c = \text{MACH}_{\text{true}}\ c_s = 38.9826\ \text{MACH}_{\text{true}}\ T^{\frac{1}{2}} \text{ kt.} \tag{5.21}$$

The implication of this result is that the airspeed is directly related to the isobaric temperature pattern.

If not otherwise stated we shall from now on presume that the navigation is based on the Constant Mach Number Technique.

The temperature values are usually provided in the form of grid point values.

These are stored together with the wind information. From a point of view of computation speed it may be recommended to convert immediately these stored grid point values into the grid point values for the airspeed, using relation (5.21). This will facilitate the programming effort.

5.5 Operational meteorological data

The horizontal fields of meteorological parameters may be represented by their values in a number of grid points or by the coefficients of truncated expansions in orthogonal functions.

The last mentioned *spectral* representation would in principle be very convenient to use in the numerical computation schemes for the evaluation of the Least Time Track. Its closed analytical form admits a simple programming and the speed of computation is favourable with respect to the grid point representation.

Although the interest in the numerical integration of the hydrodynamic equations in terms of a spectral representation is growing steadily, the integration has until now been based almost entirely on the grid point representation. The grid point representation has been accepted as a standard practice for use in the automation of Flight Planning and Air Traffic Control.

Air Traffic Services, airlines and other aeronautical users have now developed statements of requirements, setting out the necessary details including format, contents, grid system, frequency of issue and validity times of the meteorological messages. The meteorological parameters required for the processing of the time optimal track comprise mainly the forecast upper winds and temperatures at preselected (standard) pressure surfaces.

In view of the dynamic character of the system it is desired to have the disposal of at least two sets of prognostic values in digital form for successive validity times not further apart than 12 hours. This allows the equation of state to be integrated in phase with the time changing atmospheric conditions.

Note:

The compatibility of the flight performance with the expected time varying conditions could improve if the evolution of the flight control could be matched with the intermediate results as provided by each integration step — 0,5-2 hours in advance — in a numerical weather prediction model.

Wind data may be supplied in the form of grid point values of the vector wind. Then it suffices to prepare spot values by interpolation with respect to the values of the wind components in the surrounding vertices of the grid.

For many reasons, however, it is preferable not to receive the wind data sec but to receive the values of the stream function ψ or the geopotential altitude H of a standard pressure level from which the wind may be obtained indirectly.

For instance, let $\psi(\mathbf{x}, t)$ be the stream function obtained by solution of the balance equation, then the wind components may be written as

$$w^1 = -\frac{\partial \psi}{\partial y},$$

$$w^2 = \frac{\partial \psi}{\partial x}. \tag{5.22}$$

Given the geopotential altitude $H(\mathbf{x}, t)$ for a constant pressure surface, the wind may be obtained by using the geostrophic approximation:

$$w_g^1 = -\frac{1}{2\omega \sin \varphi} \frac{\partial H}{\partial y},$$

$$w_g^2 = \frac{1}{2\omega \sin \varphi} \frac{\partial H}{\partial x}. \tag{5.23}$$

Here ω denotes the angular speed of the earth's rotation, φ is the latitude. The geopotential altitude H is expressed in geopotential meters.

Both (5.22) and (5.23) are easily to manipulate with, but one could object against its use that the approximation of the real wind may, under certain conditions, be poor, especially in the case of strong cyclonic curvature of the stream lines. This effect, however, is partly compensated for by smoothing operations in the numerical integration model. Replacing (5.22) and (5.23) by the gradient wind has experimentally been shown not to yield a marked improvement. Henceforth we shall maintain the use of the equations (5.22) and (5.23).

The Jacobian of the wind vector W which affects the optimal control predominantly in real practice, assumes in case of the balance wind (5.22) the form:

$$W^T = \begin{bmatrix} -\frac{\partial^2 \psi}{\partial x \partial y} & \frac{\partial^2 \psi}{\partial x^2} \\ -\frac{\partial^2 \psi}{\partial y^2} & \frac{\partial^2 \psi}{\partial x \partial v} \end{bmatrix}. \tag{5.24}$$

Using the geostrophic approximation (5.23):

$$W^T = \frac{1}{2\omega \sin \varphi} \begin{bmatrix} -\frac{\partial^2 H}{\partial x \partial y} & \frac{\partial^2 H}{\partial x^2} \\ -\frac{\partial^2 H}{\partial y_2} & \frac{\partial^2 H}{\partial x \partial y} \end{bmatrix} + \operatorname{cotan} \varphi \, \nabla \varphi \, . \, \mathbf{w}_g^T. \tag{5.25}$$

The second term on the right may be shown to have a vanishing small effect on the control compared with the first term. The contribution of this second term will from now on be discarded.

On closer examination of the system equations, cf. section 5.2, it is seen that their solution requires the evaluation of spot values of all constituent state variables at each step of the iteration process before integration may be performed.

These spot values comprise not only the zero order terms like wind, airspeed and scale of the mapping, but also higher order terms: Jacobian of the wind vector, the isobaric gradient of airspeed and the gradient of scale.

The basic atmospheric data are assumed to be provided in the form of grid point representations. Then, given the grid point representation of a horizontal parameter field $F(x, y, t)$ we need to determine the value of F and its partial derivatives $\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial^2 F}{\partial x^2}, \frac{\partial^2 F}{\partial y^2}$ and $\frac{\partial^2 F}{\partial x \partial y}$ in an arbitrary point.

This requires the development of suitable bivariate interpolation and finite difference expressions over the grid system.

Let at a given time t the parameter field F be represented by its grid point values $F_{i,j}$ in a rectangular grid G with vertices (i, j). Let h be the mesh-width in the x-direction, k the mesh-width in the y-direction. A spot value $F(P)$ at an arbitrary point

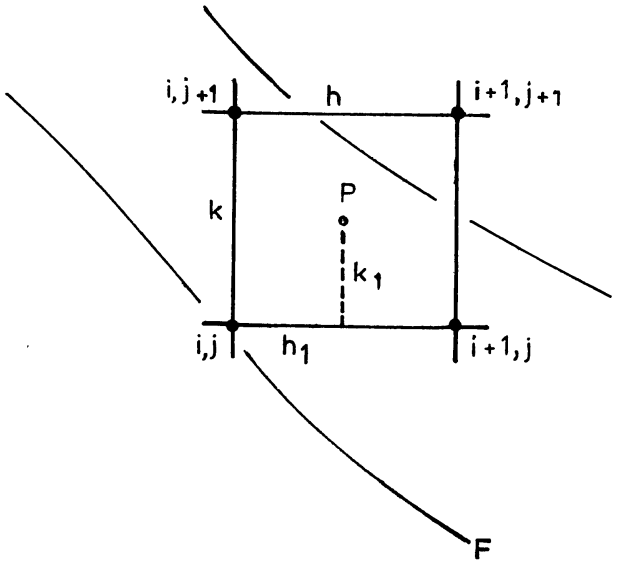

Fig. 16 Unit cell in a Cartesian grid for obtaining interpolation schemes.

$P(x, y)$ is then obtainable from the surrounding grid point values (Fig. 16) by using specific interpolation schemes [24].

In view of the accuracy attainable in present day computer forecast products simple interpolation formulas are acceptable like the following:

$$F(P) \simeq (1-\lambda)(1-\mu)F_{i,j} + (1-\lambda)\mu F_{i,j+1} + \lambda(1-\mu)F_{i+1,j} + \lambda\mu F_{i+1,j+1}. \tag{5.26}$$

Here

$$\begin{aligned} i &= \lfloor x/h \rfloor,^{*} \\ j &= \lfloor y/k \rfloor, \\ h_1 &= x - ih, \\ k_1 &= y - jk, \\ \lambda &= h_1/h, \\ \mu &= k_1/k. \end{aligned} \tag{5.27}$$

(5.27) specifies which unit cell in the grid contains the point P.

This 4-point bilinear interpolation formula may be used to compute the spot values of wind components, temperature, airspeed and geopotential altitude.

The formula might also be used in subroutines developed for computergraphic purposes, for instance to present upper air analyses via a line printer in the form of a character display. (See Fig. 24 and Fig. 25).

Finite difference expressions in terms of higher order derivatives of F may most efficiently be derived by applying the same 4-point formula to the grid point values of these derivatives over surrounding vertices. This means that the $F_{i,j}$ should be replaced by values in finite difference form itself.

Spot values of $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ may best be derived from the grid point values in a grid G' which is congruent with G and centred with respect to G (Fig. 17).

Let $F_{i,j}$ now refer to the derivative $\frac{\partial F}{\partial x}$, then a grid point value in the point $(i + \frac{1}{2}, j + \frac{1}{2})$ may be approximated by:

$$F_{i+\frac{1}{2},\, j+\frac{1}{2}} \simeq \frac{1}{2h}(F_{i+1,j+1} + F_{i+1,j} - F_{i,j+1} - F_{i,j}). \tag{5.28}$$

* $[x/h]$ signifies the next integer $\leq x/h$.

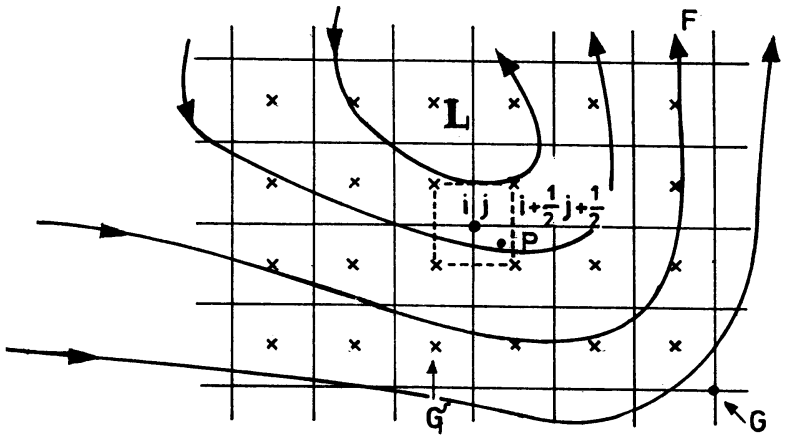

Fig. 17 Cartesian grid G together with the centred Cartesian grid G'.

In G' the unit cell which contains P is specified by:

$$
\begin{aligned}
i - \tfrac{1}{2} &= [x/h + \tfrac{1}{2}] - \tfrac{1}{2}, \\
j - \tfrac{1}{2} &= [y/k + \tfrac{1}{2}] - \tfrac{1}{2}, \\
h_1 &= x - (i - \tfrac{1}{2})h, \\
k_1 &= y - (j - \tfrac{1}{2})k, \\
\lambda &= h_1/h, \\
\mu &= k_1/k.
\end{aligned}
$$

Interchanging i by $i - \frac{1}{2}, j$ by $j - \frac{1}{2}$ in (5.26), substitution for (5.28) and rearranging terms gives

$$
\begin{aligned}
\left(\frac{\partial F}{\partial x}\right)_P \simeq{}& \frac{\mu}{2h}(F_{i,j+1} + \lambda F_{i+1,j+1} + (1-\lambda)F_{i-1,j+1}) \\
&- \left(\frac{1-\mu}{2h}\right)(F_{i,j-1} + \lambda F_{i+1,j-1} + (1-\lambda)F_{i-1,j-1}) \\
&+ \left(\frac{1-2\mu}{2h}\right)(F_{i,j} + \lambda F_{i+1,j} + (1-\lambda)F_{i-1,j}). \qquad (5.29)
\end{aligned}
$$

Similar for the y-component

$$\left(\frac{\partial F}{\partial y}\right)_P \simeq \frac{\lambda}{2k}(F_{i+1,j} + \mu F_{i+1,j+1} + (1-\mu)F_{i+1,j-1})$$
$$- \left(\frac{1-\lambda}{2k}\right)(F_{i-1,j} + \mu F_{i-1,j+1} + (1-\mu)F_{i-1,j-1})$$
$$+ \left(\frac{1-2\lambda}{2k}\right)(F_{i,j} + \mu F_{i,j+1} + (1-\mu)F_{i,j-1}). \tag{5.30}$$

These 9-point formulas may be used to determine the spot values of the isobaric gradient of the airspeed, the balance wind (5.22) or the geostrophic wind approximation (5.23). The $F_{i,j}$ are depicted from the grid point representations c_{ij}, ψ_{ij} resp. H_{ij}.

Note:
The gradient of a scalar quantity may be expressed in simpler finite difference schemes, but these lack symmetry and give rise to sharper discontinuities when passing from one unit cell to another.

The Jacobian of the wind vector in the form (5.24) or (5.25) requires the formulation of an additional interpolation scheme in terms of the second order derivatives of F.

To that aim we consider again the grid point representation $F_{i,j}$ in grid G.

Then to obtain an estimate for $\frac{\partial^2 F}{\partial x^2}$ we replace the F_{ij} value by:

$$F_{i,j} \to \frac{1}{h^2}(F_{i-1,j} + F_{i+1,j} - 2F_{ij}). \tag{5.31}$$

etc.

The unit cell in G which contains P is specified by:

$$i = \lfloor x/h \rfloor.$$
$$j = \lfloor y/k \rfloor,$$
$$h_1 = x - ih,$$
$$k_1 = y - jk,$$
$$\lambda = h_1/h,$$
$$\mu = k_1/k.$$

Substitution of (5.31) into (5.26) yields

$$\left(\frac{\partial^2 F}{\partial x^2}\right)_P \simeq \frac{\lambda\mu}{h^2}(F_{i+2,j+1} + F_{i,j+1} - 2F_{i+1,j+1}) +$$

$$\frac{(1-\lambda)\mu}{h^2}(F_{i+1,j+1} + F_{i-1,j+1} - 2F_{i,j+1}) +$$

$$\frac{\lambda(1-\mu)}{h^2}(F_{i+2,j} + F_{i,j} - 2F_{i+1,j}) +$$

$$\frac{(1-\lambda)(1-\mu)}{h^2}(F_{i+1,j} + F_{i+1,j} - 2F_{i,j}). \tag{5.32}$$

Similarly, replacing $F_{i,j}$ by an estimate of $\frac{\partial^2 F}{\partial y^2}$:

$$F_{i,j} \to \frac{1}{k^2}(F_{i,j+1} + F_{i,j-1} - 2F_{i,j})$$

and entering into (5.26) gives:

$$\left(\frac{\partial^2 F}{\partial y^2}\right)_P \simeq \frac{\lambda\mu}{k^2}(F_{i+1,j+2} + F_{i+1,j} - 2F_{i+1,j+1}) +$$

$$\frac{(1-\lambda)\mu}{k^2}(F_{i,j+2} + F_{i,j} - 2F_{i,j+1}) +$$

$$\frac{(1\lambda-\mu)}{k^2}(F_{i+1,j+1} + F_{i+1,j-1} - 2F_{i+1,j}) +$$

$$\frac{(1-\lambda)(1-\mu)}{k^2}(F_{i,j+1} + F_{i,j-1} - 2F_{i,j}). \tag{5.33}$$

Finally, replacing F_{ij} by an estimate of $\frac{\partial^2 F}{\partial x \partial y}$:

$$F_{ij} \to \frac{1}{4hk}(F_{i+1,j+1} + F_{i-1,j-1} - F_{i-1,j+1} - F_{i+1,j-1})$$

and entering into (5.26) affords:

$$\left(\frac{\partial^2 F}{\partial x \partial y}\right)_P \simeq \frac{\lambda\mu}{4kh}(F_{i+2,j+2} + F_{i,j} - F_{i,j+2} - F_{i+2,j}) +$$

$$\frac{(1-\lambda)\mu}{4kh}(F_{i+1,j+2} + F_{i-1,j} - F_{i-1,j+2} - F_{i+1,j}) +$$

$$\frac{\lambda(1-\mu)}{4kh}(F_{i+2,j+1} + F_{i,j-1} - F_{i,j+1} - F_{i+2,j-1}) +$$

$$\frac{(1-\lambda)(1-\mu)}{4kh}(F_{i+1,j+1} + F_{i-1,j-1} - F_{i-1,j+1} - F_{i+1,j-1}). \tag{5.34}$$

The expressions (5.32), (5.33) and (5.34) are vital for the computation of the elements of the matrix W in the form (5.24) or (5.25).

Since the parameter fields F are in general dependent on time the numerical integration of the system equations will necessitate the issue of at least two forecast grid point representations. The spot values of all contributing terms in the system equations should then be evaluated in both these representations followed by (linear) interpolation with respect to the time variable.

Due to the linearization this procedure may be simplified by first interpolating the F_{ij} values with respect to time followed by a space interpolation using one of the above finite difference expressions in terms of the new interpolated F_{ij} values.

It is inevitable that the use of the grid point interpolation formulas introduces some smoothing, in particular when the forecasts are issued at intervals exceeding 12 hours.

Since the grid system is assumed to be superposed on a (conformal) projection map and the quantities should be measured on the earth's sphere the grid spacings h and k should be replaced by the arc distances, see (5.19) or (5.20).

A quantity which remains to be examined is the scale of the mapping and its gradient.

In the polar stereographic projection the values for S and ∇S follow immediately from (5.1):

$$S = \frac{1.86603}{1 + \sin\varphi},$$

$$\nabla S = -\frac{1.86603}{a(1 + \sin\varphi)^2}\cos\varphi\,\mathbf{j}, \tag{5.35}$$

where $\mathbf{j}$ is a unit vector parallel to a meridian pointing to the Pole.

Thus:

$$\frac{\nabla S}{S} = -\frac{\cos\varphi}{a(1+\sin\varphi)}\,\mathbf{j}. \tag{5.36}$$

5.6 Computer implementation of the Least Time Track construction

In section 5.2 a general outline was given of the iteration method to be applied to the evaluation of the Least Time Track. In this section we shall discuss the most important features of this iteration process especially in regard to computer application.

The synthesis will be based on the interpretation of the equation of motion in conjunction with one of the control equations. The use of the refraction principle will be the subject of a separate treatment. (section 5.7).

On closer examination of the iteration process it is observed that each cycle consists of a simulation of a flight along an optimal trajectory starting from the departure point at a prescribed initial time. Since the system equations have a predictive nature each such cycle may be built up of integration steps. At each integration step the system equations should be evaluated and integrated over a small time interval into the future.

Let at the j-th step the integration have resulted in the determination of the next position P_j at time t_j under influence of the control $\mathbf{c}_j$. (Fig. 18.) Then, in order to

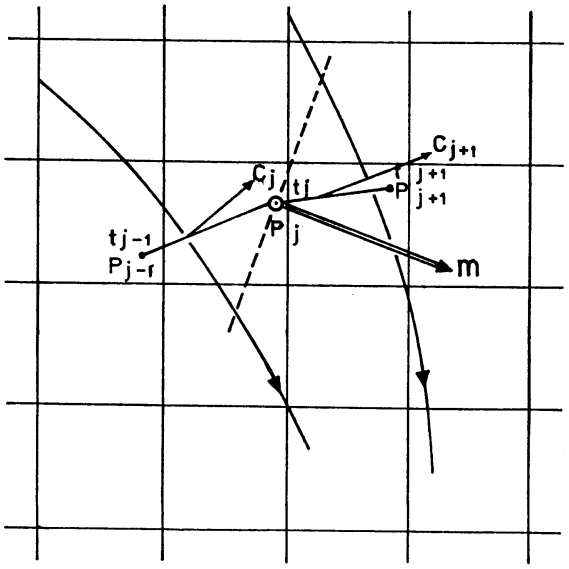

Fig. 18 Portion of track resulting from an integration step.

locally evaluate the system equations in the next step, we need first to compute the refraction normal $\mathbf{m}$ or $\mathbf{m}'$ in point $P = P_j$. For this purpose it is required to determine the spot values at P of the wind vector $\mathbf{w}_P$, the airspeed c_P, its gradient $(\nabla c)_P$, the scale of the mapping S_P and its gradient $(\nabla S)_P$.

In addition, we need to calculate the airspeed vector $\mathbf{c}_P$, which is assumed to point into the direction of the control $\mathbf{c}_j$ in the last step:

$$\mathbf{c}_P = c_P \frac{\mathbf{c}_j}{c_j}.$$

The effective airspeed in P is then:

$$(c_{\text{eff}})_P = c_P + \frac{1}{c_j} \mathbf{c}_j^{\mathrm{T}} . \mathbf{w}_P.$$

Under the assumption that the parameter fields are made available in the form of grid point representations, preferably for at least two validity times, these spot values may be derived from the interpolation and/or finite difference schemes as given in the previous section, including interpolation with respect to time.

Entering these spot values into (5.4) yields:

$$\mathbf{m}_P \simeq \frac{1}{c_P} W_P^T . \mathbf{c}_P + (\nabla c)_P + (c_{\text{eff}})_P \left(\frac{\nabla S}{S} \right)_P - \frac{(\mathbf{c}_P + \mathbf{w}_P)^T . (\nabla S)_P}{S_P} \frac{\mathbf{c}_P}{c_P}. \quad (5.37)$$

The evaluation of this refraction normal along the trajectory is a key element in the iteration process.

The next action involves the computation of the incremental change of the system control by solving one of the control equations by means of a quadrature formula, which does not include the unknown variables. Whichever equation is used for this purpose the result may be shown always expressible in terms of the airspeed vector $\mathbf{c}$, which has the property to control the motion directly via *Eq.* (5.8).

Let us first discuss the way to establish the required quadrature formulae for the various control equations.

Gradient equation

For a sufficiently small time increment Δt the integration of (5.9) gives:

$$(\nabla \tau)_{j+1} \simeq (\nabla \tau)_j - \frac{\Delta t}{(c_{\text{eff}})_P} \mathbf{m}_P. \quad (5.38)$$

As by the principle of optimality

$$\nabla\tau = \frac{1}{c_{\text{eff}}} \frac{\mathbf{c}}{c},$$

the control $\mathbf{c}_{j+1}$ in step $j+1$ might be expressed in terms of $(\nabla\tau)_{j+1}$:

$$\mathbf{c}_{j+1} = c_P(c_{\text{eff}})_P(\nabla\tau)_{j+1}. \tag{5.39}$$

Phase velocity equation

Integration of (5.10) over a small time interval Δt affords:

$$(\mathbf{c}_{\text{eff}})_{j+1} \simeq (\mathbf{c}_{\text{eff}})_j + (c_{\text{eff}})_P C_P \mathbf{m}_P \Delta t. \tag{5.40}$$

If c^1, c^2 are the components of $\mathbf{c}_P$ the matrix C_P in point P takes the form, refer to (5.3):

$$C_P = \frac{1}{c_P^2}\begin{bmatrix} c^1c^1 - c^2c^2 & 2c^1c^2 \\ 2c^1c^2 & c^2c^2 - c^1c^1 \end{bmatrix}.$$

As the *optimal* control vector $\mathbf{c}$ is pointing in the same direction as $\mathbf{c}_{\text{eff}}$ the control vector $\mathbf{c}_{j+1}$ in the $j+1$th step may be found by putting

$$\mathbf{c}_{j+1} \simeq \frac{c_P}{(c_{\text{eff}})_{j+1}}(\mathbf{c}_{\text{eff}})_{j+1}. \tag{5.41}$$

Steering equation and phase speed equation

In principle the steering equation and phase speed equation should be taken together being component equations of the adjoint system.

Let the steering angle ξ_j be known in the j-th step. Its value in the next step is then according to *Eq.* (5.11):

$$\xi_{j+1} \simeq \xi_j - \frac{\mathbf{c}_P{}^{*\mathrm{T}}}{c_P} \cdot \mathbf{m}_P \Delta t. \tag{5.42}$$

Given the efftective airspeed $(c_{\text{eff}})_j$ at the jth step the phase speed equation (5.12) yields:

$$(c_{\text{eff}})_{j+1} \simeq (c_{\text{eff}})_j + (\mathbf{c}_{\text{eff}})_P^T \cdot \mathbf{m}_P \Delta t. \tag{5.43}$$

The steering angle ξ_{j+1} and effective airspeed $(c_{\text{eff}})_{j+1}$ determine uniquely the airspeed vector $(\mathbf{c}_{\text{eff}})_{j+1}$ in the $j+1$th step:

$$(\mathbf{c}_{\text{eff}})_{j+1} \simeq ((c_{\text{eff}})_{j+1} \cos \xi_{j+1}, (c_{\text{eff}})_{j+1} \sin \xi_{j+1}). \tag{5.44}$$

Then, finally to obtain $\mathbf{c}_{j+1}$ in the $j+1$th step one can apply formula (5.41).

At this stage the question may be raised whether the optimal control may also be described by one of the component equations only. For the steering equation the answer is in the affirmative, for by approximation the controlling vector $\mathbf{c}_{j+1}$ may immediately be written down:

$$\mathbf{c}_{j+1} \simeq (c_P \cos \xi_{j+1}, c_P \sin \xi_{j+1}).$$

Using the phase speed equation, $\mathbf{c}_{j+1}$ might be derived from equation (5.2):

$$(c_{\text{eff}})_{j+1} \simeq c_P + \frac{\mathbf{c}_{j+1}^T}{c_{j+1}} . \mathbf{w}_P.$$

This relation affords *two* solutions for $\mathbf{c}_{j+1}$ of which it is impossible to decide which solution is the correct one without use of some additional information concerning the incremental change of the heading.

This finding leads to the conclusion that computer programmes may be developed for each control equation separately except for the phase speed equation!

Once the system control is known for the next step the response of it on the motion will follow from *Eq.* (5.8):

$$\Delta \mathbf{x} \simeq (\mathbf{w}_P + \mathbf{c}_{j+1}) \Delta t. \tag{5.45}$$

Next position P_{j+1} and corresponding transit time t_{j+1} follow from

$$\mathbf{x}_{j+1} \simeq \mathbf{x}_j + \Delta \mathbf{x}, \tag{5.46}$$

$$t_{j+1} = t_j + \Delta t. \tag{5.47}$$

The process is repeated in order to advance the terminal point. The elaboration of the quadrature formulae was rather straightforward, but it is clear that one could introduce various refinements in the calculation, for instance by computing the spot values of the wind in the midst of a predicted arc element in the next step rather than at the endpoint of the previous step.

The integration is carried out in terms of a small (fixed or not) time interval Δt at each step. The quadrature could also have been performed using a small space

interval $\Delta \mathbf{x}$. This may be of use in case a computer programme is developed, which provides also the visual representation of the results via an incremental plotter (or other graphical display unit). Then a fixed space interval $|\Delta \mathbf{x}|$ may be chosen being of the order of a few steps in the incremental plotter in order to assure a smooth picture drawing.

In the iteration process each cycle is started after a proper choice has been made for the initial state of control. The selection of the initial control is of prime importance, in view of the speed of convergence of the iteration process.

The termination of the process should be decided upon by the behaviour of the solution in the vicinity of the terminal point Q. (Fig. 19). Whether the iteration process should be brought to a halt or be continued depends on the 'distance' of the point of closest approach. In the case that the terminal point is reached within a certain tolerance limit ∂ the process is terminated, otherwise a new cycle is activated.

Bypassing of the terminal point $\mathbf{x}_Q$ may be depicted from a change in sign of the vector product:

$$\dot{\mathbf{x}}_{P_n}^T \cdot (\mathbf{x}_Q - \mathbf{x}_{P_n}), \tag{5.48}$$

where $\dot{\mathbf{x}}_{P_n}$ is the *ground speed* vector at the endpoint P_n of the n-th step. As soon as the switching of sign is detected the computer calculates the distance function d:

$$d = \frac{\mathbf{k} \times \dot{\mathbf{x}}_{P_n}^T \cdot (\mathbf{x}_Q - \mathbf{x}_{P_n})}{|\dot{\mathbf{x}}_{P_n}|}. \tag{5.49}$$

($\mathbf{k}$ is a unit vector normal to the projection plane).
$|d|$ represents the distance of the point of nearest approach to the terminal point.

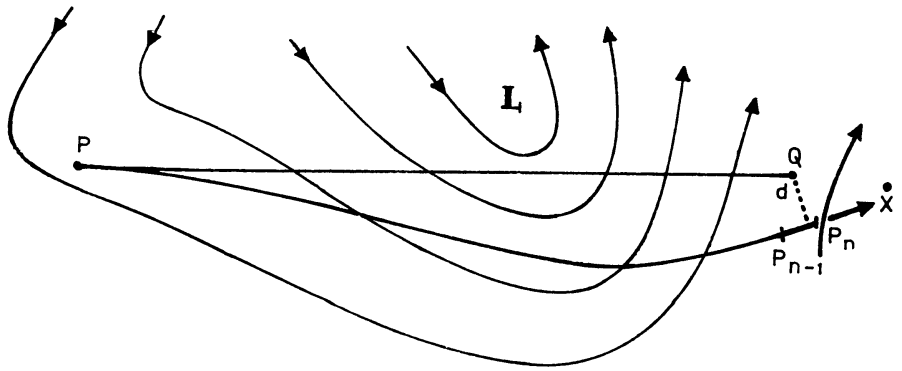

Fig. 19 Point of closest approach at the terminal point for a trial solution in an iterative process.

If $|d| \geqslant \partial$ the process is continued by starting a new cycle. Otherwise the process is stopped.

In practice, however, it may be suitable in that case to restart the last cycle in order to calculate and present all detailed information of the final optimal track.

The very nature of the iteration scheme presupposes a proper selection of the initial state already at the start. The first choice may be a vector $\mathbf{c}_P$ at P pointing to the terminal point:

$$\mathbf{c}_P = c_P \frac{\mathbf{x}_Q - \mathbf{x}_P}{|\mathbf{x}_Q - \mathbf{x}_P|}, \tag{5.50}$$

where c_P is the spot value of the airspeed in the point of departure.

Another choice may be based on the well-known drift formula of Bellamy associated with the technique of *single heading navigation.*

To that aim, let ψ be the stream function in a fixed time prognostic chart. Let ψ_P and ψ_Q denote the values of ψ in the endpoints (Fig. 20). Further, let PQ be the straight line connection on the projection plane with arc distance D, then the drift formula of Bellamy reads:

$$\sin \alpha = \frac{\psi_Q - \psi_P}{cD}. \tag{5.51}$$

α is the drift angle between $\mathbf{c}$ and $\overline{PQ}$.

After the first cycle is completed without a 'near target hit' the next and following search cycles are activated by correcting the previous initial control(s). It is logical then to correct $\mathbf{c}_P$ in dependency of the distance function d, formula (5.49).

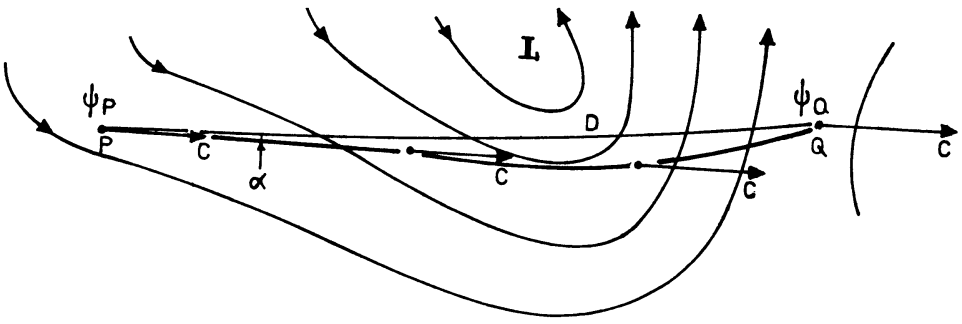

Fig. 20 Schematic illustration of a single heading flight operation.

Examples:

a Let ξ_{i-1} be the steering angle corresponding to the initial control $\mathbf{c}_{i-1}$ at the start of the $(i-1)$th cycle. Then we may put

$$\xi_i = \xi_{i-1} + \frac{d}{|\mathbf{x}_Q - \mathbf{x}_P|} \tag{5.52}$$

for the initial control $\mathbf{c}_P$ ($c_P \cos \xi_i$, $c_P \sin \xi_i$) in the next cycle.

b A technique which is frequently used in modern computer algorithms is Newton's *regula falsi* method.

Suppose that the $(i-1)$th cycle commences with the initial steering angle ξ_{i-1} and that the terminal point is bypassed at a 'perigee' distance $|d_{i-1}|$. Let the i-th cycle have started with the initial steering angle ξ_i and let $|d_i|$ be the perigee distance in this cycle. Then the next cycle $i+1$ should commence with the steering angle:

$$\xi_{i+1} = \frac{d_i \xi_{i-1} - d_{i-1} \xi_i}{d_i - d_{i-1}} \tag{5.53}$$

and corresponding initial control $\mathbf{c}_P$ ($c_P \cos \xi_{i+1}$, $c_P \sin \xi_{i+1}$). This procedure obviously requires two independently assigned values for ξ in the first two trial solutions.

Failure of convergence is not ruled out, in particular if there exists no simple relationship between the perigee distance and the initial control. This may happen for instance if the optimal trajectories tend to diverge strongly in the vicinity of the terminal point.

The systematic algebraic schemes can readily be programmed. Numerical schemes may be developed for the gradient equation, phase velocity equation or steering equation. The last possibly combined with the phase speed equation.

The programmes may also be used to generate *single heading tracks* between a given pair of endpoints. This may be accomplished by keeping the initial heading $\mathbf{c}_P$ fixed with respect to the Cartesian coordinate system in the projection plane or with respect to the local meridians.

Applied to single heading flights the speed of convergence should be rapid: 1 to 3 cycles are satisfactory.

To check the correct functioning of a programme it may be brought to execution for some specialized cases. Program execution in the case of zero wind conditions should result in the production of the *greatcircle track* in the projection map. When the single heading track is evaluated in terms of a constant heading with respect to the meridians in zero wind conditions the computer should produce a loxodrome.

5.7 Use of the refraction formulae

To make the solution of the navigation problem accessible for computer application it was required to discretisize the system parameters and system equations. The

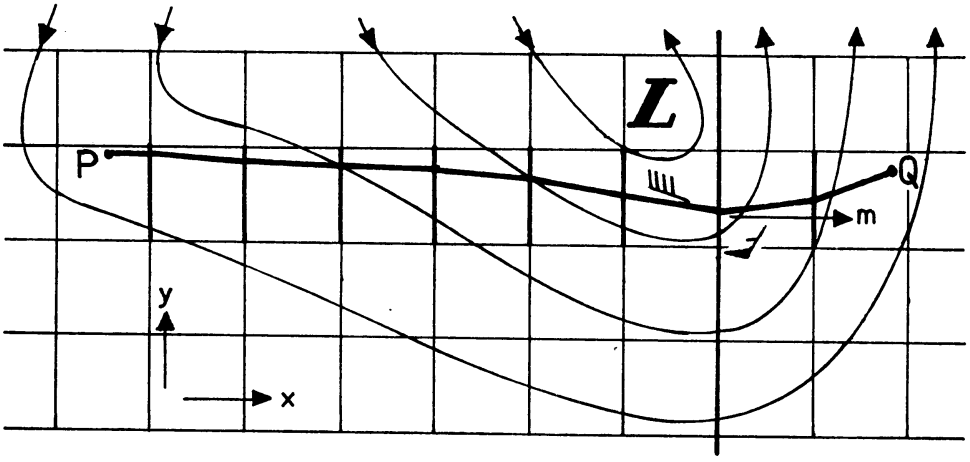

Fig. 21 Use of 'meteorological' grid lines as lines of discontinuity in a mode of refraction.

system equations may also be discretisized in their own right, without reference to automata. In this respect the formulation of the refraction formulae forms a good example and from a conceptual viewpoint it might look that the refraction formulae are particularly suited to be processed by using high speed digital computers. An analysis of the method reveals however that its applicability is rather restricted. An over-simplification may easily cause a misinterpretation of the desired solution [26]. To what this may lead may be illustrated by the following example:

Let by way of approximation be assumed that Von Mises' refraction formula (5.13) be applied with respect to the y-grid lines of the meteorological grid used as lines of discontinuity (Fig. 21).

Let further be assumed that the windfield be uniform locally on both sides of a gridline.

It may then be expected that a step by step integration by jumping from gridline to gridline in an iteration process will produce a fairly accurately positioned Least Time Track between the terminals P and Q.

Computer experiments clearly show however that this method under circumstances may cause marked deviations from the real solution. This discrepancy may be explained by the following analysis.

Consider a square grid in a plane surface ($S \equiv 1$). The y-gridlines are assumed to represent the refraction lines in terms of which the refraction formula due to Von Mises may be applied.

In addition, we assume that the airspeed c is constant in the whole area of interest.

Then the wind must be approximated by

$$w^1 = f(x, t),$$
$$w^2 = g(x, t).$$

The Jacobian of the wind vector becomes:

$$W^T = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial g}{\partial x} \\ 0 & 0 \end{bmatrix}.$$

The refraction normal, cf. (5.4), reads

$$\mathbf{m} = \frac{1}{c} W^T . \mathbf{c} = \begin{bmatrix} \frac{\partial f}{\partial x} \cos \xi + \frac{\partial g}{\partial x} \sin \xi \\ 0 \end{bmatrix}.$$

m is normal to the y-grid lines, as expected.
The steering equation (5.11) takes the form

$$\dot{\xi} = -\frac{1}{c} \mathbf{c}^{*\mathbf{T}} . \mathbf{m} = [\sin \xi, -\cos \xi] \begin{bmatrix} \frac{\partial f}{\partial x} \cos \xi + \frac{\partial g}{\partial x} \sin \xi \\ 0 \end{bmatrix},$$

or

$$\dot{\xi} = \sin \xi \left(\frac{\partial f}{\partial x} \cos \xi + \frac{\partial g}{\partial x} \sin \xi \right).$$

For $\xi = 0$ this equation reduces to

$$\dot{\xi} = 0.$$

stating that the Least Time Track degenerates into the single heading track.

In practice the orientation of the grid system is chosen such that the x-grid lines are mainly parallel to the main stream of traffic. Then the control will also be mainly parallel to the x-grid lines ($\xi \simeq 0$). This implies that the proposed method will produce an *approximation to the single heading track instead of the desired Least Time Track.*

The misinterpretation lies in the choice of the refraction normal, which was assumed

to be constantly pointing in the x-direction. From the theory it is known however that this normal continuously varies all the way along the track.

To avoid a misconception of the refraction principle it is advisable not to attach too much importance to the connection of the refraction principle with the grid geometry.

The computational algorithm in terms of the refraction formulae may ultimately be elaborated along the same lines as described in the previous sections using an iterative searching technique for the solution in the 'continuous system'.

Let $\mathbf{c}_j$ be the control at the j-th step (Fig. 22). Integration of (5.8) in the j-th step gives the segment $P_{j-1}P_j$. Then we may compute the refraction normal $\mathbf{m}$ from (5.37) in point P_j. The refraction will be considered with respect to a line of discontinuity normal to $\mathbf{m}$. Applying the first mode of refraction the refraction formula (5.13) may be written:

$$K = \frac{\sin i_j}{c_j + w_j^1 \cos i_j + w_j^2 \sin i_j} = \frac{\sin i_{j+1}}{c_{j+1} + w_{j+1}^1 \cos i_{j+1} + w_{j+1}^2 \sin i_{j+1}}. \quad (5.54)$$

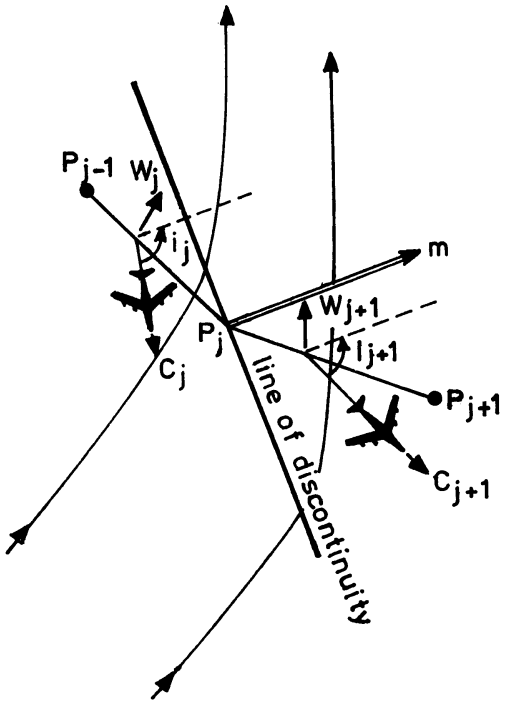

Fig. 22 Portion of track determined by an integration step using a refraction formula.

Here the angle of incidence i_j is the angle between $\mathbf{c}_j$ and $\mathbf{m}$ measured in counter clockwise direction. For c_j, w_j^1, w_j^2 the spot values may be taken in the midst of segment $P_{j-1}P_j$. c_{j+1} and w_{j+1}^1, w_{j+1}^2 may be determined in the midst of a first estimate of the segment P_jP_{j+1}. Thus the only unknown variable in (5.54) is the angle of refraction i_{j+1} between $\mathbf{c}_{j+1}$ and $\mathbf{m}$.

The equation to be solved is a goniometric equation, quadratic in $\sin i_{j+1}$.

The equation admits therefore *two solutions* for the refraction angle:

$$(\sin i_{j+1})_{1,2} = \frac{(K - w_j^2)\, c_j \pm w_j^1 ((K - w_j^2)^2 + (w_j^1)^2 - c_j^2)^{\frac{1}{2}}}{(K - w_j^2)^2 + (w_j^1)^2}. \tag{5.55}$$

This finding expresses the property that Von Mises' refraction formula in a sense describes a form of *double refraction.*

The roots may coincide or may become both real or complex. For real roots one of these should be discarded as being physically unrealistic. Which of the roots should be expelled might be inferred from the requirement that by transit of the interface the refracted 'ray element' P_jP_{j+1} in the 'second medium' should be a continuation of the incident ray element $P_{j-1}P_j$ in the first medium. This may also be expressed by the condition that the components of the ground speed vector normal to the line of discontinuity in both media should have equal sign:

$$\operatorname{sign}(c_{j+1} \cos i_{j+1} + w_{j+1}^1) = \operatorname{sign}(c_j \cos i_j + w_j^1).$$

The case of equal roots corresponds with the optical phenomenon that the refracted ray element P_jP_{j+1} is parallel to the interface.

For roots becoming both complex, i.e. for:

$$(K - w_j^2)^2 + (w_j^1)^2 - c_j^2 < 0, \tag{5.56}$$

the integration process will be cut off. This very unsatisfactory situation is typical of the solution of the navigation problem using one of the modes of refraction. It is caused by the particular type of discretization applied to the governing control equation. The case of complex roots may occur at portions of the track where the change in the control vector is most pronounced and the refraction normal happens to become nearly perpendicular to the control vector $\mathbf{c}$. It is not easy to remove this abrupt termination in the integration process. The introduction of smaller steps is no remedy, for it increases the risk of obtaining complex roots.

The best strategy seems to be a perforce rotation of the refraction normal about a (fixed) angle Δi_j such that the roots of equation (5.55) resume real values and the step by step integration may be continued. But this measure goes to the expense of

the precision of the construction, as the integration process is re-established at the most sensitive portions of the desired track.

After the angle of refraction i_{j+1} is known one may compute $\mathbf{c}_{j+1}$ from c_{j+1} and i_{j+1} and continue the process with computation of the segment P_jP_{j+1}. Repetition of this scheme will finally result in a solution by iteration.

The first mode of refraction may also be applied in terms of the refraction normal $\boldsymbol{\mu}$, cf. (5.6) together with the refraction formula (5.14). The elaboration of this alternative method meets with the same difficulties in practice as in the foregoing method.

The second mode of refraction in the form (5.15) or (5.16) becomes even more complicated to process than the first, due to the fact that the refraction formula turns out to be a goniometric equation of the *fourth* degree in $\sin i_{j+1}$.

5.8 Computational aspects of the graph method

We now turn to the completely different sort of problem involving the search for the shortest path in the 'navigation graph'. The nodes and arcs in this navigation graph may either be defined in the free air space or be prescribed by the elements of a given route pattern.

If for each arc the arc distance d_{ij} between a pair of nodes i and j would be a fixed value, the matrix procedure as explained in section 3 would be very convenient to use by digital computer. In practice, however, the arc distance may be variable due to the non-stationary character of the airflow. Then it is recommended to use an algorithm as indicated by the scheme (3.4). This algorithm requires a proper ordering of the graph points, for instance by arranging the nodes on the y-grid lines of the meteorological grid (Fig. 31) or a set of concentric circles about the departure point (Fig. 32).

The procedure is then centred on the computation of the arc distance d_{ij}.

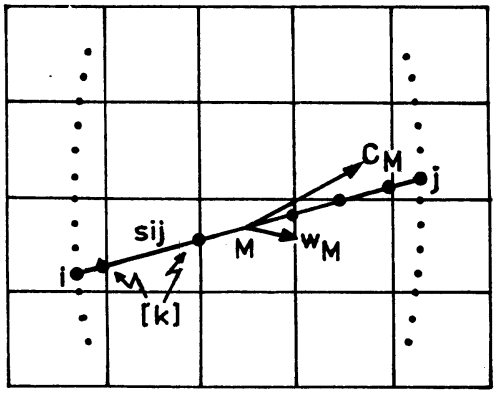

Fig. 23 Arc between two nodes of a graph within a Cartesian grid.

Consider an orthogonal system Oxy with coordinate axis parallel to the grid lines of a Cartesian grid superposed on a conformal chart projection. In the case of time optimal control the 'arc distance' d_{ij} will represent the flight time t_{ij} between a pair of nodes. Let us assume that any arc is traversed as a single heading flight by keeping the heading ξ constant with respect to the x-axis. This heading may be derived from Bellamy's drift equation (5.51). For stretches of a few hundred miles we may approximate the flown track by a greatcircle arc of length s_{ij} (Fig. 23). Let M be the midpoint of the arc, T_M the spotvalue of the temperature at M, c_M the corresponding value of the airspeed and w_M the wind at the point M. In order to include the effect of the time variability of winds and temperature it is advisable to determine a time estimate t_M for reaching the midpoint M. All these spotvalues may be obtained by using one of the interpolation formulae of section 5.5.

The flight time t_{ij} becomes approximately:

$$t_{ij} \simeq \frac{s_{ij}}{|\mathbf{c}_M + \mathbf{w}_M|} \tag{5.57}$$

$$\text{with } \mathbf{c}_M = (c_M \cos \xi, c_M \sin \xi).$$

For longer stretches the flight time may be determined more accurately by considering the intersection points $\{k\}$ of the straight line connection with the grid lines.

Then, if $\mathbf{c}_{M;k}$, $\mathbf{w}_{M;k}$ represent the values in the midpoints of the subarcs we obtain:

$$t_{ij} = \sum_{(k)} t_{ij;k} = \sum_{(k)} \frac{s_{ij;k}}{|\mathbf{c}_{M;k} + \mathbf{w}_{M;k}|}. \tag{5.58}$$

Representative values for the wind and airspeed may be found by taking a weighted mean over the sub arcs in terms of the flight time:

$$\mathbf{w}_{i,j} = \left(\frac{1}{t_{ij}} \sum_{(k)} t_{ij;k}\, w^1_{M;k}, \frac{1}{t_{ij}} \sum_{(k)} t_{ij;k}\, w^2_{M;k} \right), \tag{5.59}$$

$$c_{ij} = \frac{1}{t_{ij}} \sum_{(k)} t_{ij;k}\, c_{M;k}. \tag{5.60}$$

Given the 'arc distance' t_{ij} the shortest path problem may be solved simply by application of the scheme for optimization (3.4) after performing a suitable ordering of the nodes.

In case provision is made for the wind data sec in the form of a grid point representation the controlling airspeed cannot be derived from a single drift formula. The airspeed c_M is then found by the condition that the ground speed vector is directed along an arc:

$$\frac{\mathbf{w}_M + \mathbf{c}_M}{|\mathbf{w}_M + \mathbf{c}_M|} = \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|}. \tag{5.61}$$

The graph algorithm is easy to programme. The method is principally one of track selection albeit that this selection takes place in a vast multitude of admissible paths. The number of tracks runs in the millions with increasing density of nodes.

The increase of a well ordered point set in the graph will also promote the shortest path to converge to the 'exact' solution.

An improved approximation might in practice however be hindered by a limitation of available storage capacity and computer running time. In the next section we will mention some additional advantages and disadvantages of this method.

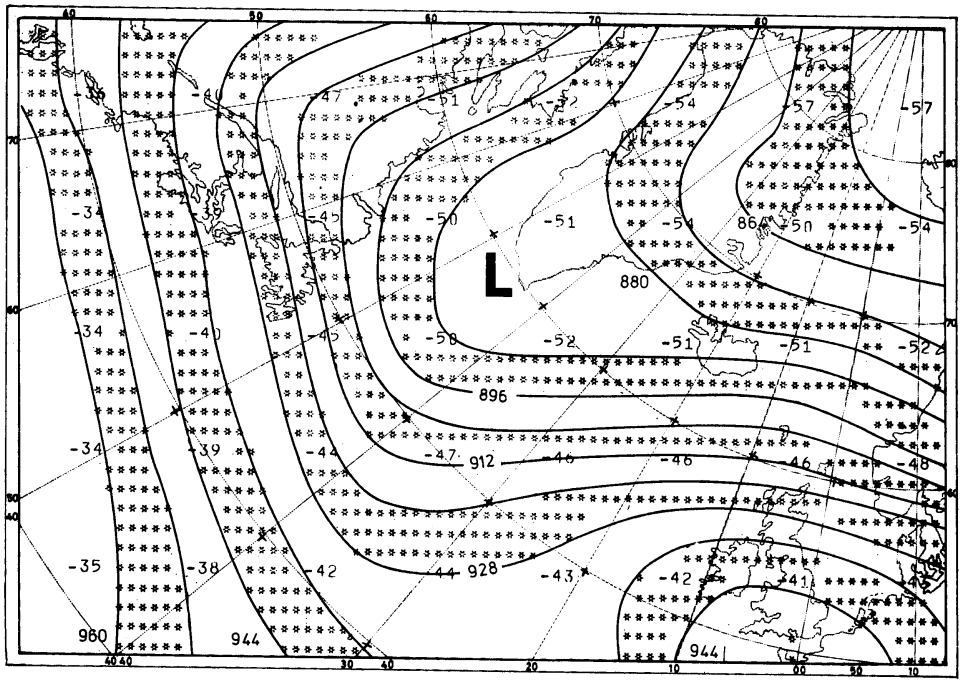

Fig. 24 Line print of an analysis of isobaric topography, 300 mbar for 00.00 GMT, 4 May 1970. Contours labelled in gpdam (interval every 8 gpdam). Grid point values of temperature in °C

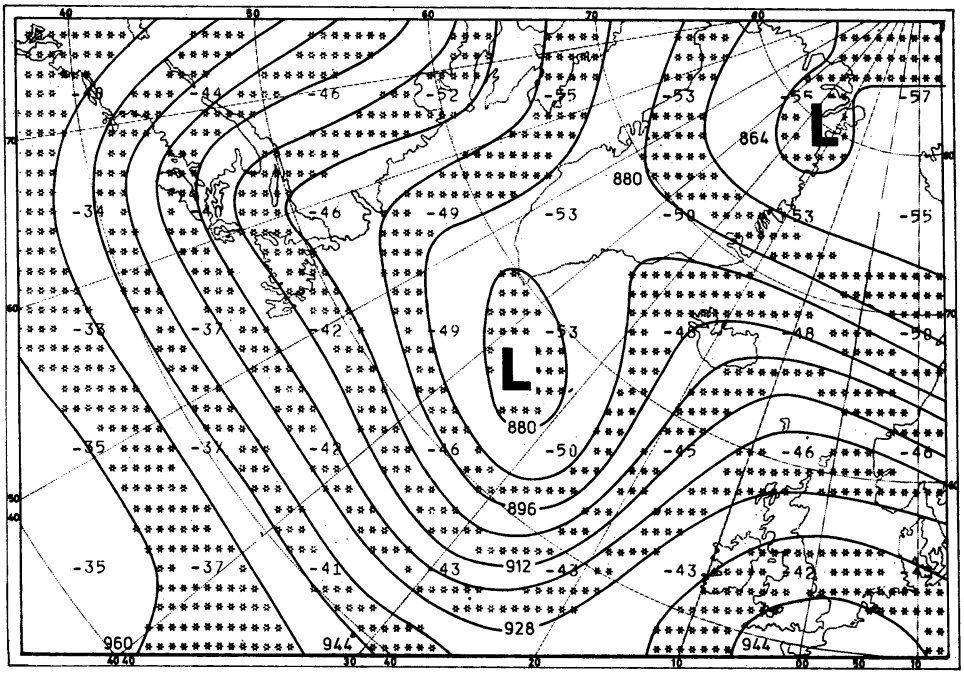

Fig. 25 Same as Fig. 24. 300 mbar, 12.00 GMT, 4 May 1970.

5.9 Numerical results

The purpose of this section is primarily to investigate the practical applicability of the various approaches by computer to the solution of the navigation problem in the time domain.

Efforts have been concentrated on the development of mechanized iterative procedures involving various versions of the control equations including the refraction formulae. In addition some programmes were written for application of the graph method.

The numerical experiments have been carried out for transatlantic crossings in order to bring them up to date with present day's operation.

The meteorological information consists of the grid point values for geopotential and temperature in the 300 mbar standard pressure level. The wind data are obtained by applying the geostrophic wind equation (5.23). The grid is a section out of the well-known northern hemispheric octagonal grid system used by the National Meteorological Center, Suitland, Md. This Cartesian grid is superposed on a conformal polar stereographic chart projection with true scale at 60^0 N. The spacing at 60^0 N is 381 km. The y-axis of the grid runs parallel with the 80^0 meridian.

In order to account for the dynamical character of the atmospheric circulation provision is made for the 300 mbar grid point representations for two validity times 12 hours apart.

Fig. 24 and 25 show the retouched line prints of the upper air analysis used in all experiments to be described.

All experiments are based on the constant Mach Number Technique for Mach True = 0.803. Acceleration and deceleration in the climb and descent phase of the flight are discarded whence it follows that the resulting track solutions should be considered as *overhead-overhead* Least Time Tracks.

The programs written for the mechanization of the iterative procedures use a synthesis of the computational schemes explained in the previous sections. They include the processing of the gradient equation, the phase velocity equation, Zermelo's steering equation and Von Mises refraction formulae.

The inclusion of the spherical domain in this system is obtained by inserting the corrective terms associated with a mapping of the sphere onto a polar stereographic chart, cf. (5.35) and (5.36).

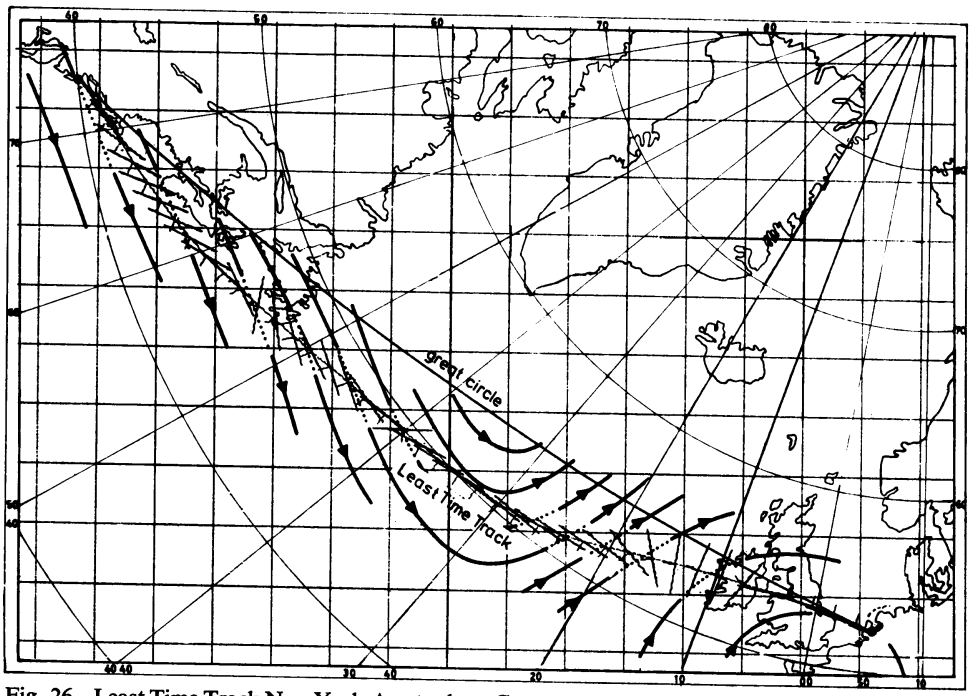

Fig. 26 Least Time Track New York-Amsterdam. Constant Mach (M = 0.803). Computer produced using Zermelo's steering equation. Time of departure 03.00 GMT. 'Streamlines' indicate airflow in phase with flight as depicted from information in Fig. 24 and 25. Minimum flight time: 6h 10′ 19″. Track distance: 3235 N.M. Also shown: refraction normal at every integration step (short bars). Longer bars represent the lines of 'discontinuity'.

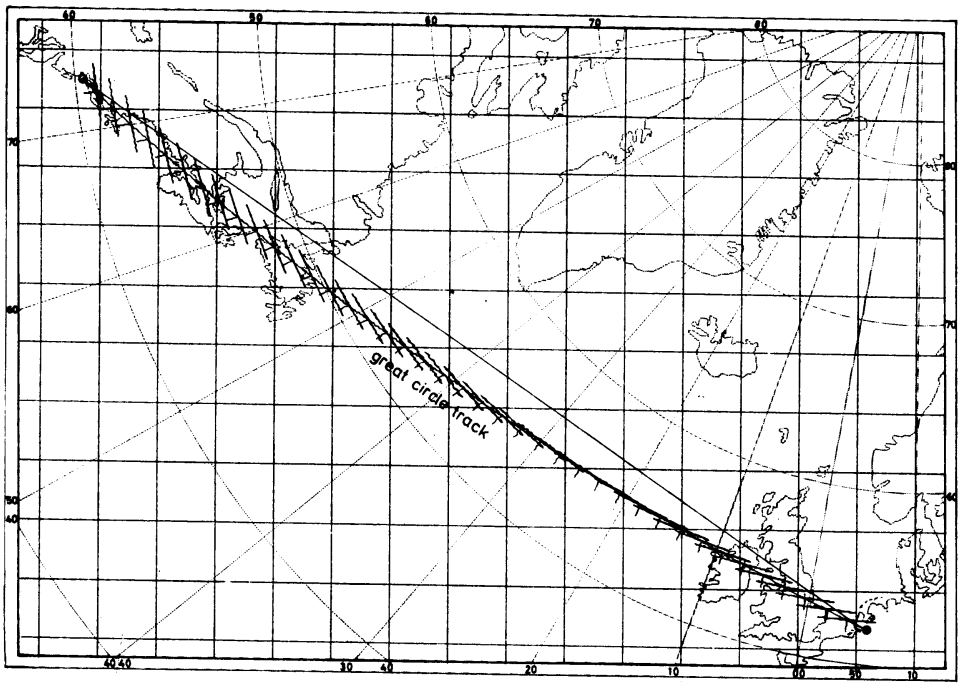

Fig. 27 Great Circle Track New York-Amsterdam. Computer produced using the gradient equation. Zero wind, constant temperature. Track distance: 3187 N.M. Short bars denote the refraction normal. Longer bars represent the lines of 'discontinuity'. Note that these bars are parallel to the local meridians and latitude circles.

Fig. 26 shows the result of an iterative search obtained for a flight between John F. Kennedy Airport, New York, to Schiphol Airport, Amsterdam, using Zermelo's steering equation. The curve plotter picture shows the track together with the lines of discontinuity and direction of the refraction normal at each integration step. The iteration took 6 cycles before the destination was reached within 5 n.m.

To check the correct performance of the programmes use can be made of an 'exact' solution as a test extremal. Fig. 27 has been generated by computer for the case of zero wind conditions to yield the great circle track between New York and Amsterdam, eastbound as well as westbound. The lines of discontinuity are parallel to the local latitude circles at each step and the refraction normal is directed along the local meridian. If the corrective terms for the scale S would have been abandoned from the control equations the programme would not have generated the great circle but the straight line connection on the projection chart instead.

In the iteration process difficulties may arise due to the ill-conditioned mapping between the initial and final states. The process may indeed be complicated by extremely high sensitivity to small changes in the initial heading. Even the most successful

procedure depends on first obtaining a trial solution whose terminal values lie somewhere in the vicinity of those specified. In the navigation problem the preliminary search for such a trial solution does not consume much time and effort as the initial heading is bound to lie within a comparatively small sector, but if 'tight' tolerances are employed for reaching destination the process is oscillatory and convergence may be slowed. If convergence fails to come within a prescribed time limit the tolerance may be weakened, but for practical reasons such 'loose' tolerances should remain within acceptable bounds, e.g. within 10 n.m. from the terminal for long range flights.

The overall success of the process will depend on the mapping between initial heading and final distance of approach at the terminal. This mapping may be very complicated. The mapping may be studied qualitatively by the investigation of a one-parameter family of tracks emanating from the departure point.

Fig. 28 shows such a one-parameter set of tracks originating from New York with an incremental change of the initial heading of 2 degrees. It is a curve plotter picture generated by a programme including the gradient equation as control equation.

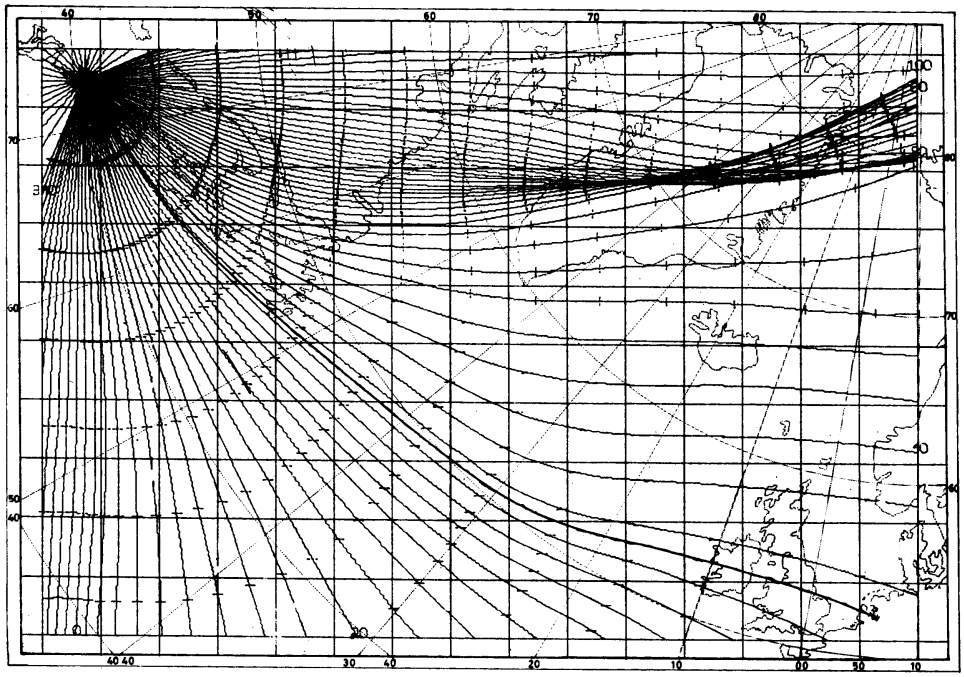

Fig. 28 One-parameter family of Least Time Tracks with New York as origin. Mach true = 0.803. Computer produced by means of an incremental plotter using the gradient equation. Meteorological information depicted from Fig. 24 and 25. Time of departure: 03.00 GMT. Short cross bars indicate the half-hourly transit times. Note the focussing effect near Greenland and Norway.

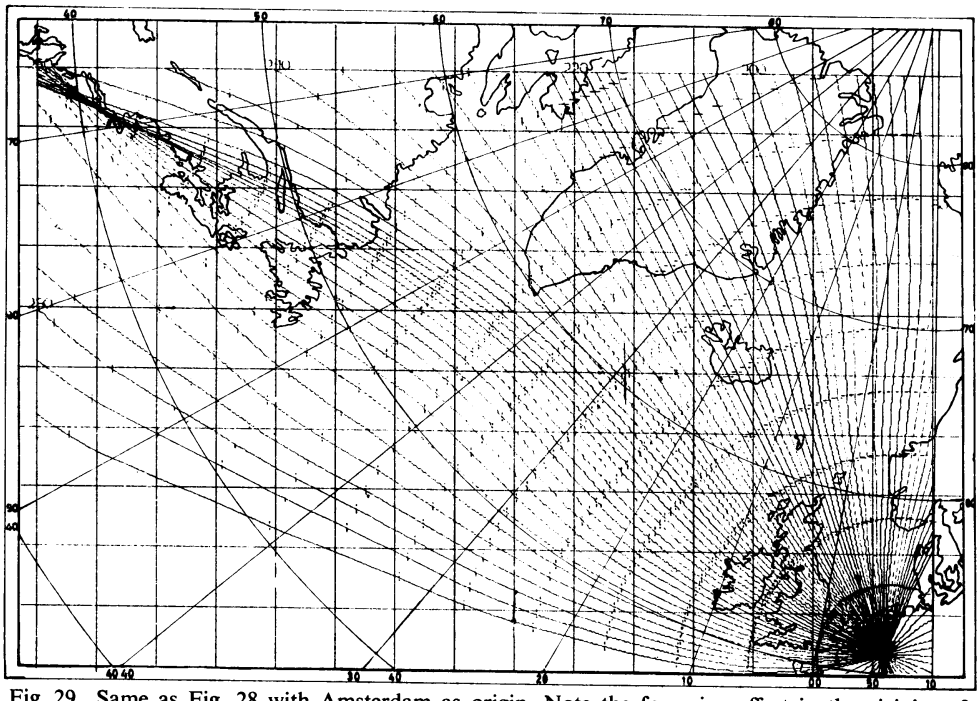

Fig. 29 Same as Fig. 28 with Amsterdam as origin. Note the focussing effect in the vicinity of New York.

In regular field areas the mapping referred to above may be well-conditioned especially where the crowding of tracks is pronounced. The iteration process will then be successful most of the time, using a simple convergence scheme like the regula falsi method of Newton, cf. (5.53). In regions where the crowding of tracks happens to be poor the sensitivity to small changes in the initial heading will be high, causing instability in the convergence process due to discretization, truncation- and round-off errors.

The regularity of the field of tracks is disturbed by a focussing effect over Greenland and Norway. In the region bounded by the enveloping caustica there occurs an overcrossing of tracks and time fronts. This feature destroys the uniqueness of the solution. For a terminal located in the region enclosed by the caustica two separate solutions exist, each giving a minimum with respect to adjacent trajectories.

The mapping function referred to above becomes complicated especially for terminals at or near a focus or caustic. For instance, if the focus is a terminal point, any trial solution is likely to reach the focus within the prescribed tolerance distance. For a terminal within the area enclosed by the caustica convergence may be fast, but

it is not known whether the resulting solution gives an absolute or a relative minimum.

Fig. 29 represents the one-parameter family of optimal tracks upstream with Amsterdam as point of departure. The wind and temperature patterns are the same as in Fig. 28. New York terminal is located within an area where the optimum tracks tend to focus.

Fig. 30 gives another example of a one-parameter family of extremals, which clearly demonstrates how complex the solution of the two point boundary problem may be when using iteration. The figure shows the more or less hypothetical case of a flight with Mach True = 0.5 within an idealized jet stream model. The wind profile is Gauszian with a core speed of 180 knots. The figure exemplifies the phenomenon of beating up against the wind, wellknown to yachtsmen.

If the aircraft is not allowed to make wide excursions from the jet axis the optimum tracks become sinusoidal.

The computational experience with this type of two point boundary problems is very limited at the time of writing. Its elaboration awaits further studies.

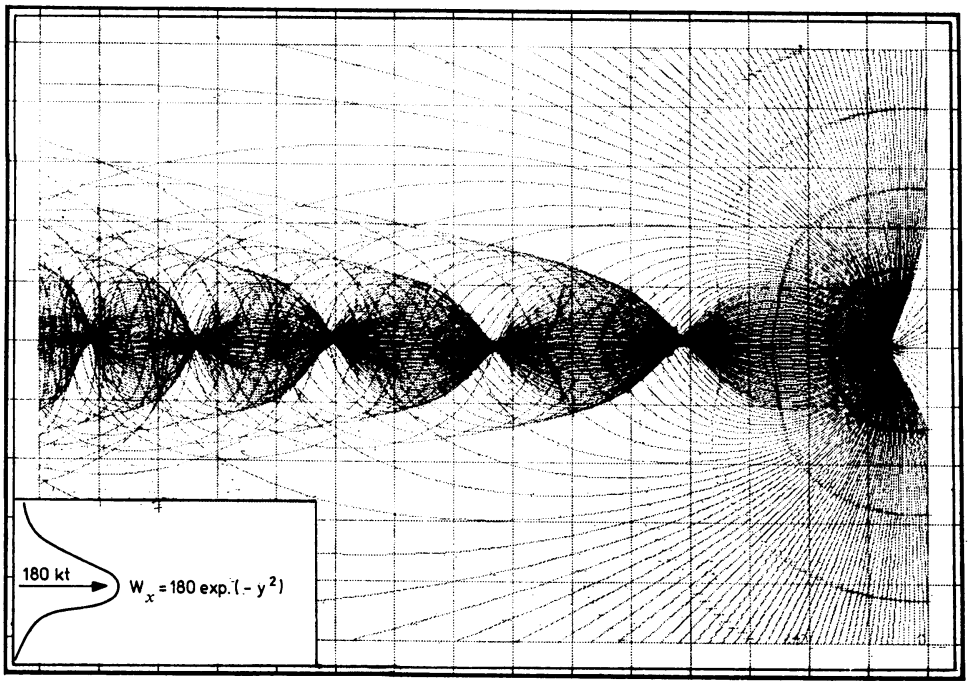

Fig. 30 One-parameter family of Least Time Tracks for Mach true = 0.5 with origin at the axis of a hypothetical jet stream. The wind profile is Gauszian with a core speed of 180 kts. Computer produced using the phase speed equation. Note the oscillatory character of the solutions in the vicinity of the axis.

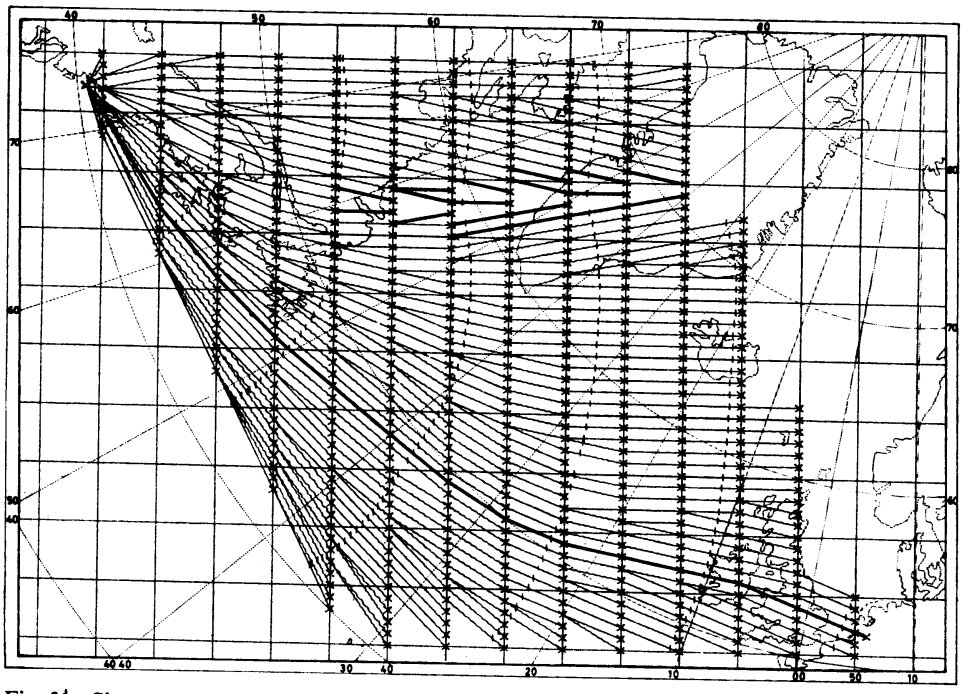

Fig. 31 Shortest spanning tree with New York as root of the tree. Computer produced using the graph method. Constant Mach (M = 0.803). Meteorological information depicted from Fig. 24 and 25. The 'shortest' path between New York and Amsterdam is accentuated by the heavier line. Time of departure: 03.00 GMT. Minimum flight time: 6 h 10′ 41″. Track distance: 3248 N.M.

When we direct our attention to the application of the graph theoretical approach for solving the two point boundary problem an entirely different picture evolves.

The graph procedure performs a track selection within a specified network of points and connecting arcs. It keeps away from delicate mathematical questions as continuity, convergence, focussing, etc.

The procedure affords a unique solution all the time.

For example, if the terminal lies within an area of overcrossing optimal tracks the graph method yields always a solution which warrants an absolute minimum with respect to admissible paths in the graph.

Fig. 31 shows the shortest spanning tree in the graph with New York as the root of the tree. This shortest spanning tree corresponds to the one-parameter family of optimum tracks as shown in Fig. 28, which was produced by the iterative programme. The graph points have been chosen to lie on the y-grid lines of the meteorological grid. Note the peculiar structure of the graph in the region which reveals the focussing effect. The particular solution between New York and Amsterdam within this graph

may be back-tracked starting at the terminal (thick line). This track corresponds to that in Fig. 26, obtained by Zermelo's method. In another version involving the graph method the graph points are located at the circumference of a set of concentric circles about the origin. The shortest spanning tree associated with this graph structure is shown in Fig. 32. The graph solution between New York and Amsterdam is again a good approximation of the 'exact' solution.

A disadvantage of this graph method is its moderate speed. The running time increases proportionally with the number of subsets of nodes and quadratically with the number of nodes in the sub sets. The processing requires the passing on of the whole spanning tree.

The graph algorithm lends itself extremely well for comparative studies, for instance to determine the potential gain in flight time with respect to another track for comparison.

Example: Let it be intended to compare the flight time and distance along the optimal track with the corresponding elements along the geometric shortest path in the graph. This geometric shortest path which is an approximation of the great circle arc between the terminals may first be determined by activating the program under zero atmospheric conditions.

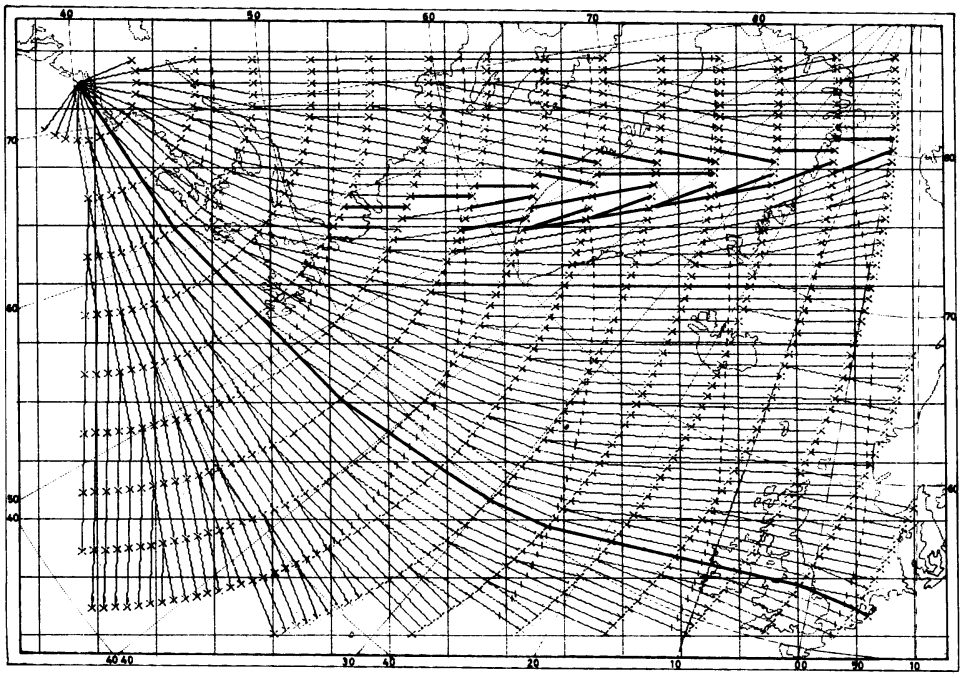

Fig. 32 For explanation see legend to Fig. 31. Minimum flight time: 6h 11′ 01″. Total distance: 3251 N.M.

Given the indices of the nodes of this path the program may be restarted to produce the desired elements in real atmospheric conditions. This may be accomplished by putting the lower and upper bounds in the j-cycle of the scheme equal to the indices corresponding to the nodes of the geometric shortest path.

An important question is to draw a comparison between the various computational techniques, for instance in regard to speed and ease of computation. Computational experience with the various iteration adjustment processes reveals that the schemes based on the gradient equation, phase velocity- and the steering equation of Zermelo are practically alike. There are no indications whatsoever which of these schemes should be preferred. They produce curves barely distinguishable from each other. The iterative procedure involving Von Mises' refraction formula is more complex in view of the selection of the correct refraction angle and the removal of complex roots in the governing quadratic goniometric equation. The tric to rotate the refraction normal in order to remove the complex roots distorts the solution considerably and the speed of convergence may be low.

In so far as comparison between the various iterative schemes and the graph method is concerned, the active competition appears to be won by the graph method. The former require a complicated correction cycle to obtain a solution for the two point boundary problem, which is not necessary with the latter. The graph method will always afford a unique solution, which guarantees an absolute minimum.

Although the speed of computation required for the graph algorithm is unfavourable in regard to the fast convergent iteration processes, this is not considered an unsurmountable obstacle with modern computer facilities.

PART II

ASPECTS OF 3-DIMENSIONAL FLIGHT PLANNING

6. 3-Space graph solution of the optimal flight path

6.1 General

Flight planning and the assignment of tracks to individual aircraft are made in accordance with the navigational capability of the aircraft concerned. The navigational capability varies with the cruise-performance characteristics and the time changing conditions of the atmospheric environment. The influence of the environmental conditions on the performance is exploited by the companies to obtain the assignment of the most advantageous flight path, advantageous not always in the sense of the speediest fligth as described in the previous sections. The criterium of optimality may be entirely different, for the ultimate goal is to produce the optimal plan for each flight which will fulfil the companies' policy of carrying the offered payload in the most economical manner under the safest possible operating conditions.

One of the most pronounced features of the optimization problem in flight planning is its 3-dimensional character, as may be inferred from possible diversions of the ground track in the horizontal and the adherence to a continuous climb or stepped altitude profile in the vertical (Fig. 33).

Operationally the practices and procedures have evolved in a direction whereby the optimization is carried out in a two-stage process. First an optimal track selection is performed in terms of a time-optimal criterium (Least Time Track). Secondly an optimization procedure is applied in the vertical in terms of the same or another

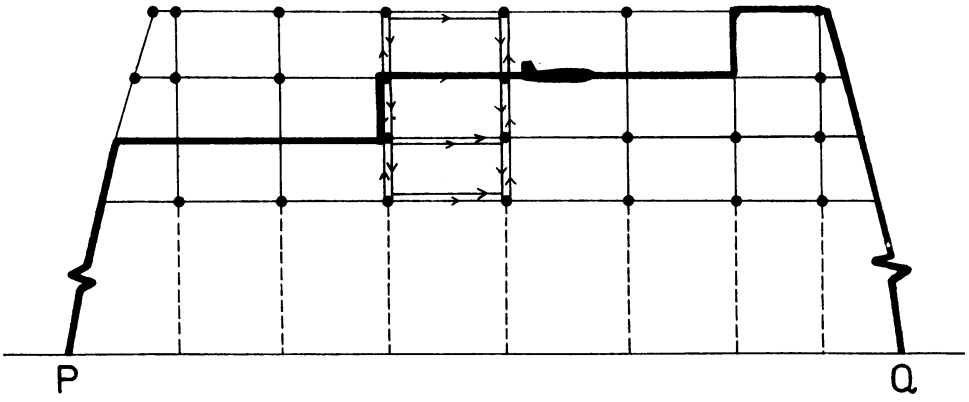

Fig. 33 (Optimal) stepped climb profile within a 2-space graph consisting of regulated flight levels and intermediate climbs (descents) at check points.

criterium to determine the optimal altitude step schedule.

Note:
In some Air Navigation Regions like the NAT Region the first stage is the responsibility of the ATC centres.

In current practice the 3-dimensional problem has in this way been split up in two separate 2-dimensional problems. From a point of view of logistics this approach seems to be somewhat fortuitous.

Since we may dispose now of powerful computer processing systems it would appear worthwhile to attack the problem directly in 3-space. For it may be expected that a 3-space solution by the effects of horizontal and vertical temperature gradients and wind shear, may be 'better' than a two-stage 2-space solution. Especially in the flight régime of turbo-jet aircraft the prevailing meteorological conditions may be so pronounced (jet stream, tropopause) that the 3-space solution may indeed be better than a combination of the Least Time Track and optimal step climb profile in the vertical (See section 6.4).

In addition, if we succeed in finding an optimization algorithm for the 3-space problem it is a simple matter to use this same algorithm as a tool for solving the above-mentioned 2-stage problem. This simplifies also the design of a well-ordered process architecture and program structure, which makes computerized flight planning more comprehensive and flexible.

In this extent it is necessary to point to the fact that the flexibility is manifest also in the degrees of freedom of the system. The operator has the choice of various cruise systems: Long Range, Standard Cruise (Constant Mach), High Speed, etc. The optimization criterium may be specified in terms of minimum operating time, minimum fuel flow and minimum operating costs. The best strategy is trying to find a process algorithm which solves the optimization problem in the most general terms possible. Subsequently, when arrangements are made to develop an *operational* process due account is given to the specific navigational and performance aspects encountered in real practice.

The purpose of this chapter is to present a computational algorithm, which is suited to be used in more advanced electronic flight planning processes. The main intention is to prove its usefulness by experimentation.

6.2 Optimal track selection in 3 dimensions

6.2.1 The graph method

The optimization problem in '3-dimensional' flight planning could in principle be studied in the context of modern optimal control theory. This approach, however,

has not been explored in any depth. Its elaboration is hampered by the complexity of the flight performance in the vertical. In the régime of varying atmospheric conditions of winds and temperature the problem becomes even a space-time problem.

Zermelo [1] using the classical tools of variational calculus succeeded in formulating a solution in 3-space, but under now unacceptable assumptions of climb- and descent parameters. The criterium used for optimization was minimum operating time. The solution was presented in the form of a set of steering equations similar to that found in the 2-space navigation problem.

A simpler and more direct approach is to investigate the flight performance within the structure of a graph in 3-space. In section 3 the graph method was explained for the navigation problem in 2-space. The 2-space we had in mind there was a constant pressure level, be it a flight level or a meteorological standard pressure level. Another graph in 2-space is encountered when the performance is studied in a vertical plane through a prescribed flight track. The graph is specified here by the set of flight levels and step-climbs or step-descents (Fig. 33). The location of the steps is supposed to be fixed at check points. The nodes of the graph consist of the intersection points of the vertical lines through check points and the set of flight levels. The arcs of the graph consist of the flight segments *including steps*. The point of departure is the entry point. Destination is the exit point. The graph is a directed graph. It is also a multigraph in view of the multiplicity of arcs between a pair of nodes. A special feature in this graph is the arrangement of the nodes in ordered point sets by virtue of which the graph is divided into zones, a structure which is often encountered in transport theory. The performance optimization in this 2-space graph may be elaborated using the same principles as explained for the navigation graph in section 3, resulting in the identification of the optimum altitude step profile.

In the two-stage processing of modern computer flight planning it has now become general practice to perform the optimal track selection in the 2-space graph in the horizontal and to identify the optimal altitude step schedule in the 2-space graph in the vertical.

Now we shall focus our attention on the possibility of solving the optimal flight planning problem directly in the navigation régime of a graph in 3-space. This navigation régime may be considered to have the appearance of a crystal lattice (Fig. 34). The check points constitute the lattice points. In continental flight regions these check points are located at NAVAID points, aerodromes, transition points, etc. In oceanic regions these check points consist of the one degree crossing points at 5 or 10 degrees standard meridians. The flight levels specify the lattice planes. The purpose of the optimization problem is to search for the 'shortest' path in the 'navigation crystal' along which a specified performance index attains an (absolute) minimum or maximum value. This performance index or 'figure of merit' is associated with an entity like minimum operating time, minimum fuel consumption, minimum operating costs,

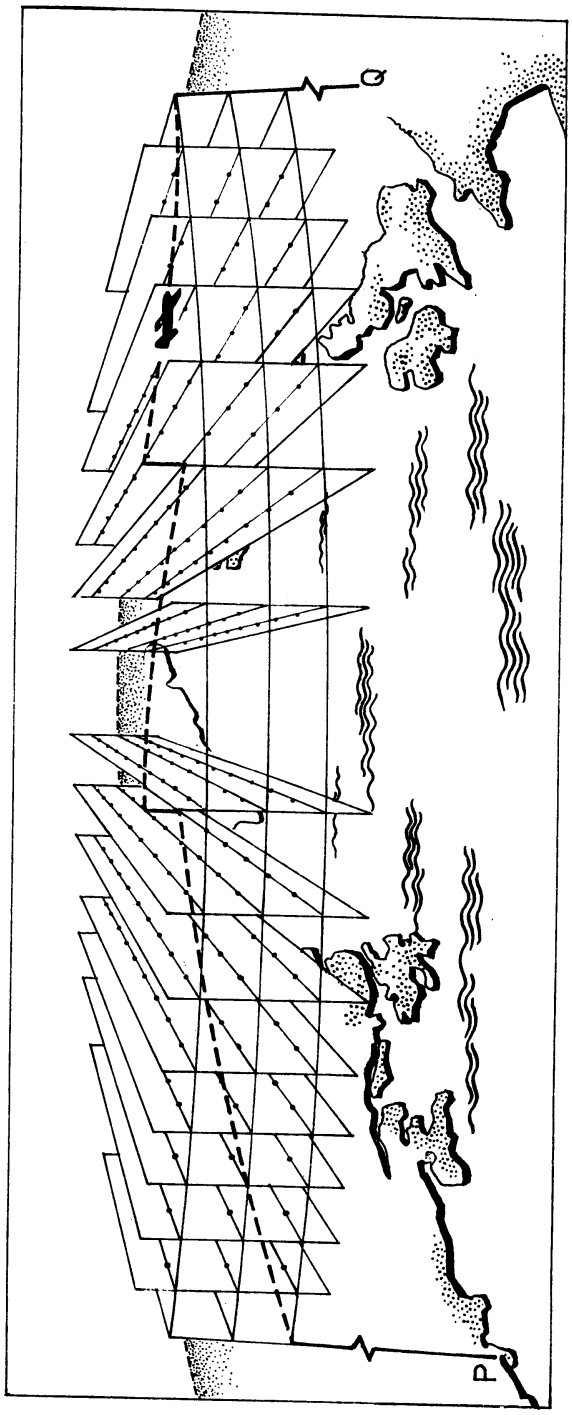

Fig. 34 Special graph — 'navigation crystal' — to identify the 'shortest path' in 3-space flight planning.

etc. in a preselected cruising system. The shortest path problem may then first be formulated in terms of the performance index without reference to the specific desiderata of flight performance. These last features come into play only when a definite flight planning process is developed in accordance with the company's standard practices.

Note:
The structure of the graph as depicted in Fig. 34 presupposes that the graph is partitioned in zones of air space. This requires the arrangement of nodes in point sets similar to the ordered sets of oceanic check points along the standard meridians. This grouping may always be realized by adding, if necessary, some fictitious check points. Without such grouping in the graph configuration it would become extremely complicated to find a shortest path solution.

6.2.2 Optimization scheme

Let the nodes in the 3-space graph be arranged in the point sets X_i, $i = 0, \ldots, n + 1$ with the following properties. X_0 is the entry point P, X_{n+1} the exit point Q. A node in the set X_i is specified by its 'coordinates' (i, k, l), a node in the set X_{i+1} by the coordinates $(i + 1, K, L)$ (Fig. 35).

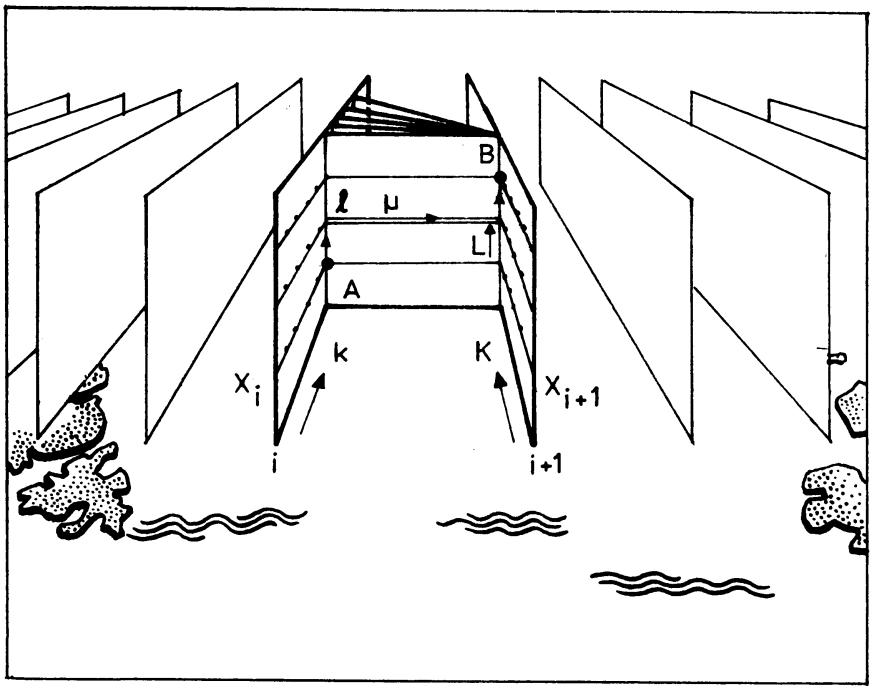

Fig. 35 Portion of the 3-space graph indicating the parameters used for developing an optimization scheme.

Let us assume that the optimum value of the performance index J is known for all nodes in set X_i. Further, let $A(i, k, l)$ be an arbitrary node in X_i with optimum value $J^{opt}_{i,k,l}$. Next consider a node $B(i+1, K, L)$ in X_{i+1}. The vertical plane through A and B is the support of a subgraph. This subgraph is a multigraph, as the nodes A and B may be connected by multiple altitude segments, steps-up and steps-down included.

The value of J in B along an individual altitude segment μ becomes:

$$J_{i,k,l,K,L,\mu} = J^{opt}_{i,k,l} + (\Delta J)_{i,k,l,K,L,\mu} \tag{6.1}$$

Here $(\Delta J)\ldots$ denotes the incremental change of J along the stepped altitude segment μ. Then, to obtain an optimal value of the performance index $J^{opt}_{i+1,K,L}$ we first search for an optimum in the subgraph:

$$J^{opt}_{i+1,K,L,k} = \underset{(l)}{opt}\{\underset{(\mu)}{opt}(J^{opt}_{i,k,l} + (\Delta J)_{i,k,l,K,L,\mu})\}.$$

This is a sub-optimum value only. In order to acquire the real optimum it is required to rotate the planar support about node B and to repeat the optimization process in terms of k:

$$J^{opt}_{i+1,K,L} = \underset{(k)}{opt}\{J^{opt}_{i+1,K,L,k}\}$$

$$= \underset{(k)}{opt}\{\underset{(l)}{opt}\{\underset{(\mu)}{opt}(J^{opt}_{i,k,l} + (\Delta J)_{i,k,l,K,L,\mu})\}\}.$$

$$\mu = \mu_1(i), \ldots, \mu_2(i)$$

$$l = l_1(i), \ldots, l_2(i)$$

$$k = k_1(i), \ldots, k_2(i)$$

The set of critical index values k^{crit}, l^{crit}, μ^{crit} for which $J^{opt}(B)$ is obtained defines a node in X_i and arc, which appertain to the optimum path. It is required to keep a record of these critical indices during execution of the process algorithm. In the zonal airspace between X_i and X_{i+1} the above described processing should be repeated for all nodes B in X_{i+1}. Thereafter the process is continued in consecutive zones. It terminates after the exit point Q has been reached.

In its totality the process may be formulated as follows:

$$\left[\left[J^{opt}_{i+1,K,L} = \underset{(k)}{opt}\{\underset{(l)}{opt}\{\underset{(\mu)}{opt}(J^{opt}_{i,k,l} + (\Delta J)_{i,k,l,K,L,\mu})\}\}\right]\begin{matrix}K = K_1(i+1)\ldots K_2(i+1)\\ L = L_1(i+1)\ldots L_2(i+1)\end{matrix}\right] i = 0\ldots n \tag{6.2}$$

$$\mu = \mu_1(i), \ldots, \mu_2(i)$$
$$l = l_1(i), \ldots, l_2(i)$$
$$k = k_1(i), \ldots, k_2(i)$$

together with the set of critical indices:

$$[k^{\text{crit}}(i,K,L),\ l^{\text{crit}}(i,K,L),\ \mu^{\text{crit}}(i,K,L)\ |\ J^{\text{opt}}_{i+1,K,L}]\ i = 0 \ldots n. \qquad (6.3)$$

The ultimate result of a process run (6.2) is the (accumulated) value of the optimized performance between take-off and landing.

The corresponding path information should be derived from the set of critical indices (6.3). The nodes and arcs which specify the geometry of the optimal path are identified by means of a search of the relevant indices in (6.3) starting in the exit point and 'backtracking' until the point of departure has been reached.

To appreciate the algorithm as a very useful tool in computer flight planning it is appropriate to consider its properties and possible applications in practice.

6.2.3 Properties

a. In practice the optimality criterium is usually a minimum criterium: minimum fuel flow, minimum operating costs, minimum operating time. In case of fuel expenditure, however, it is customary to express this quantity in terms of loss of aircraft weight. Taking the aircraft weight as a performance index the intention is to search for a maximum weight.

The principle of dynamic programming in the spatial graph preserves its validity if the performance index also depends on absolute time. Especially under control of rapidly changing wind and temperature patterns due account should be given to these factors in the evaluation of J and ΔJ.

The principle equally holds when ΔJ depends on J itself. For instance if the optimality criterium is based on fuel economy the fuel flow will depend on the fuel consumed already since the moment of take-off. Operating costs are predominantly expressed as a (linear) combination of fuel flow and operating time. In a computer flight planning system this is a strong argument to program the scheme on the basis of a cost index. The system then provides great flexibility to explore not only cost savings but also time- or fuel savings merely by letting vanish the values of some coefficients.

b. To start the process it is required to know the initial value $J_{0,0,0} = J(P)$ in the entry- or departure point. In practice, however, it may occur that this initial value is not known in advance to the operator. For example, in case the performance is

considered in terms of fuel consumption, J is taken as aircraft weight. It may then happen that the take-off weight $J(P)$ is not specified, but instead the (maximum) landing weight $J(Q)$ in the exit point.

It is a favourable circumstance that even then the algorithm is still applicable, provided it is performed in retrograde order starting at the exit point and terminating in the entry point.

The retrograde scheme may be formulated as follows:

$$\left[\left[J^{\text{opt}}_{i,k,l} = \underset{(K)}{opt}\{\underset{(L)}{opt}\{\underset{(\mu)}{opt}(J^{\text{opt}}_{i+1,K,L} - (\Delta J)_{i,k,l,K,L,\mu}\}\}\right]\begin{matrix}k = k_1(i)\ldots k_2(i)\\ l = l_1(i)\ldots l_2(i)\end{matrix}\right] i = n\ldots 0$$

$$\begin{aligned}&\mu = \mu_1(i),\ldots,\mu_2(i)\\ &K = K_1(i+1),\ldots,K_2(i+1)\\ &L = L_1(i+1),\ldots,L_2(i+1)\end{aligned} \qquad (6.4)$$

with the set of critical indices:

$$[K^{\text{crit}}(i+1,k,l),\ L^{\text{crit}}(i+1,k,l),\ \mu^{\text{crit}}(i+1,k,l)\,|\,J^{\text{opt}}_{i,k,l}]\ i = n\ldots 0. \qquad (6.5)$$

After the process (6.4) has been finished the track information is drawn from the set (6.5) by screening the relevant critical index values starting at the entry point P and proceeding to the exit point Q.

c. The principle of dynamic programming is not affected by reducing the airspace even when the bounds of one or more index intervals are made to coincide. As a consequence the algorithm may still be used if the graph in 3-space is made to degenerate into a 2-space graph or even a 1-space graph c.q. a prescribed track and altitude profile. Identification of the lower and upper bounds of $l(i)$, $L(i)$ and $\mu(i)$ intervals affords the optimization scheme for the optimal track solution in the 2-space graph of a predetermined flight level. Identification of the bounds in the $k(i)$ and $K(i)$ intervals affords the scheme for obtaining the optimum altitude step schedule in a vertical plane. If carried out one after the other this processing simulates the two-stage process as described in section 6.1.

By a proper identification of the interval bounds for the indices $k(i)$, $K(i)$, $l(i)$, $L(i)$ and $\mu(i)$ the graph is made to degenerate into a specified flight path. Execution of the algorithm (6.2) then merely results in a performance computation along this flight path. The latter specialization is extremely suited for comparison of the performance data along the optimal path and concurrent tracks like the Great Circle Track, an alternate track in the organized track system, etc.

Note:

In a computerized process of optimal track selection it is for several reasons to be recommended to repeat the calculation in lower dimensions as soon as the first 3-space run terminates. One reason is that in the process one should keep a record of the critical index values k^{crit}, l^{crit} and μ^{crit}, which specify the geometry of all optimal paths emerging from the entry- or exit point. To save core storage it may be recommended to store only the index values k^{crit}. These values fix the location of the (projected) ground track associated with the optimum path. Then the subroutine may be put into action a second time applied in the 2-space graph in the vertical plane through this ground track. This will result in the identification of the optimum altitude profile. This time the l and μ indices are stored. Finally, for the purpose of the computation and presentation of all desired flight plan entries the subroutine may be put into action once more, applied in the 1-space graph specified by the k, l and μ-indices.

Another advantage of this stepwise processing procedure is the possibility of restarting the subroutine with improved values for the take-off weight or landing weight resulting in an improved solution of the two-point boundary problem. (See 6.3.5.)

d. For the purpose of long distance flight planning the spatially graph may be established in the flight region in its most comprehensive form on the basis of air traffic patterns and regulated flight levels. In practice the operations are often confronted with all kinds of airspace limitations. Most common are restrictions imposed by ATC (restricted climbs, airways and flight level restrictions) and military blockages. In the spatially graph certain nodes and arcs need then to be blocked, so that the optimization problem can be solved in a subgraph. Blocking may also appear a necessity during the execution of the algorithm due to cruise thrust limitations. At each node of the optimal path the aircraft weight should not exceed a maximum operating weight, which is known beforehand for each type of aircraft depending on aircraft weight and the temperature deviation from standard (See 6.3.5).

This in turn has the consequence that application of the scheme requires a separate calculation of the weight performance even when the optimality criterium is in terms of cost saving or time saving. It may be noticed in passing that these weight limitations may cause that an actual computer run fails to give a solution, for instance when the available flight levels are too high for a heavy loaded aircraft.

The removal of graph elements may be so drastically that the remaining subgraph admits but a few concurrent paths.

Example: Current practices and procedures relating to the airspace organization in the NAT Region involve the establishment and application by ATC of an organized track system in parts of this Region.

The organized track system itself is the product of an optimal track selection established and updated at regular intervals using forecast MET conditions.

The tracks are designated in order of preference A B C ... routes. After authorization to use this route structure or part of it the operators can take this structure into account in their flight planning. The organized track system together with the continental track system determine a (spatially) subgraph, which may be used by the airline to perform a supplementary and final track selection. The ordering of routes, route A as the 'best' route, route B as the 'next best' route and so on, obtained in this final processing may be entirely different from that established by ATC.

The blocking of nodes and arcs may easily be accomplished in the optimization algorithm without distortion of the scheme.
Problems may arise, however, which in order to solve them require a further stylizing of the scheme. To take an example, we once again refer to the establishment of an organized track system. The algorithm may first be put into action to obtain the best route. Once this route has been identified, the process may be generated once again to determine the next best route. The best route should first be removed from the graph, but in order to fulfil the (lateral) separation standard the next best route should approach the best route by a distance not closer than 2 degrees north or south of this route. A detailed analysis of this problem reveals that it may be solved by additional blocking of nodes and arcs and an appropriate partitioning of the index intervals in subintervals.
A last but most important restriction is related to the weight limitation at the point of departure and destination. Both landing weight and take-off weight are bound to remain below prefixed weight limits.

6.2.4 *Reduced optimization scheme*

According to the formulations (6.2) and (6.4) the optimization scheme involves 6 cycles and a supplementary one to assess the desired flight path information.

The execution time is proportional to $N_i N_k^2 N_l^3$, where N_i, N_k and N_l are representative values of the number of stages, the number of check points per stage and the number of regulated flight levels. The number of arcs to be scanned in a spatially graph for the North Atlantic may run in the millions. This is quite an obstacle for the implementation of the method, even when use is made of a powerful computer. It is therefore from a point of view of computation speed of paramount importance to investigate whether the scheme may be relaxed in terms of the cycling process without introducing gross errors in the overall performance.

Consider in (6.1) the argument:

$$J^{\text{opt}}_{i,k,l} + (\Delta J)_{i,k,l,K,L,\mu}$$

The incremental change in performance ΔJ along an arc between a pair of nodes A and B (Fig. 35) consists of a contribution along the altitude segment μ plus a contribution due to steps-up and steps-down required to reach A and B. The steps are multiples of unit-steps. For example, if regulated altitudes are used with a separation of 2000 feet the unit-step may be taken as 2000 feet. Now let us assume that the contribution to ΔJ per unit-step is a fixed value ε and is equal to the contribution to ΔJ per unit step-down with opposite sign, then between A and B we have:

$$\begin{aligned}(\Delta J)_{i,k,l,K,L,\mu} &= (\Delta J)_{i,k,K,\mu} + (\mu - l)\,\varepsilon - (\mu - L)\,\varepsilon \\ &= (\Delta J)_{i,k,K,\mu} + (L - l)\,\varepsilon. \end{aligned} \tag{6.6}$$

For turbo-jet aircraft ε is of the order of 25 kg/1000 ft.

Referring to the formulation (6.2) we have

$$J^{\text{opt}}_{i+1,K,L} = \underset{(k)}{opt}\,\{\underset{(l)}{opt}\,\{J^{\text{opt}}_{i,k,l} + \underset{(\mu)}{opt}\,(\Delta J)_{i,k,K,l,L,\mu}\}\}. \tag{6.7}$$

Next, let us assume that

$$J^{\text{opt}}_{i,k,l} = J^{\text{opt}}_{i,k} + l\varepsilon, \tag{6.8}$$

then we can prove that this statement holds at each step of the optimization procedure.

Indeed, entering (6.6) and (6.8) into (6.7) delivers:

$$\begin{aligned} J^{\text{opt}}_{i+1,K,L} &= \underset{(k)}{opt}\,\{\underset{(l)}{opt}\,\{J^{\text{opt}}_{i,k} + l\varepsilon + \underset{(\mu)}{opt}\,(\Delta J)_{i,k,K,\mu} + (L-l)\varepsilon\}\} \\ &= \underset{(k)}{opt}\,\{J^{\text{opt}}_{i,k} + \underset{(\mu)}{opt}\,(\Delta J)_{i,k,K,\mu}\} + L\varepsilon \end{aligned}$$

or

$$J^{\text{opt}}_{i+1,K,L} = J^{\text{opt}}_{i+1,K} + L\varepsilon. \tag{6.9}$$

If it can be shown that the statement is also valid in the climb and descent zone, the statement holds for the complete 3-space graph.

Climb zone

In the climb zone the subgraph has a simpler structure, so that (6.6) should be

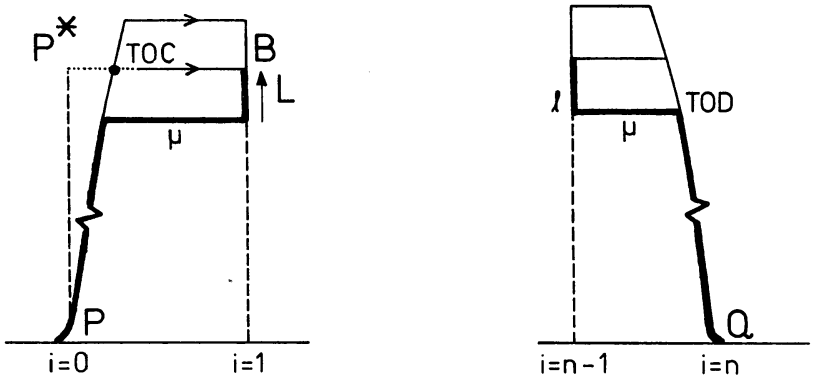

Fig. 36 *a*, *b* Climb and descent zone.

considered for an arc between the entry point and a point B over the first check point (Fig. 36a).

$$(\Delta J)_{0,0,0,K,L,\mu} = (\Delta J)_{0,0,K,\mu} + (L-\mu)\,\varepsilon. \tag{6.10}$$

Entering (6.10) into (6.7) affords:

$$J^{\text{opt}}_{1,K,L} = J(P) + \underset{(\mu)}{opt}\,\{(\Delta J)_{0,0,K,\mu} - \mu\varepsilon\} + L\varepsilon.$$

$$J^{\text{opt}}_{1,K,L} = J^{\text{opt}}_{1,K} + L\varepsilon. \tag{6.11}$$

Here $J(P)$ denotes the initial value of the performance index at the entry point. The increment (ΔJ) involves the climb performance and the performance along the portion of the altitude segment μ remaining between *TOC* (top of climb) and the first check point.

Descent zone

The subgraph has a structure similar to that in the climb zone. Referring to (6.6) and Fig. 36b we obtain:

$$(\Delta J)_{n,k,l,0,0,\mu} = (\Delta J)_{n,k,0,\mu} + (\mu - l)\,\varepsilon. \tag{6.12}$$

And according to (6.9):

$$J^{\text{opt}}_{n,k,l} = J^{\text{opt}}_{n,k} + l\varepsilon. \tag{6.13}$$

Entering (6.12) and (6.13) into (6.7) gives

$$J^{\text{opt}}_{n+1,0,0} = \underset{(k)}{opt}\,\{J^{\text{opt}}_{n,k} + \underset{(\mu)}{opt}\,\{(\Delta J)_{n,k,0,\mu} + \mu\varepsilon\}\},$$

which formally may be written as:

$$J^{\text{opt}}_{n+1,0,0} = J_{n+1,0} + L\varepsilon \quad (L = 0). \tag{6.14}$$

resulting in the optimum value of the performance index in the exit point. Here ΔJ involves the performance along the portion of an altitude segment μ between the last check point and *TOD* (top of descent) augmented with the descent performance.

Using (6.9), (6.11) and (6.14) the scheme (6.2) then reduces to:

$$\left[\left[\left[J^{\text{opt}}_{i+1,K} = \underset{(k)}{opt}\{J^{\text{opt}}_{i,k} + \underset{(\mu)}{opt}(\Delta^* J)_{i,k,K,\mu}\}\right] + L\varepsilon\right] K = K_1(i+1)\ldots K_2(i+1)\right] i=0\ldots n$$

$$k = k_1(i),\ldots,k_2(i) \qquad (6.15)$$

$$\mu = \mu_1(i),\ldots,\mu_2(i)$$

where $(\Delta^* J)_{i,k,K,\mu}$ denotes:

in the climb zone: $(\Delta J)_{0,0,K,\mu} - \mu\varepsilon$,

in the descent zone: $(\Delta J)_{n,k,0,\mu} + \mu\varepsilon$,

elsewhere: $(\Delta J)_{i,k,K,\mu}$

Note that in the last stage, in the descent zone, $L = 0$.

To identify the critical path it is necessary to store the critical index values:

$$[k^{\text{crit}}(i,K),\ \mu^{\text{crit}}(i,K),\ |\ J^{\text{opt}}_{i+1,K} + L\ \varepsilon]\ i = 0\ldots n. \qquad (6.16)$$

In (6.16) l^{crit} is missing. This indicates that the 3-parameter family of optimized paths emerging from the entry point reduces to a 2-parameter family.

The algorithm (6.15) involves 4 cycles as distinct from the 6 cycles in the general approach. The computation time is now proportional to $N_i N_k^2 N_l$, which is a factor N_l^2 less than in the algorithm (6.2) – for 10 flight levels a factor 100. – . This is a reason why the application of (6.15) in practice appears to be most promising although it should be borne in mind that the solution is approximative only. The usefulness of the algorithm must be checked by experimentation.

A point which should not be overlooked here is the possibility that the consistency of extreme values for different options of the critical path might be disturbed. So it may happen that for example the fuel consumed along an optimal cost-saving profile is less than the 'minimal' fuel consumed along an optimal fuel profile (See 6.4, experiment 4).

The optimization is apparently very sensitive to a slight modification of the process algorithm. However, be it somewhat irritating, this is not considered a serious objection against the use of (6.15) and (6.16) as long as the solution remains close to the 'exact' solution.

The approximate scheme may further be specialized along the lines previously described for the general process algorithm.

Example: Let $k_1(i) = k_2(i)$, $K_1(i) = K_2(i)$ and $\varepsilon = 0$. Then the 3-space graph degenerates to a 2-space graph in a vertical plane through a prescribed ground track with negligible step performance. Then the scheme (6.15) reduces to

$$\left[J_{i+1}^{\text{opt}} = J_i^{\text{opt}} + \underset{(\mu)}{opt} (\Delta J)_{i,\mu} \right] i = 0 \ldots n, \quad \mu = \mu_1(i) \ldots \mu_2(i) \tag{6.17}$$

with track information

$$\left[\mu^{\text{crit}}(i) \,\middle|\, J_{i+1}^{\text{opt}} \right] i = 0 \ldots n. \tag{6.18}$$

The relation (6.17) is recursive, so that

$$J^{\text{opt}}(Q) = J_{n+1}^{\text{opt}} = J(P) + \underset{(\mu)}{opt}\, (\Delta J)_{0,\mu} + \ldots \underset{(\mu)}{opt}\, \Delta J_{n,\mu}. \quad \mu = \mu_1(0) \ldots \mu_2(0) \quad \mu = \mu_1(n) \ldots \mu_2(n) \tag{6.19}$$

The processing is very simple and may be used for finding an optimal altitude step schedule in terms of minimum operating time.

6.3 Implementation of the spatial graph algorithm

6.3.1 General

This section discusses the numerical aspects of an operational flight planning process which is founded on the previously described principles of a spatial graph algorithm. Some of the initial steps are high-lighted, necessary for the application of the mathematical technique as formulated by the optimization scheme (6.2) or the scheme in relaxed form (6.15). The feasibility of the spatial optimization process will then be illustrated by some examples in the next section.

The interdisciplinary nature of flight planning makes that the design of an overall system is not so simple as could be guessed from the optimal control formulation. Indeed, the synthesis of the spatial navigation problem involves the preprocessing of a voluminous amount of operational data: space-time meteorological information, geographic and geometric data and flight-performance data.

The preprocessing of this information has already been discussed in part in chapter 5 when dealing with the synthesis of the time optimal problem in two dimensions. The extension of the problem in three dimensions and the generalization of the optimal control in terms of cost- or fuel-saving demand however for the elaboration of some additional numerical details.

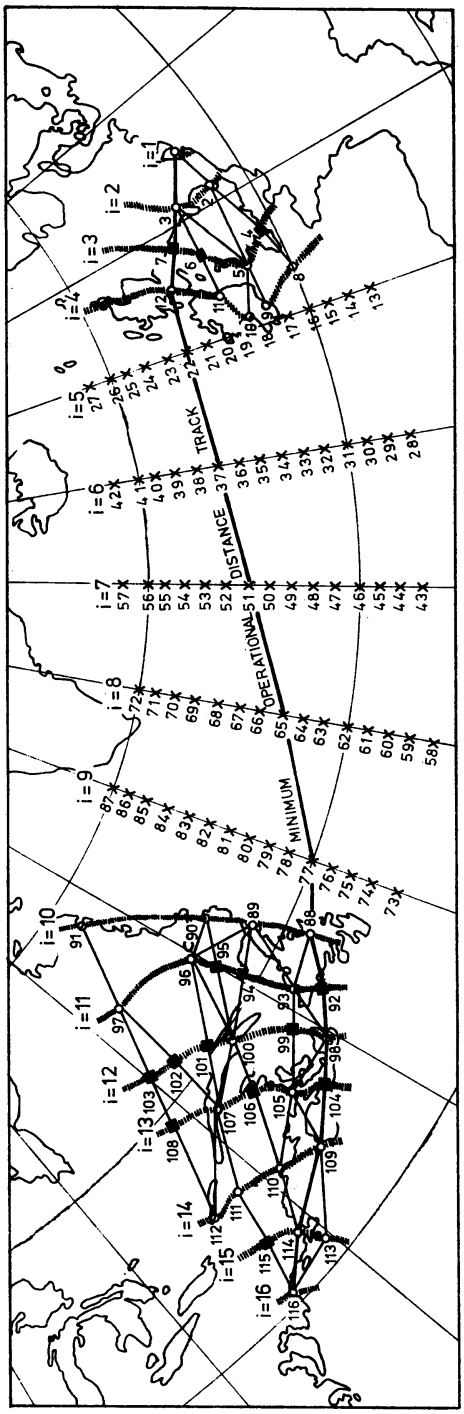

Fig. 37 Graph structure chosen in the NAT-region for experimentation. Minimum operating distance track.

6.3.2 *Geographic and geometric data*

We may perform the computations with respect to a fixed coordinate system $Oxyp$ (p for pressure in the vertical) defined in a regular geographic or Cartesian grid superposed on a conformal chart projection. The parameters which specify the grid geometry may best be depicted from a grid already existing and used in numerical weather prediction models. The necessary coordinate transformations, for instance to convert the geographical into Cartesian and vice versa, may be inferred from expressions like (5.17) and (5.18). Involved with the synthesis of the navigation problem in a (spatial) graph we need as a next step to define the check points, which constitute the nodes of the graph. These nodal points are arranged in ordered groups. The geographical position coordinates are converted into xy-coordinates and the points are labelled with the indices (i, k, l) as required to perform the computational algorithm (6.2). To obtain a zonal graph structure it may be necessary to add some fictitious graph points different from NAVAID points, reporting points etc.

The arrangement of check points used in the experiments for the NAT-region is shown in Fig. 37.

To complete the graph structure a set of flight levels is specified and provided with the index μ. The lowest level is FL 290 as turbo-jet aircraft operate in the airspace at or above this level. In view of the presently applied vertical separation standards the levels are 2000 ft apart.

Note:
In regions where *composite separation* is applied half the value of the vertical separation is introduced (1000 ft).

The algorithm is put into action for a sub-set of these levels only, for in practical operations the number of flight levels available is usually restricted.

As regards the flight segments between pairs of nodes some geometric data are essential for the computation: distance, location and orientation (bearing). The conformity of the projection makes that these data may easily be determined in reference to the grid coordinates. The great circle distance of the segment may be derived by using one of the formulas (5.19) or (5.20).

6.3.3 *Meteorological grid point values*

Computerized flight planning requires the provision of a voluminous amount of meteorological data. In case of spatial flight planning for long distance flights there is a need to receive forecasts for at least three standard levels, say 300, 250 and 200 mbar together with tropopause and wind shear data for at least two validity times 12 hours apart. The forecast should preferably be made available in the form of grid point values and be issued in digital form and coded in a suitable format.

In practice, Area Forecast Centres should be capable of meeting these requirements and have adequate communication systems to disseminate its products preferably using the facilities and services available under World Weather Watch. This point is not discussed further here.

As regards the preprocessing of the MET information within standard pressure levels we may refer to the discussion in section 5.5. The information thus obtained, however, is unsuited to be applied directly in the graph algorithm, as in the spatial graph we should have the information available in regulated flight levels instead of standard pressure levels. The preprocessing should therefore be extended to cover the desired grid point values in these flight levels. The additional processing comprises all parameters: geopotential, wind and temperatures at all flight levels from FL 290 onwards and for all forecast times.

The geopotential and other relevant data may be deduced from the 300, 250 and 200 mbar forecasts by means of interpolation taking due account of tropopause and wind shear information.

The winds may be calculated by using the geostrophic approximation (5.23) and the finite difference expressions (5.29) and (5.30). Atmospheric temperatures affect various performance data (cf. 6.3.4). In the derivation of these performance data account is taken of the temperature effects by inclusion of a corrective term for off-standard conditions. The grid point values of temperatures therefore may best be expressed and stored as deviations from standard. Temperatures are also involved in the speed control associated with a specified cruise system.

Using for example Standard Cruise, constant Mach (cf. 2.1) the speed control is even exclusively governed by the environmental temperatures. The integration of the equations of motion may then be facilitated by calculation of the grid point values of the true airspeed, using formula (5.21) in the whole grid area and storing the data sets for each flight level.

Enabling the processing to be accomplished in space-time requires that the operational flight level data be prepared separately for the validity times of the forecasts.

The preprocessing results in the production of a great series of data sets outnumbering the data sets of the forecasts from which they were derived. The preprocessing does also a strong appeal to core memory storage, but this is compensated for by a substantial gain in runtime when the time consuming graph algorithm procedure is put into operation.

6.3.4 Performance data

As regards the performance aspects of long distance flight planning the manual procedure involves the following phases: selection of the cruise system, determination

of the take-off weight (TOW), determination of the altitude schedule, calculation of time and fuel and compilation of the flight plan.

It is outside the scope of this study to consider all these phases in detail. The subsequent rules and instructions are described in the companies' Operations Manuals. Besides, when exploring the flight plan calculation in a spatial system by computer the procedure is different from that quoted. The computation scheme starts with a known Landing Weight in the retrograde mode (6.4) and the determination of the altitude schedule forms an integral part in the track selection.

In the computational schemes (6.2) or (6.4) or (6.15) the calculation of fuel (and costs) takes place in step with that of flight time although the optimization is carried through for one of these items only. The partitioning of the spatial graph requires the processing of the performance to be considered in three phases: en route, the climb phase and the descent phase.

a. *en route*

time

The time of flight follows from the equation of motion:

$$dt = \frac{ds}{|\mathbf{c} + \mathbf{w}|}. \tag{6.20}$$

The wind vector $\mathbf{w}$ and the controlling True Air Speed (TAS) vector $\mathbf{c}$ together determine the ground speed vector $\mathbf{c} + \mathbf{w}$. Applying a Standard Cruise system (Constant Mach) the TAS $|\mathbf{c}|$ is bound to fulfil the relation (5.21). What then remains is a directional control for $\mathbf{c}$ such that the ground speed vector is directed along a flight segment (Fig. 38). For stretches of a considerable length (6.20) may be integrated by summing all contributions in the unit cells (j) of the meteorological grid, which are crossed by the flight segment.

The flight time between the end points A and B of a segment μ in the graph may then be approximated by:

$$\Delta t = \sum_{(j)} (\Delta t)_j \approx \sum_{(j)} \frac{\Delta s_j}{|\mathbf{c}_j + {}_j\mathbf{w}|}, \tag{6.21}$$

where Δs_j represents the great circle distance of a subarc in unit cell (j). The spot values $\mathbf{c}_j$ and $\mathbf{w}_j$ may be taken in the midpoints of these subarcs.

To account for the time variability of $\mathbf{w}$ and $\mathbf{c}$ the spot values may be taken with respect to the estimated time of arrival in the midpoints.

All this presupposes that the meteorological information is made available in the form of grid point values. The spot values then may be assessed using suitable space-

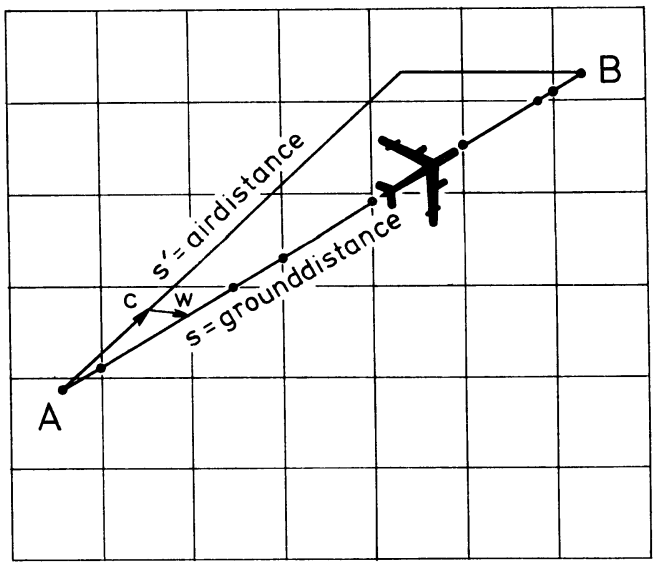

Fig. 38 Flight segment between two nodes of the graph within the meteorological grid.

time interpolation formulae in the forecast sets stored for each flight level and forecast time (See 6.3.3).

In view of the costly calculation speed we may try to simplify the summation (6.21) by defining a *fixed* directional control for **c** along the whole flight segment instead of adjusting the control in each unit cell. This may be accomplished by using the well-known principle of single heading flight between A and B. The directional control is then governed by the drift formula of Bellamy (5.51). The advantages are obvious not only from a practical point of view (fixed heading per flight segment), but also theoretical because of the property that for relatively short stretches the single heading flight is a better approximation to the optimal track than the great circle track. Adherence to a constant drift angle has no marked effect on the evaluation of the time of travel (6.21).

If the stretches in the graph are indeed not too long (6.21) may further be approximated simply by putting:

$$\Delta t = \frac{\Delta s}{|\mathbf{c} + \mathbf{w}|},$$

where Δs is the ground distance between A and B and the denominator denotes some suitable mean value of the ground speed along the flight segment, for example a geometric or arithmetic mean of the endpoint values.

fuel

The air distance covered by a turbo-jet aircraft per unit of fuel consumption – *the specific range* – is an important economic index. It can be shown that this datum is almost independent of temperature and depends on the aircraft weight W only at a prefixed flight level. Values indicative for this specific range $a\,(W)$ are known for each type of aircraft. Table I lists these values for the DC 8-D3. The fuel consumption may be expressed as loss of aircraft weight, so that by definition:

$$dW = -\frac{ds'}{a}, \tag{6.22}$$

where the *air distance* s' is related to flight time and true air speed according to

$$ds' = c\,dt. \tag{6.23}$$

Integration of (6.22) may be performed in step with the summation (6.21) along a flight segment μ between the pair of nodes (AB):

$$\Delta W = \sum_{(j)} (\Delta W)_j \approx -\sum_{(j)} \frac{c_j \cdot (\Delta t)_j}{a_j\,(W_j)}. \tag{6.24}$$

As indicated it is required to relate the spot value a_j to an estimated weight W_j, say, halfway the air distance covered in the unit cell (j).

A good estimation is:

$$W_j = W - \frac{\frac{1}{2}c_j \cdot (\Delta t)_j}{a}, \tag{6.25}$$

where W and thus a are known values at the begin point of a subarc.

In the fight against the clock we may approximate the weight-loss further by putting:

$$\Delta W = -\frac{\bar{c}\ \Delta t}{a(\overline{W})}. \tag{6.26}$$

Here Δt is the time of travel derived from (6.21).

$\bar{c}$ is a suitable mean value along the flight segment AB and $a(W)$ is the specific range approximately halfway the total air distance covered between A and B.

For $\overline{W}$ we may put:

$$\overline{W} = W_A - \tfrac{1}{2}\bar{c}\,\Delta t(r_1 + r_2\,W_A), \tag{6.27}$$

where r_1 and r_2 denote the coefficients of a straight line fit in a $\left(\frac{1}{a}, W\right)$ graph at constant pressure altitude (slicing method).

Table II lists values for these coefficients applicable to the DC 8-D3

TABLE I. *Specific range (NM/10.000 kg)*

WEIGHT (kg)	FL 250	FL 270	FL 290	FL 310	FL 330	FL 350	FL 370	FL 390	FL 410
144000	643	655	658	648	622	—	—	—	—
140000	654	668	674	670	651	—	—	—	—
136000	665	681	690	691	679	—	—	—	—
132000	678	694	705	711	706	674	—	—	—
128000	687	706	720	730	732	710	—	—	—
124000	697	718	734	749	758	744	—	—	—
120000	707	729	748	767	782	777	741	—	—
116000	717	740	761	784	805	809	784	—	—
112000	727	751	774	800	827	839	826	—	—
108000	736	762	786	815	848	868	865	830	—
104000	745	772	798	830	868	895	903	881	—
100000	754	782	809	845	887	921	939	930	884
96000	763	792	820	858	906	946	973	977	946
92000	771	801	831	872	923	970	1005	1021	1007
88000	779	811	841	884	940	992	1035	1063	1065
84000	787	820	852	896	956	1013	1064	1102	1119
80000	795	829	861	908	971	1033	1091	1139	1169
76000	803	837	871	919	985	1052	1117	1174	1217
72000	811	846	881	931	999	1070	1141	1207	1262
68000	819	855	890	941	1013	1088	1164	1237	1303

TABLE II.

FL	r_1	r_2
290	6.71	0.0567
310	4.94	0.0693
330	2.32	0.0900
350	0.54	0.1045
370	0.66	0.1010
390	—0.65	0.1170

We finally obtain:

$$W_B = W_A + \Delta W. \tag{6.28}$$

When the optimization scheme is executed in terms of fuel saving we may put in (6.2) or (6.4) or (6.15):

$$J = W, \Delta J = \Delta W$$

and put the optimization algorithm into operation in terms of a maximum criterium.

b. *climb*

As confirmed by computer experiments it may occur that the ground track of a 3-space solution deviates from the Least Time Track merely as a consequence of the climb performance characteristics. For flights with turbo-jet aircraft 10 to 15% of

the trip fuel is already consumed in the climb zone. This fuel expenditure may vary under influence of the meteorological conditions in *different directions* after take-off.

This suggests that the processing of the climb performance should be carefully studied. To calculate the effect of the climb data on the overall performance use is made of a set of climb tables. These are characteristic for a specific type of aircraft.

The tables III, IV, V, VI, VII and VIII list the data for climb time, climb air distance and climb fuel up to the top of the climb (TOC) for the DC 8-D3.

For each parameter we need to consult two tables. One affords the values under standard conditions and zero wind as a function of pressure altitude and take-off weight (TOW). The second table lists the corrections to be applied for off-standard temperature conditions. The desired information may numerically be obtained by applying suitable table look-up and table screening procedures.

TABLE III. *Climb time (1/1000 hrs)*

WEIGHT (kg)	FL 250	FL 270	FL 290	FL 310	FL 330	FL 350	FL 370	FL 390
148000	333	377	439	501	608	—	—	—
144000	317	359	414	466	553	—	—	—
140000	302	341	391	435	506	—	—	—
136000	287	325	369	408	467	590	—	—
132000	274	309	350	384	434	524	—	—
128000	261	294	331	362	405	473	638	—
124000	249	280	314	342	381	435	554	—
120000	237	267	298	324	359	406	491	—
116000	226	254	283	307	340	382	443	624
112000	216	242	269	292	322	362	406	532
108000	205	230	255	277	305	343	377	466
104000	195	219	243	263	289	325	352	418
100000	186	207	230	249	273	306	331	383
96000	176	196	218	235	258	286	310	355
92000	167	186	207	238	243	266	290	330
88000	158	176	196	210	229	245	270	307
84000	150	166	185	198	214	227	252	284
80000	142	157	174	186	200	211	235	263

TABLE IV. *Climb time (1/1000 hrs)*

Climb time (1/1000 hrs) (St. temp.)	St—20	St—15	St—10	St—5	St+5	St+10	St+15	St+20
50	45	46	47	48	53	56	59	62
150	129	134	140	145	158	165	175	185
250	205	213	222	236	267	285	307	330
350	273	287	300	325	390	430	475	520
450	335	352	375	412	512	585	675	820
550	387	407	443	493	640	767	950	1295
650	425	453	504	568	782	985	1300	—
750	454	490	556	640	933	1245	—	—

TABLE V. *Climb distance (N.M.)*

WEIGHT (kg)	FL 250	FL 270	FL 290	FL 310	FL 330	FL 350	FL 370	FL 390
148000	105	126	151	181	232	—	—	—
144000	99	118	142	167	207	—	—	—
140000	94	111	133	154	187	—	—	—
136000	89	105	125	144	170	226	—	—
132000	85	99	118	135	156	198	—	—
128000	80	94	111	127	145	177	253	—
124000	76	89	105	119	135	160	216	—
120000	73	85	100	113	127	147	188	—
116000	69	80	95	107	119	136	167	250
112000	66	76	90	101	112	128	151	209
108000	63	73	85	96	106	120	139	179
104000	59	69	81	91	100	113	129	158
100000	56	65	77	86	94	107	120	143
96000	53	62	73	81	89	100	113	131
92000	51	58	69	76	83	94	105	121
88000	48	55	65	71	77	87	97	113
84000	45	52	61	67	72	81	90	104
80000	42	49	57	63	68	75	83	96

TABLE VI. *Climb distance (N.M.)*

Climb distance (N.M.) (St. temp.)	St—20	St—15	St—10	St—5	St+5	St+10	St+15	St+20
0	0	0	0	0	0	0	0	0
50	38	40	43	47	54	59	68	75
100	73	79	84	93	110	123	141	157
150	105	113	123	136	168	190	222	268
200	131	144	160	179	230	260	322	460
250	150	168	191	218	296	334	440	—
300	161	187	219	254	372	417	—	—

The calculation of performance data in the climb zone is straightforward if the computational algorithm starts at the point of departure for a given TOW.

Let P^*B (Fig. 36a) be a flight portion at a given pressure altitude in the climb zone of the graph. Then the performance data are composed of the climb performance up to TOC and the en route performance between TOC and first check point B.

The calculation is somewhat complicated by the fact that the location of TOC is not fixed. The position where the flight levels off may vary due to the influence of the winds in the climb phase. The actual climb distance including wind effect is:

$$(\Delta s)_{\text{climb}} = |\bar{c}_{\text{climb}} + \bar{w}_{\text{climb}}| \cdot (\Delta t)_{\text{climb}}, \tag{6.29}$$

TABLE VII. *Climb fuel (kg)*

WEIGHT (kg)	FL 250	FL 270	FL 290	FL 310	FL 330	FL 350	FL 370	FL 390
148000	4240	4670	5150	5640	6500	—	—	—
144000	4050	4440	4890	5310	5980	—	—	—
140000	3870	4230	4640	5010	5540	—	—	—
136000	3700	4030	4410	4730	5160	5950	—	—
132000	3540	3840	4190	4480	4830	5440	—	—
128000	3380	3670	3990	4250	4550	5020	6210	—
124000	3240	3500	3800	4040	4300	4680	5490	—
120000	3090	3340	3620	3840	4070	4380	4950	—
116000	2960	3190	3450	3650	3870	4130	4530	5570
112000	2830	3050	3290	3470	3680	3910	4210	4930
108000	2700	2900	3130	3300	3500	3710	3950	4450
104000	2580	2770	2980	3140	3320	3520	3720	4090
100000	2460	2630	2840	2980	3150	3340	3520	3800
96000	2340	2500	2690	2830	2980	3160	3330	3570
92000	2230	2380	2560	2690	2820	2980	3140	3350
88000	2120	2250	2420	2540	2650	2800	2950	3140
84000	2010	2130	2290	2400	2500	2630	2750	2940
80000	1900	2010	2160	2270	2350	2470	2570	2750

TABLE VIII. *Climb fuel (kg)*

Climb fuel (kg) (St. temp.)	St—20	St—15	St—10	St—5	St+5	St+10	St+15	St+20
500	460	470	480	490	515	530	545	560
1500	1430	1445	1460	1480	1550	1600	1645	1690
2500	2220	2280	2340	2420	2620	2740	2870	3000
3500	2950	3050	3180	3350	3700	3920	4200	4470
4500	3680	3850	4050	4280	4850	5150	5670	6320
5500	4370	4590	4880	5150	6000	6670	7530	8520
6500	5050	5260	5630	5950	7200	8500	9900	12150
7500	5630	5900	6280	6720	8700	11050	—	—

$\bar{\mathbf{c}}_{\text{climb}}$ is the mean projected TAS vector in the climb, $\bar{\mathbf{w}}_{\text{climb}}$ the mean climb vector wind:

$$|\bar{\mathbf{c}}_{\text{climb}}| = \frac{(\Delta s')_{\text{climb}}}{(\Delta t)_{\text{climb}}}, \tag{6.30}$$

For $\bar{\mathbf{w}}_{\text{climb}}$ we may take 60 or 70% of the 300 mbar vector wind. The direction of $\bar{\mathbf{c}}$ may be inferred from the drift formula of Bellamy using the geopotential altitude in P overhead (P^*) and B. The climb time $(\Delta t)_{\text{climb}}$ is usually reduced by a few minutes to account for take-off and acceleration.

We may now proceed as follows:

1. determine the en route flight time $(\Delta t)_{P^*B}$ between P overhead and B using (6.21);
2. estimate the temperature deviation from standard at the given pressure altitude, for example at P overhead;
3. determine the climb air distance $(\Delta s')_{\text{climb}}$, climb time $(\Delta t)_{\text{climb}}$ and climb fuel $(\Delta W)_{\text{climb}}$ entering the tables III till VIII;
4. determine the climb distance $(\Delta s)_{\text{climb}}$ using (6.29);
5. determine the ground distance Δs between P^* and B;
6. determine from (3) the weight at TOC:

$$W_{\text{TOC}} = \text{TOW} - (\Delta W)_{\text{climb}}.$$

7. determine from (1), (4) and (5) the flight time en route between TOC and B:

$$(\Delta t)_{\text{en route}} = \frac{\Delta s - (\Delta s)_{\text{climb}}}{\Delta s}(\Delta t)_{P^*B}.$$

8. determine from (6) and (7) the fuel en route between TOC and B:

$$(\Delta W)_{\text{en route}} = \frac{c\ (\Delta t)_{\text{en route}}}{a}.$$

9. determine the flight time between P and B including climb time:

$$\Delta t = (\Delta t)_{\text{climb}} + (\Delta t)_{\text{en route}}.$$

10. determine the fuel consumption between P and B including climb fuel:

$$\Delta W = (\Delta W)_{\text{climb}} + (\Delta W)_{\text{en route}}.$$

11. determine the aircraft weight in B:

$$W_B = \text{TOW} - \Delta W.$$

It should be stressed that this scheme is applicable only if the take-off weight is known. When the algorithm is carried out using the retrograde version the climb performance calculation constitutes the last stage of the process. Then the weight W_B is known, but the TOW remains to be evaluated. To that aim an iteration search for TOW is put into action. Starting with an estimate of TOW (e.g. the maximally allowable TOW) a preliminary value W_B' is derived and compared with W_B.

In a next iteration step a correction is applied to TOW proportionally with $W_B' - W_B$ and a new value for W_B' determined. This process is repeated until $W_B' - W_B$ comes within prescribed bounds (5 à 10 kg).

c. *descent*

The descent performance is no critical factor in the search for an optimum. As regards the processing in the descent zone the method to be followed is quite similar to the one applied in the climb. A distinction is that no use is made of tabulated values for the performance data, but of some simple analytical expressions instead. These are specified in accordance with the company's planning policy. In our experiments for the DC 8-D 3 we have used the following expression:

$$\begin{aligned} (\Delta t)_{\text{descent}} &= 6.1 + 0.41\,z \text{ (min.)}. \\ (\Delta s)_{\text{descent}} &= (1 + 0.004\,\Delta T).(9 + 3\,z) \text{ (NM)}. \\ (\Delta W)_{\text{descent}} &= 940 + 8.5\,z \text{ (kg)}. \end{aligned} \tag{6.31}$$

z denotes flight altitude in 10^3 feet and ΔT the temperature deviation from standard (^{0}C).

d. *costs*

The cost-saving potential is of a substantial economic significance. The economic benefits improve both by a reduced fuel burn-out and time-saving. The costs may mathematically be expressed as a function of fuel burn and flight time.

The simplest way is to express Δc linearly in Δt and ΔW:

$$\Delta c = c_1\,\Delta t + c_2\,\Delta W \tag{6.32}$$

where the coefficients c_1 and c_2 are specified in accordance with the company's flight planning policy.

In practice the calculation of Δc takes place in step with the fuel- and time calculation, so that the optimization in terms of a cost criterium might be performed simply by using the subroutine for fuel- or time optimization.

It is to be expected as confirmed by experiments that the vertical cost profile will be somewhere between the fuel- and time profiles. (See Fig. 41).

6.3.5 The graph algorithm

To apply the graph algorithm in the form (6.2), (6.4) or (6.15) one single computer

routine may be developed. To that aim the programme might be structured so as to admit the execution of all these modes simply by deleting or reversing some cycles and switching a few plus or minus signs in the computation scheme.

The objective function J is identified with the planned quantity to be optimized: aircraft weight, time of travel or costs.

The process evolves in stages starting either in the climb zone or descent zone. The calculation of the performance data is straightforward except in the last stage. The descent- or climb performance in the last step should be matched to the en route performance by means of an iteration procedure (See 6.3.4. *b.*).

A salient feature is the requirement to protect the process algorithm against a possible transgression of the maximum cruise thrust due to weight limitations.

At each point on the optimal path the aircraft weight is bound to remain below a maximally allowable weight. This maximally allowable weight is dependent on flight altitude and temperature deviation from standard. Table IX records these weight limits as a function of flight level and off-standard temperatures for the DC 8-D 3. In the computer routine the control with respect to this weight restriction is carried out for each flight segment in the endpoint where the performance weight is expected to be the highest. As soon as the actual weight surpasses the maximally allowable weight derived from table IX, the segment is blocked and the processing continued. The weight control should be applied in all flight segments as well as in the TOC and TOD.

It is also required to extend the control to the landing weight and take-off weight, which are bound to remain below the corresponding maximally allowable weight limits. Weight limits appropriate in the climb and descent are accounted for already in the tables III till VIII (empty spaces).

Under circumstances mainly attributable to unfavourable meteorological conditions and highly restricted airspace utilization it may occur that the computer routine fails to find a solution. The flight happens to become payload critical. In that case the TOW or LW should be lowered and a new computer run be performed.

TABLE IX. *Maximum weight (kg) per flight level*

	St—10	St—5	Stand.	St + 5	St + 10	St + 15	St + 20
FL 410	97700	96000	94200	92500	90900	89000	85800
FL 390	108200	106300	104300	102600	101000	99100	95600
FL 370	119900	117900	115600	113800	112100	100300	106300
FL 350	131800	129300	126800	124800	122800	120700	116900
FL 330	max.	140100	137000	134900	133200	130600	126600
FL 310	max.	max.	max.	max.	max.	141000	135800

A salient feature is further that the process should in any case entail a calculation of the fuel burn or weight performance uncoupled with the optimality criterium.

As a consequence it should be decided upon whether the process should start in the point of departure or in the point of arrival. Obviously this depends on whether the take-off weight or landing weight is known in advance.

To comment on this we consider the scheme of possible weight- and fuel desiderata, which are indicative for the main entries in a flight plan.

trip fuel	flight plan sub-total	ramp weight (RW)
		taxi fuel (TF)
		start take-off run (TOW)
		climb consumption
		top of climb weight (TOC)
		cruise consumption
		top of descent weight (TOD)
		descent consumption
		landing weight for performance (LW*)
		regularity percentage $r\%$ of (RW-LW*)
		landing weight (LW)
		reserves
		zero fuel weight (ZFW)

This list is only tentative. All items in this scheme are variable with the exception of LW. This quantity involves the zero fuel weight and reserves or more specifically: operating weight empty, payload and reserves.

One could, of course, begin with a prescribed TOW, at the risk however of exceeding the maximally allowable landing weight. It is more profitable to use the 'pivotal' term LW instead and to perform the algorithm in the retrograde mode.

Here again, a difficulty arises as the algorithm should not start with LW sec, but with the landing weight for performance LW*.

LW* is not precisely known however, so that we should try to find an estimate $\hat{\text{LW}}^*$ dependent on type of aircraft and mileage covered. For the DC 8-D3 we have used the expression:

$$\hat{\text{LW}}^* = \text{LW} + r\left(\frac{\text{TAS} \cdot \Gamma}{465}(0.08\,\text{LW} - 500) - 0.15\,\text{LW} + \text{TF} + 7300)\right) \quad (6.33)$$

Γ is an estimate of the flight time, TAS the mean air speed en route. TAS . Γ is therefore an estimate of the air distance covered between the endpoints. LW*, LW and taxi-fuel are given in kg.

Starting with $\hat{\text{LW}}^*$ a first run is completed in the basic 3-space graph. For reasons explained in section 6.2.3 *c.* the first run is not considered decisive, but it is advised to continue with a second process run within the 2-space graph in the vertical plane through the optimum ground track, which was identified in the first run. If this second run would start with the TOW as obtained in the first run, the routine would merely afford a duplicate of the optimal step altitude profile and associated performance

data. The purpose of the recomputation is however to start with an improved TOW or LW* in order to remove the coarse approximations of LW* in the first run.

For turbo-jet flights the following relation holds to a good approximation:

$$R = \frac{\text{LW*}}{\text{TOW}} = \text{constant} \tag{6.34}$$

Furthermore by definition:

$$\text{LW*} - \text{LW} = r(\text{RW} - \text{LW*}) \tag{6.35}$$

and

$$\text{RW} = \text{TOW} + \text{TF} \tag{6.36}$$

For a new estimate of LW^* we may pose

$$\hat{\text{LW}}^* = \text{LW} + r\,(\hat{\text{TOW}} + \text{TF} - \hat{\text{LW}}^*)$$

and

$$\text{R} = \frac{\hat{\text{LW}}^*}{\hat{\text{TOW}}}$$

so that the estimate for TOW becomes:

$$\hat{\text{TOW}} = \frac{LW + r\,TF}{(1+r)R - r} \tag{6.37}$$

The algorithm is performed in the second run starting with this TOW estimate. After completion of the second run the routine is activated once more, but now in the 1-space graph consisting of the optimal altitude profile as identified in the second run. The intention of this third run is merely to make a final calculation of all entries in the flight plan and to present these in an appropriate flight plan format. The speed of computation of the third run is negligible compared with the first and second run. In passing the third run may be used to start the computation with another improved estimate for the take-off weight.

In order to investigate the feasibility of automation of this process in practice we have developed an experimental software package in ALGOL and tested this on a second generation computer. In view of its experimental character we abstain here from giving a detailed description of the process architecture and programme structure. The results of some of these experiments are given in the next section.

6.4 Experimentation

A series of computer experiments was planned to achieve an insight in the potentialities of the space-time flight planning process. The main purpose of these experiments was to prove that, in every respect, the spatial algorithm is a more powerful tool for accurate track selection than the conventional 2-dimensional method. In one sense this is a trivial question, for the conventional method is nothing but a special case of the spatial process (cf. 6.2.3 *c*.). The question is however not trivial if we consider the possible limitations in the practical application of the process in terms of computation time, core memory capacity, ease of programming, etc. Apart from this it is interesting to investigate what effect the meteorological conditions have on the accuracy of the optimal track selection and whether the extra gain in performance obtained is worth the price and effort. Finally, if the practical evaluation of the algorithm in its most primitive form (6.2) or (6.4) might appear illusive with the present art of systems design and technology, then it is very important to demonstrate that the approximate scheme (6.15) is a useful interim solution.

The experiments were set up for transoceanic flights over the North Atlantic, especially for flights between Amsterdam and New York vice versa.

In the process use is made of upper air forecasts for 300, 250 and 200 mbar for two validity times 12 hours apart (See Fig. 39a, b, c, d, e, f).

The meteorological situation as depicted from Fig. 39 is characterized by a pronounced meridional circulation with 'blocking' high west of Ireland. Off the Canadian coast there is a deep trough extending far south over the Bermuda area moving slowly to the east. The winds are moderate, so that it may be expected that the track diversion and gain in performance will not be spectacular compared with the great circle route. Notable is the strong temperature variation along the route. The amplitude of the temperature wave is of the order of 25 ^{0}C, which probably will affect the altitude profile.

The MET information is supplied in digital form. Geopotential values and temperature are depicted from the grid points in a Cartesian grid superposed on a polar stereographic chart projection with standard parallel at 60 ^{0}N. This grid is part of the well-known octagonal grid prepared by the National Meteorological Center, Suitland, Md. The y-axis runs parallel with the 80 ^{0}W meridian. The area covered is a rectangle comprising 12 × 16 grid points.

The number of operational MET-data (temperature and geopotential) to be supplied is therefore

$$2 \times 2 \times 3 \times 12 \times 16 = 2304.$$

The spatial graph structure may be specified by the sets of check points and flight levels. Fig. 37 shows in projection the groups of check points. The flight levels selected for these experiments were FL 310, FL 350 and FL 390.

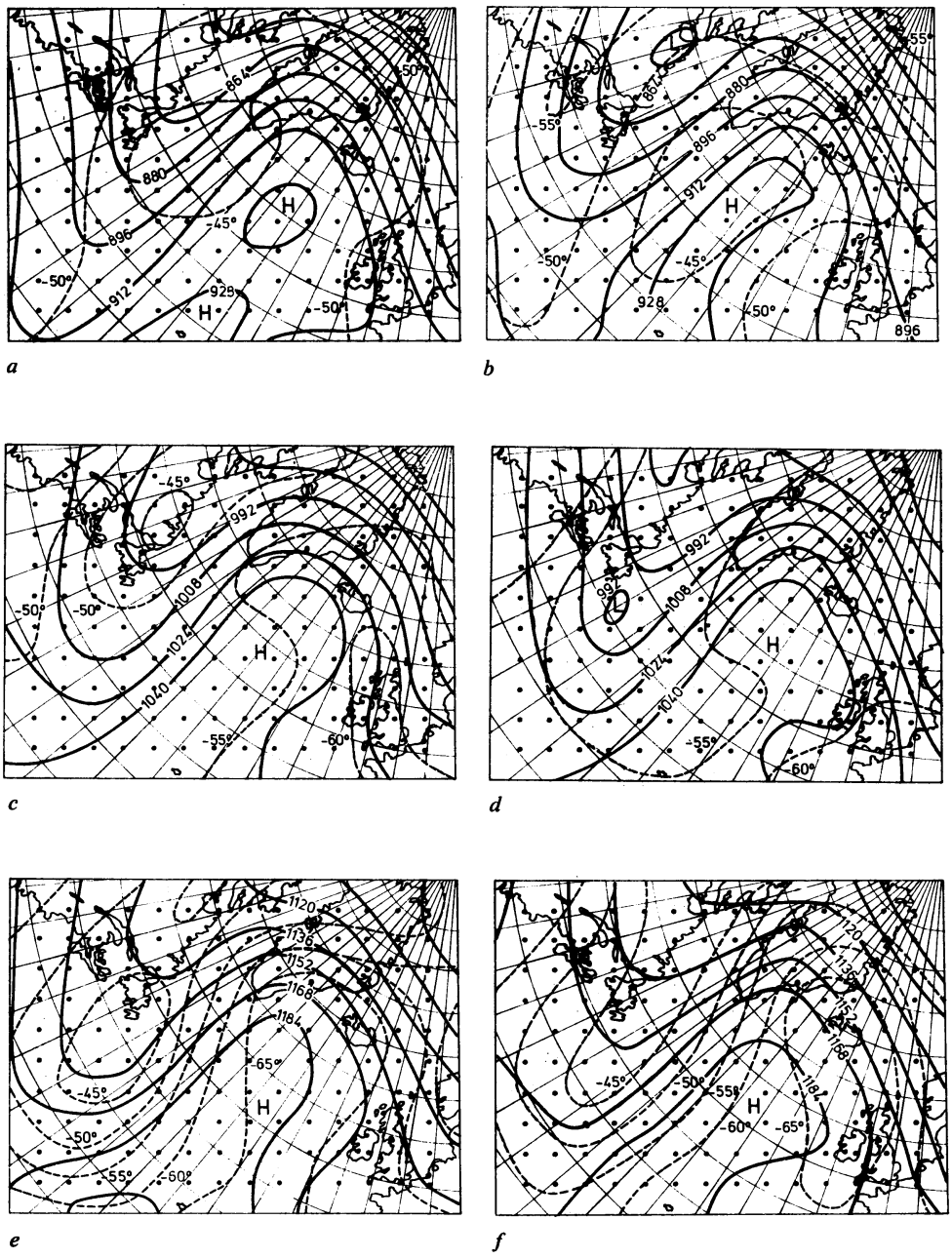

Fig. 39 Analysis of isobaric topographies. Contours labelled in gpdam (heavy lines). Temperature labelled in °C (dashed lines).

a 300 mbar 00.00 GMT
b 300 mbar 12.00 GMT
c 250 mbar 00.00 GMT
d 250 mbar 12.00 GMT
e 200 mbar 00.00 GMT
f 200 mbar 12.00 GMT

The input of performance data comprises the tabulated values for the specific range (Table I, II), maximum cruise weight (Table IX) and six tables for the climb time, climb fuel and climb distance. The tabulated values are representative for the type of aircraft DC 8-D 3.

The selected cruise system is standard cruise (Constant Mach number technique).

At the time the series of trials was planned it was feared that the limited storage capacity of the available computer system would be insufficient for mastering the processing. For that reason measures were taken to keep the amount of computational work to a strict minimum. The number of check points and flight levels as well as the area of the rectangle were reduced as much as possible. In addition, a measure was taken to restrict the scanning of arcs between nodes (i, k, l) and a given node $(i + 1, K, L)$ within a *sector* about the main flight direction. This was considered acceptable inasmuch *zigzag routes* are highly improbable.

In the last few years, however, the technology has made enormous progress, so that in modern complex machinery these measures might have become obsolete.

In the series of experiments use is made of the following desiderata for the DC 8-D 3 aircraft:

standard cruise, Mach		0.8034
taxi fuel		1500 kg
zero fuel weight		75000 kg
reserves		10000 kg
maximal take-off weight		142900 kg
maximal landing weight		93000 kg
regularity %		3 %
step up	extra burn	25 kg/1000 ft
step down	extra burn	−20 kg/1000 ft
flight levels		310, 350, 390
cost coefficients		c_1 = 6.0/min
(currency unspecified)		c_2 = 0.02/kg
time of departure		0300 GMT

All trials made a start in the retrograde mode of (6.2) or (6.15) with the following preliminary value of the landing weight: LW = 85000 kg (zero fuel weight + reserves, 75000 + 10000 kg).

1. The use of a computer programme needs a reasonable assurance of the reliability of the programme. A kind of minimal check is to run the programme under standard atmospheric conditions and zero wind. The optimal track should then consist of the 'optimal operational distance track' in the graph as a good approximation of the great circle

track between the endpoints. This result should be independent of the chosen optimality criterium.

In the vertical the altitude step profile and en route performance data should then be standard as may be verified by inspection of diagrams like the flight planning check chart and time-fuel chart. The results should be unaffected when the programme is run for a flight in opposite direction between the same endpoints (westbound versus eastbound).

All runs produced the optimal operational distance track as shown in Fig. 37.

Table X lists the results of a computation based on fuel economy for the traject Amsterdam - New York, westbound, obtained by computer and manually using a flight planning check chart and time fuel chart.

TABLE X. *Standard meteorological conditions, zero wind. Graph algorithm (6.2). Criterium: fuel saving. Amsterdam-New York, Standard Cruise, Mach 0.803.*

method	LW (kg)	TOW (kg)	distance (NM)	duration* (hrs, min.)	trip fuel** (kg)
computer	85057	124446	3193	7h 04′	39389
manual***	85100	124446	3193	7h 05′	39350

* incl. 2 min. for take-off and acceleration.
** excl. taxi fuel, incl. 3% regularity.
*** flight planning check chart was entered with TOW = 124446 kg and distance: 3193 NM.

The time fuel chart indicates that two steps are required, 25′ and 4h 15′ after takeoff. The programme required two *fixed* step locations – coinciding with check points – at 44′ and 4h 30′ after take-off.

Table XI summarises some flight plan totals for calculations in terms of time-, fuel- and cost saving, westbound and eastbound (6 runs).

TABLE XI. *Standard meteorological conditions, zero wind. Graph algorithm (6.2). Amsterdam-New York, Standard Cruise, Mach = 0.803.*

optimality criterium	westbound			eastbound		
	duration* (hrs, min.)	fuel** (kg)	cost	duration (hrs, min.)	fuel (kg)	cost
time saving	6h 58′	41700	10784	6h 58′	41732	10788
fuel saving	7h 04′	38242	10331	7h 04′	38229	10329
cost saving	7h 04′	38242	10331	7h 04′	38229	10330

* incl. 2 min. for take-off and acceleration.
** trip fuel excl. taxi fuel, excl. regularity %.

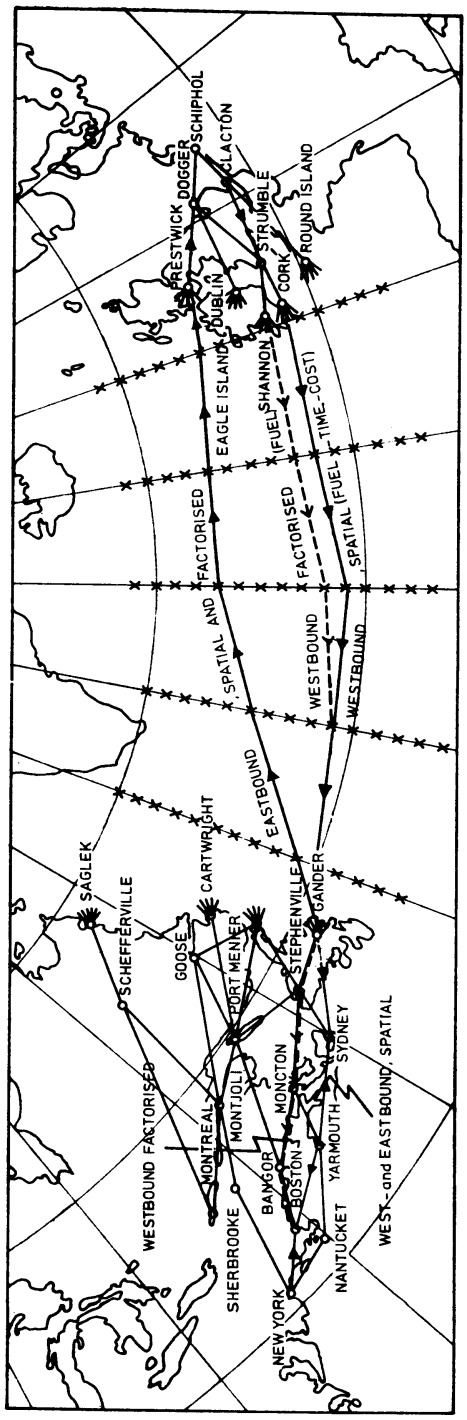

Fig. 40 Ground tracks associated with the spatial optimum flight paths based on time-, fuel- or cost-saving. New York-Amsterdam, east -and westbound (heavy lines). Dashed line: ground track of the solution produced by factorization in terms of least time (horizontally) and least fuel consumption (vertically). For meteorological conditions see Fig. 39.

The speediest flights should take place constantly at the lowest flight level, where the highest temperature prevails (FL 310). This was confirmed experimentally. The results are in fairly good agreement. Some minor discrepancies may be attributed to discretization-, interpolation- and round-off-errors.

The correctness of the program may also be proved by comparing the computer products in off-standard meteorological conditions with the corresponding results obtained by manual dispatch procedures.

2. Intriguing of all is to explore how the process functions when the optimization is performed in non-standard atmospheric conditions. Again six runs were made using the general graph algorithm (6.2) for flights between Amsterdam and New York, eastbound and westbound, in terms of time-, fuel- and cost-saving. Some of these results are shown in Fig. 40 and 41 and in table XII.

First we see from Fig. 40 that the ground tracks westbound as well as eastbound are in triplo. Apparently the criterium of optimality had no effect at all on the location of the ground track. This may be pure chance.

TABLE XII. *Graph algorithm (6.2). Amsterdam-New York, Standard Cruise, Mach 0.803.*

optimality criterium	westbound			eastbound		
	duration* (hrs, min.)	fuel** (kg)	cost	duration (hrs, min.)	fuel (kg)	cost
time saving	*7h 11'*	40485	10739	*6h 43'*	39932	10357
fuel saving	7h 14'	*39347*	10608	6h 51'	*36366*	9502
cost saving	7h 14'	39384	*10606*	6h 51'	36366	*9502*

* incl. 2 min. for take-off and acceleration.
** trip fuel excl. taxi fuel, excl. regularity %.

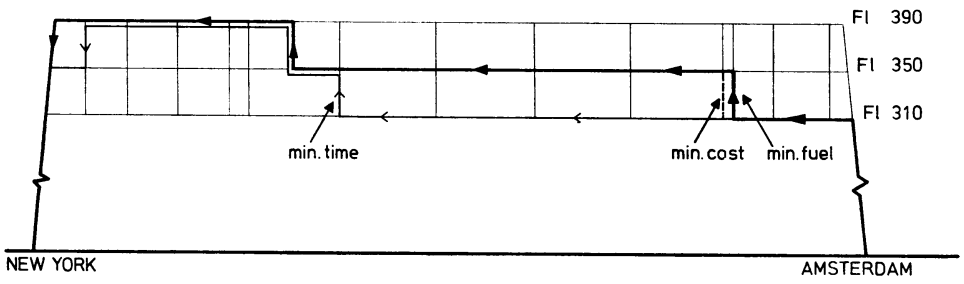

Fig. 41 Optimum altitude profile associated with the spatial optimum flight paths based on fuel-, time- or cost-saving. New York-Amsterdam. For the corresponding ground track see Fig. 40. Steps are located at check points.

The graph in the vertical (Fig. 41) westbound (eastbound is not shown) clearly indicates that the corresponding optimal step altitude profiles are different. The fuel and cost profiles have both the normal appearance: approximation of a continuous cruise climb with one or two 4000 ft steps. The flight time profile is more irregular. This is caused by the response of the aircraft's speed to the vertical temperature gradients. The level of highest temperatures is sought, which is clearly demonstrated in the last portion of the flight in the American Airway zone. Temperatures are lower there at 300 mbar than higher aloft in the stratosphere.

The values taken for the cost coefficients c_1 and c_2 make that the fuel costs take approximately 75% of the total costs. This explains why the cost profile is often coincident with the fuel profile or hardly deviates from it. This accounts also for the small differences in the corresponding performance data (see table XII) in terms of fuel- and cost-saving. It is worthwhile to observe the consistency of the extreme values in the table (in italics). The consistency is required as an intrinsic property of the general optimization algorithm.

3. The advantages of introducing a spatial flight planning process will become obvious when proving its efficiency with respect to the conventional method. This method consists of a separate track selection in the horizontal followed by a search of the optimal altitude profile in the vertical.

An experiment was planned whereby the spatial process was forced to produce the Least Time Track (at FL 310) and subsequently the optimal fuel profile in the vertical plane through the ground track. The outcome was then compared with that obtained in the 3-dimensional approach in terms of fuel saving. It turned out that, eastbound, both the optimal flight path and performance were exactly the same. Westbound, however, the results differed as indicated by a diversion of ground tracks (see dashed line in Fig. 40) and the content of table XIII.

TABLE XIII. *Graph algorithm (6.2). Amsterdam-New York, westbound, Standard Cruise, Mach 0.803.*

method	duration (hrs, min.)	fuel (kg)	cost
spatial fuel saving	7h 14′	39347	10608
Least Time Track plus fuel saving	7h 15′	39449	10632

This experiment clearly proves that a solution in 3-space may indeed be rewarding as compared with a (factorized) 2-space solution. The *extra* gain in performance is not inconsiderable. This is the result of one case study only. It would require a long series

of trials to get an insight in the efficiency of the spatial flight planning process as compared with the traditional method.

4. Another experiment involves a test of the usefulness of the graph algorithm in reduced form. Its significance in practice is the substantial reduction of computation time provided that the loss in accuracy remains within acceptable bounds.

The algorithm is applicable if the saving in fuel burn in a step-down be taken equal to the extra fuel burn in a step-up (−25 kg/1000 ft). The computer runs are the same as described in 2). Table XIV shows the results of these runs.

TABLE XIV, *Graph algorithm (6.15). Amsterdam-New York, Standard Cruise, Mach 0.803.*

optimality criterium	westbound			eastbound		
	duration (hrs, min.)	fuel (kg)	cost	duration (hrs, min.)	fuel (kg)	cost
time saving	*7h 11'*	40476	10736	*6h 43'*	39931	10358
fuel saving	7h 14'	39401	10633	6h 51'	*36378*	9909
cost saving	7h 14'	39348	*10607*	6h 51'	36378	*9909*

The differences observed between the flight plan totals in table XII and table XIV may be considered immaterial from an operational point of view. The optimal paths were identical with those identified by algorithm (6.2) in experiment 2 (See Fig. 41). The execution time contrasted well with the general graph algorithm: 3 minutes against 25 minutes on a 1 μs core assess time machine. A feature which is somewhat disturbing is the possibility that the consistency of extreme values may be destroyed. (See westbound: fuel saving). The optimization is apparently very sensitive to slight modifications of the general process scheme.

Nevertheless, when the computation speed of the general algorithm forms an obstacle we consider the specialized scheme as a good substitute for it.

ANNEX

The control equations associated with the conformal mapping of a uniform great circle motion

The control equations governing the navigation along the Least Time Track have been generalized in section 4, in order to apply them in a conformal mapping of the earth's surface onto a plane surface.

The resulting equations involve a corrective term in terms of the scale of the mapping.

In the special case of a flight with constant airspeed and zero wind conditions these control equations should describe the mapping of a uniform great circle motion onto the plane.

Referring to section 4, equations (4.10) the component equations to be considered are:

$$\dot{\xi} = -\mathbf{c}^{*\mathbf{T}} . \frac{\nabla S}{S}, \tag{1}$$

$$\dot{c} = 0. \tag{2}$$

These equations hold with respect to a fixed Cartesian coordinate system in the image plane.

The proof will be given in terms of the Lambert conformal conic projection. The proof is then valid also for the polar stereographic projection and the cylindrical Mercator projection, being special cases of the former Lambert projection.

All these mappings may be conceived as conformal mappings onto a plane surface, since the cone and cylinder may be unrolled to form the plane map.

First we shall review the mathematical description of the Lambert conformal projection [23].

Consider Fig. I, 1.

In a conformal conic projection with a tangent cone enveloping the northern hemisphere the projection of a latitude circle (latitude φ) has radius:

$$r = a\frac{S_0}{p}\{\tan(\pi/4 - \varphi/2)\}^p. \tag{3}$$

In this case $p = \sin\beta$, where β represents half the angle of the vertex of the cone. a is the earth's radius and S_0 denotes the map scale at standard parallel:

$$S_0 = \cos\beta \left\{ \tan\left(\frac{\pi}{4} - \beta/2\right)\right\}^{-\sin\beta} \tag{4}$$

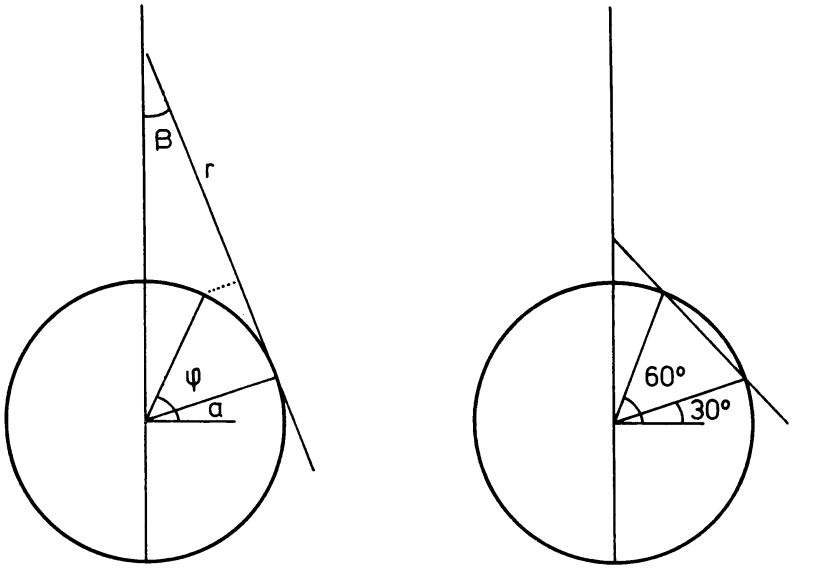

Fig. I, 1 Scheme of the conic polar and Lambert conic conformal projections.

Differentiation of (3) with respect to φ gives

$$\frac{dr}{d\varphi} = -\frac{a\,S_0}{(\cos\varphi)^{1-p}(1+\sin\varphi)^p}$$

The scale S, which on a conformal map must be the same in all directions through any point, is by definition:

$$S = \frac{1}{a}\left|\frac{dr}{d\varphi}\right| = \frac{S_0}{(\cos\varphi)^{1-p}(1+\sin\varphi)^p} \tag{5}$$

The WMO recommends use of the so-called 'secant' projections instead of tangent projections. In secant projections the image surface intersects the surface of the earth. There are two standard parallels in the Lambert projection, 30 and 60^0 N (Fig. I, 1) and Mercator projections and only one in the polar stereographic. The expression (5) for the scale of the mapping applies to all these mappings.

$p = 0.716$ for the Lambert conformal projection;
$p = 0$ for the Mercator projection;
$p = 1$ for the polar stereographic projection.

Using (5) a straightforward calculation affords:

$$\frac{1}{a}\frac{dS}{d\varphi} = \frac{S}{a}\left(\frac{\sin\varphi - p}{\cos\varphi}\right).$$

Let **j** denote a unit vector pointing northwards along a meridian on the sphere (Fig. I, 2), then we may establish the gradient of the scale factor:

$$\nabla S = \frac{1}{a}\frac{dS}{d\varphi}\mathbf{j}$$

or

$$\nabla S = \frac{S}{a}\left(\frac{\sin\varphi - p}{\cos\varphi}\right)\mathbf{j}. \tag{6}$$

Furthermore we need an expression for $\frac{S}{r}$, compare (3) and (5):

$$\frac{S}{r} = \frac{p}{a\cos\varphi} \tag{7}$$

Next, assume that a vehicle moves along a great circle arc on the sphere with constant velocity c. At an arbitrary point P we consider a Cartesian coordinate system Pxy with x-axis pointing to the east and y-axis towards the north.

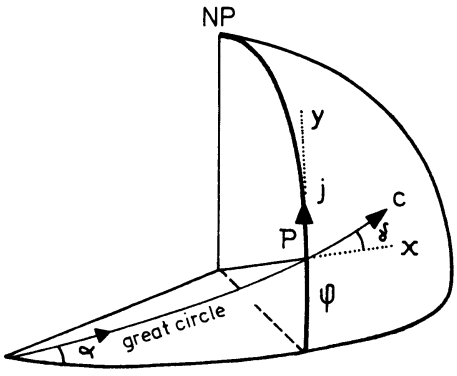

Fig. I,2 Uniform great circle motion on the sphere.

Then from Fig. I, 2 we derive

$$\cos\alpha = \cos\gamma\cos\varphi. \tag{8}$$

The great circle motion does not affect the angle α between the plane of motion and the equator:

$$\dot{\alpha} = 0.$$

Differentiation of (8) with respect to t gives:

$$\dot{\gamma} = -\cot\gamma\tan\varphi\,\dot{\varphi},$$

but

$$a\dot{\varphi} = c\sin\gamma,$$

so that

$$\dot{\gamma} = -\frac{c}{a} \cos \gamma \tan \varphi. \tag{9}$$

The uniformity of the great circle motion is expressed by

$$\dot{c} = 0. \tag{10}$$

The set of equations (9) and (10) constitute the set of control equations for the uniform great circle motion with respect to a comoving *Pxy* system on the sphere.

If the great circle motion is mapped conformally onto a plane surface then the form of equation (9) is preserved, but it should be borne in mind that the equation holds with respect to the mapping of the *xy* system onto the plane. To derive the corresponding control equation with respect to a *fixed* coordinate system $Ox'y'$ on the image plane it is necessary to find a corrective term.

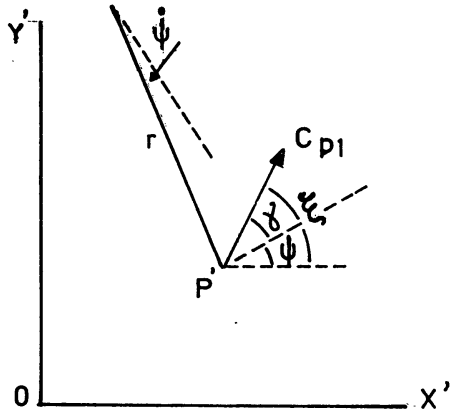

Fig. I, 3 Scheme of the great circle motion on a conformal projection map.

From Fig. I, 3 we see that

$$\xi = \gamma + \psi,$$

where ξ denotes the angle between $\mathbf{c}_{\mathrm{pl.}}$ and the x'-axis.

$\mathbf{c}_{\mathrm{pl.}}$ is the projection of $\mathbf{c}$ on the image plane.

ψ is the angle between the latitude circle and the x'-axis.

Then

$$c_{\mathrm{pl.}} = Sc,$$

$$\dot{\xi} = \dot{\gamma} + \dot{\psi}, \tag{11}$$

but

$$\dot{\psi} = \frac{Sc}{r} \cos \gamma. \tag{12}$$

Substitution of (9) and (12) into (11) affords

$$\dot{\xi} = \frac{c}{a}\left(\frac{aS}{r} - \tan \varphi\right) \cos \gamma. \tag{13}$$

According to (7)

$$\dot{\xi} = \frac{c}{a}\left(\frac{p}{\cos\varphi} - \tan\varphi\right)\cos\varphi,$$

$$\dot{\xi} = \frac{c}{a}\left(\frac{p - \sin\varphi}{\cos\varphi}\right)\cos\gamma.$$

But in view of the conformity of the mapping we may put:

$$\cos\gamma = \frac{\mathbf{c}^{*\mathbf{T}}}{c}.\mathbf{j},$$

resulting in

$$\dot{\xi} = -\frac{1}{a}\left(\frac{\sin\varphi - p}{\cos\varphi}\right)\mathbf{c}^{*\mathbf{T}}.\mathbf{j}$$

or, in view of (6):

$$\dot{\xi} = -\mathbf{c}^{*\mathbf{T}} \cdot \frac{\nabla S}{S}. \tag{14}$$

Comparison of the set of equations (10) and (14) with (1) and (2) proves that by specialization the great circle track may indeed be generated by the equations which govern the navigation along the Least Time Track.

This result applies equally well for the polar stereographic and Mercator conformal projection being special cases of the Lambert conformal conic projection.

Returning to the control equations (1) and (2) they might be put in the form

$$\dot{\xi} = c\frac{|\nabla S|}{S}\sin i, \tag{15}$$

$$\dot{c} = 0.$$

Where i denotes the angle taken in counter clockwise direction between $\mathbf{c}$ and ∇S (Fig. I, 4).

For the polar stereographic projection ($p = 1$) the control set becomes, see (6):

$$\dot{\xi} = -c\frac{\sin i\ \cos\varphi}{a(1 + \sin\varphi)}, \tag{16}$$

$$\dot{c} = 0.$$

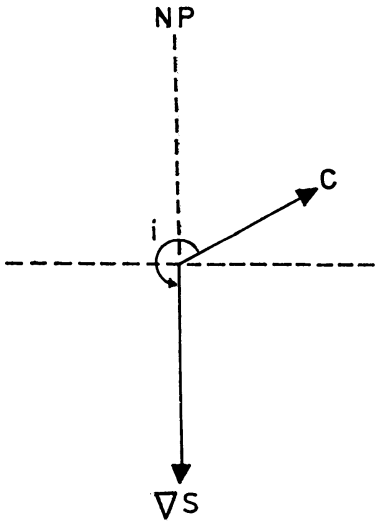

Fig. I, 4 See text.

and for the cylindrical Mercator projection ($p = 0$)

$$\dot{\xi} = \frac{c}{a} \sin i \tan \varphi, \tag{17}$$

$$\dot{c} = 0.$$

For $i = 0, \pi$ the control equations generate the straight meridians in all three mappings.

For $i = \pi/2, -\pi/2$ and $\varphi = 0$ they generate the equator.

It is interesting to note that the control equation (11) may also be used to evaluate the rhumbline in the projection map, a track which intersects all meridians at a fixed angle.

We have then

$$\dot{\gamma} = 0$$

which yields

$$\cos \gamma = \text{const.} = k.$$

This entered into (12) and (11) affords the control equation

$$\dot{\xi} = \frac{Sck}{r}. \tag{18}$$

Substituting (18) for (7) gives:

Lambert chart: $$\dot{\xi} = \frac{pck}{a \cos \varphi}.$$

Polar stereographic chart: $$\dot{\xi} = \frac{ck}{a \cos \varphi}.$$

Mercator chart: $$\dot{\xi} = 0.$$

The last relation states the well-known fact that in a Mercator chart all straight lines are rhumblines.

REFERENCES

[1] ZERMELO, E., 1931: Über das Navigationsproblem bei ruhender oder veränderlicher Windverteilung. Zts. angew. Math. & Mech., 11, pp. 114-124.

[2] LEVI-CIVITÀ, T., 1931: Über Zermelo's Luftfahrtproblem. Zts. angew. Math. & Mech., 11, pp. 314-322.

[3] VON MISES, R., 1931: Zum Navigationsproblem der Luftfahrt. Zts. angew. Math. & Mech., 11, pp. 373-381.

[4] CARATHEODORY, C, 1935: Variationsrechnung und partielle Differentialgleichungen. Tl. 2, pp. 234-242.

[5] BESSEMOULIN, I., PONE, R., 1949: Détermination des routes aériennes à durée minimum. La Météorologie, pp 101-121

[6] GRINGORTEN, I. I., 1948: The theory and computation of single heading flights. Bull. Amer. Meteor. Soc., 29, pp. 343-351.

[7] BIK, F. C., DE JONG, H. M., 1953: A report on the theory and application of the minimum flight path. Royal Dutch Airlines K.L.M. - Holland.

[8] DE JONG, H. M., 1956: Theoretical aspects of aeronavigation and its application in aviation meteorology. Med. & Verh. Kon. Ned. Meteor. Inst., publ. 102 - no. 64.

[9] PONTRYAGIN, L. S., BOLTYANSKI, V. G., GAMKRELIDZE, R. V., MISCHENKO, E. F., 1962: The mathematical theory of optimal processes. New York. (Engl. transl.)

[10] DESOER, C. A., 1961: Pontryagin's maximum principle and the principle of optimality, Journ. Franklin Inst., 271, no. 5, pp. 413-426.

[11] BERGE, C., 1970: Graphes et hypergraphes. Paris.

[12] BERGE, C., GHOUILA-HOURI, A., 1962: Programmes, jeux et réseaux de transport. Paris.

[13] HARRIS, B. (Ed.), 1970: Graph theory and its applications. New York.

[14] NAVEZ, J. E., 1966: Transatlantic electronic flight planning. SABENA.

[15] SIMPSON, S., BASHIOUM, D. L., and CARR, E., 1964: Computer flight planning in the North Atlantic. Aero Performance, Inc., Manhasset, New York.

[16] HALKIN, H., 1963: The principle of optimal evolution. In: 'Non-linear differential equations and non-linear mechanics', ed. by J. P. Lasalle and S. Lefschetz. New York.

[17] LEITMANN, G. (Ed.), 1967: Topics in optimization. New York.

[18] BELLMAN, R., GLICKSBERG, I., GROSS, O., 1956: On the 'Bang-Bang' control problems. Quart. journ. appl. math., 14, pp. 11-18.

[19] DANTZIG, G. B., 1960: On the shortest route through a network. Management science, 6.

[20] DIJKSTRA, E. W., 1959: A note on two problems in connection with graphs. Numer. math., 1, pp. 269-271.

[21] POLLACK, M., WIEBENSON, W., 1960: Solution of the shortest route problem. A review. Operation Research, 8.

[22] FARBRY, B. A., LAND, A. H., MURCHLAND, J. D., 1965: The cascade algorithm for finding a minimum distance. Rep. London School of Economics, LSE-TNT-19.

[23] SAUCIER, W. J., 1955: Principles of meteorological analysis. Chicago.

[24] ABROMOWITZ, M., STEGUM, J. A., Ed., 1964: Handbook of mathematical functions. Appl. math. series National Bureau of Standards, No. 55, pp. 877-899.

[25] 1964: Manual of the ICAO standard atmosphere, ICAO, Doc. 7488/2.

[26] RÄNICKE, G., 1965: Flugwege geringsten Zeitbedarfs. Berechnung mittels eines Brechnungsgesetzes. Luftfahrttechnik-Raumfahrttechnik, 11, pp. 114-118.

SUMMARY

By mere coincidence, some twenty years ago, aviation industry was enriched with some revolutionary novelties: the advent of regular turbo-jet operations, the advent of the high speed electronic computers and the development of the theory of optimal control processes in mathematics.

It was soon realized that this scientific and technological development would have a tremendous impact in the problem area of air navigation.

It was expected that, for example, these innovations would get a firm grip on existing procedures and practices of long distance flight planning. Here the new aids would be very useful in the elaboration of the problem of optimal track selection.

In the present study this problem is reconsidered in a unified approach taking into account the most general space variable and time changing meteorological conditions and performance characteristics.

In case of the free utilization of air space the problem was solved using the principles of optimal control processes. The criterium of time optimality, specific for flights along the Least Time Track in two dimensions, has led to the derivation of generalized forms of the well-known control equations like the steering equation of Zermelo, the refraction formula of Von Mises and the author's partial differential equation for the air navigation problem.

Owing to the unified approach of the problem solving it was possible to introduce some alternatives of the governing system equations in the form of a gradient equation for the time of transfer and a phase velocity equation for the effective air speed.

These mathematical tools have also formed the basis for the design of some 'constructive' algorithms suitable for the practical evaluation of the two-point boundary solution of the Least Time Track by means of iteration processes. The applicability of these new aids was proven by experimentation.

Before the computer experiments could be accomplished it was however required to modify the control equation as being valid in a conformal map projection.

In case of restricted airspace utilization the optimization problem could be tackled by using the principles of graph theory. This was exploited in two and three dimensions. The analysis of this problem in a discrete system in a space-time environment of meteorological conditions has led to the synthesis of special graph algorithms, which are suited to be programmed for 3-dimensional flight planning purposes. The feasibility of these graph solutions was also demonstrated experimentally.