Conjugate Gradient Methods with an Application to V/STOL Flight-Path Optimization

R. K. MEHRA* AND A. E. BRYSON JR.†
Harvard University, Cambridge, Mass.

Conjugate gradient methods have recently been applied to some simple optimization problems and have been shown to converge faster than the methods of steepest descent. The present paper considers application of these methods to more complicated problems involving terminal constraints. As an example, minimum time paths for the climb phase of a V/STOL aircraft have been obtained using the conjugate gradient algorithm. In conclusion, some remarks are made about the relative efficiency of the different optimization schemes presently available for the solution of optimal control problems.

I. Introduction

HESTENES and Stiefel[1] in 1952 introduced the method of conjugate gradients for solving linear sets of equations. The same method was used by Fletcher and Reeves[2] in 1964 to solve nonlinear programming problems. Hayes[3] extended the method in 1954 to the solution of linear problems on Hilbert spaces. Antosiewicz and Rheinboldt[4] derived, in 1962, convergence rates for these problems, and showed that convergence is obtained at a geometrically fast rate for the linear-quadratic problem. Improved estimates of rates of convergence were obtained by Daniel[5] in 1965. Lasdon, Mitter, and Warren[6] applied this method, in 1966, to the solution of optimal control problems. They showed that the conjugate gradient method converged faster than the steepest-descent method on a number of problems. Sinnot and Luenberger[7] recently used another variant of the conjugate gradient method and gave similar results. In addition, they extended the method to handle linear terminal constraints. However, most of the optimal control problems solved so far[6,7] using conjugate gradient methods have been simple in structure, involving either no or very few terminal constraints.

II. Conjugate Gradient Methods

a. Parameter Optimization

Conjugate gradient methods have the property that they minimize a quadratic function of n variables in n steps. They do so by generating a set of n directions known as conjugate directions, which span the n-dimensional space. Let the function to be minimized be $J = \frac{1}{2}(x-h)^T A(x-h)$ and let $p_0, p_1, \ldots, p_{n-1}$ be n vectors in Euclidean n space. They will be called "A-orthogonal" or "A-conjugate," if and only if

$$p_i^T A p_j = 0, \qquad i \neq j \tag{1}$$

where A is a positive definite matrix.

Therefore,

$$p_i^T A p_i > 0, \qquad \text{if } p_i \neq 0 \tag{2}$$

It is easy to show that n "A-conjugate" vectors are linearly independent and form a basis for the n-dimensional space. If x_0 is the initial guess, then $(h - x_0)$ can be expressed in terms of this basis as follows:

$$h - x_0 = \sum_{i=0}^{n-1} \alpha_i p_i \tag{3}$$

where

$$\alpha_i = -p_i^T A(x_0 - h)/p_i^T A p_i = -p_i^T g_0/p_i^T A p_i \tag{4}$$

where $g_0 = A(x_0 - h)$ is the gradient vector $\partial J/\partial x|_{x=x_0}$.

All conjugate gradient algorithms generate conjugate directions in some manner. They can be generated, for example, by a Gram-Schmidt orthogonalization procedure starting from any arbitrary set $v_0, v_1, \ldots, v_{n-1}$ of vectors. It can be shown that if v_i are the coordinate vectors, then the conjugate gradient method is functionally equivalent to the gaussian elimination procedure. A convenient choice for v_i is the negative gradient vectors or the residue vectors r_i;

$$r_i = -g_i = A(h - x_i) \tag{5}$$

This choice leads to a number of simplifications and, finally, the following algorithm is obtained. Details of the proof can be found in Beckman[14];

$$x_0 \text{ arbitrary}, \quad g_0 = g(x_0), \quad p_0 = -g_0,$$

$$x_{i+1} = x_i + \alpha_i p_i \qquad \text{where } \alpha_i = -p_i^T g_i/p_i^T A p_i \tag{6}$$

$$p_{i+1} = -g_{i+1} + \beta_i p_i \qquad \text{where } \beta_i - \|g_{i+1}\|^2/\|g_i\|^2 \tag{7}$$

This algorithm can be used for nonlinear programming problems as well. However, the matrix A is no longer a constant matrix, and must be computed at each step. One can avoid this by noting that, if J is minimized along the direction $(x_i + c_i p_i)$ with respect to c_i, the optimum value of c_i is exactly α_i.[14] Notice that, if $\beta_i = 0$, the conjugate gradient method becomes a steepest-descent method.

The conjugate gradient algorithm has a number of interesting properties. Rutishauser[8] compares it with other gradient methods and shows that it is the best method in a class of iterative gradient procedures for solving linear sets of equations. If ϵ_i denotes the error vector $(h - x_i)$, it can be shown that $\|\epsilon_{i+1}\| < \|\epsilon_i\|$ for all i. Also, it can be shown that J is decreased at each step. Geometrically, p_i is the projection of the negative gradient vector g_i on to the subspace spanned by $p_i, p_{i+1}, \ldots, p_{n-1}$. Thus we successively reduce the dimension of the subspace onto which $-g_i$ is projected.

Received June 10, 1968; revision received September 6, 1968. This work was supported by NONR 1866(16) and NASA Grant NGR 22-007-068 extended to Harvard University, Cambridge, Mass. The authors gratefully acknowledge the help and suggestions received from R. H. Miller and his students at the Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass.

* Research Assistant, Division of Engineering and Applied Physics; presently Research Engineer, The Analytic Sciences Corporation, Reading, Mass.

† Professor of Mechanical Engineering, Division of Engineering and Applied Physics; presently Professor of Applied Mechanics, Stanford University, Stanford, Calif. Associate Fellow AIAA.

b. Optimal Control Problems

Conjugate gradient methods can be readily extended to Hilbert spaces.[3,5] Consider the Mayer problem in the Calculus of Variations:
Find $u(t)$ to minimize

$$J = \phi[x(t_f)], \quad \text{subject to } \dot{x} = f(x, u, t) \tag{8}$$

$x(t_0)$ and t_f are given, but $x(t_f)$ is free. x is an $n \times 1$ state vector and u is an $r \times 1$ control vector, both functions of time variable t.

The Hamiltonian of the system is

$$H = \lambda^T f \tag{9}$$

and the adjoint equations are

$$\dot{\lambda} = -f_x{}^T \lambda \tag{10}$$

$$\lambda(t_f) = \phi_x[x(t_f), t_f] \tag{11}$$

Let

$$g(t) = \partial H/\partial u = \lambda^T(\partial f/\partial u) \tag{12}$$

g is a vector of functions and relates δJ to δu (Ref. 10., Chap. II)

$$\delta J = \int_{t_0}^{t_f} g \delta u \, dt \tag{13}$$

g plays the role of gradient vector in the finite dimensional case.

The same algorithm [Eqs. (6) and (7)] applies except that the scalar multiplications are changed to integrations, e.g.,

$$||g_i||^2 = \int_{t_0}^{t_f} g_i{}^T g_i \, dt$$

It is necessary to store a direction of search to calculate the next direction of search. A cubic interpolation scheme[2] is used for one-dimensional search. It uses all the information available, i.e., $J(u_i)$, $J(u_{i+1})$, $\partial J(u_i)/\partial \alpha_i$, $\partial J(u_{i+1})/\partial \alpha_i$ to fit the "smoothest curve" through the points u_i and u_{i+1}, i.e., the curve which minimizes the integral

$$\int_0^{\alpha_i} \frac{d^2 J}{d\sigma^2} d\sigma$$

where α_i is the step size.

III. Terminal Constraints

The conjugate gradient algorithm as given previously applies only to unconstrained minimization problems. Modifications to the algorithm are necessary when there are constraints on the problem. A fairly general optimization problem with terminal constraints can be stated as follows: Find $u(t)$ to minimize

$$J = \phi[x(t_f), t_f] \tag{14}$$

subject to

$$x = f(x, u, t); \qquad x(t_0) \text{ given} \tag{15}$$

and

$$\psi[x(t_f), t_f] = 0 \qquad q \text{ terminal constraints} \tag{16}$$

$$\Omega[x(t_f), t_f] = 0 \qquad \text{stopping condition for determining } t_f \tag{17}$$

In effect, there are $(q + 1)$ terminal constraints. Any one of these can be chosen as a stopping condition. This is an unnecessary, arbitrary, but useful device.

Two of the numerical methods for solving such problems are given below.

a. Penalty Function Method

The objective function J is modified using a quadratic penalty function

$$\bar{J} = J + \psi^T K \psi \tag{18}$$

where K is a positive definite matrix of penalty function constants. A sequence of unconstrained $\bar{J}$ problems is solved with increasing values of K. In the limit as $K \to \infty$ we get $\psi \to 0$, $J \to J_{\text{opt}}$, $u \to u_{\text{opt}}$. To check the efficiency of this method, it was used to solve a number of problems. The method worked quite well on linear-quadratic problems and simple nonlinear problems. Examples 1 and 2 of Ref. 6 were solved in one computer run by using a large enough value of K. The minimum-time Earth-to-Mars orbit transfer problem of Ref. 9 converged in 18 iterations starting from a stepped nominal and using about 1 min of IBM 7094 computer time.

However, when this method was tried on flight-path optimization problems involving aerodynamic drag and lift terms, the method ran into difficulties whenever the number of terminal constraints was increased beyond two. For a typical problem involving three state variables, V (velocity), h (altitude), and γ (flight-path angle), the convergence was extremely slow if terminal constraints were put on all the three state variables simultaneously. Since, for most of these problems, the terminal time is not specified, some sort of stopping condition is needed to determine t_f at each iteration. In this way, one of the constraints is automatically satisfied. It was found that the penalty function method could be used to handle, at most, two terminal constraints. If there were more terminal constraints, the convergence was extremely slow.

Various other types of penalty functions can be used. However, there is one common difficulty, viz., addition of penalty functions may change the problem completely creating narrow valleys in the state space. An example of this type is given in Ref. 10, Chap. 1. It is well known that gradient procedures converge very slowly under such conditions. Ho[15] discusses the difficulties that may arise in such cases if the steepest-descent method is used. Use of the conjugate gradient method seems to remedy these difficulties only partially. Even though it works well on linear-quadratic problems, and simple nonlinear problems, it works poorly on complicated nonlinear problems of the type indicated previously.

b. Gradient Projection Method

Gradient projection methods have been used in parameter optimization[11] and in optimal control problems[10,16] using a steepest-descent method. The same method can be used with the conjugate gradient method to handle linear terminal constraints.[7] If the step size a_i is small enough so that linearization is valid, the same method should work for nonlinear constraints as well.‡

An expression for the projected gradient $\bar{g}$ is given in Ref. 10;

$$\bar{g} = f_u{}^T(\lambda_\phi - \lambda_\psi I_\psi I_{\psi\psi}{}^{-1} I_{\psi\phi}) \tag{19}$$

where

$$\dot{\lambda}_\phi = -f_x{}^T \lambda_\phi; \quad \lambda_\phi(t_f) = \phi_x{}^T[x(t_f), t_f] \tag{20}$$

$$\dot{\lambda}_\psi = -f_x{}^T \lambda_\psi; \quad \lambda_\psi(t_f) = \psi_x{}^T[x(t_f), t_f] \tag{21}$$

$$I_{\psi\psi} = \int_{t_0}^{t_f} \lambda_\psi{}^T f_u f_u{}^T \lambda_\psi \, dt \tag{22}$$

$$I_{\psi\phi} = \int_{t_0}^{t_f} \lambda_\psi{}^T f_u f_u{}^T \lambda_\phi \, dt \tag{23}$$

Directions of search $\bar{p}_i$ may be generated using $\bar{g}_i$, $\bar{p}_{i-1}$, and Eqs. (6) and (7).§ If a change $d\psi$ is desired in the con-

‡ The authors do not have computational experience with this method.

§ Directions $\bar{p}_i$ will not be conjugate in general except for the linear quadratic problem with linear terminal constraints (see Ref. 17).

straint level ψ, the control change $\delta\psi$ is given by

$$\delta u = f_u{}^T \lambda_\psi I_{\psi\psi}{}^{-1}\, d\psi \tag{24}$$

The conjugate gradient algorithm is modified as follows:

$$\bar{u}_{i+1} = u_i + m_i \alpha_i \bar{p}_i \tag{25}$$

1) Start with $m_i = 1$ and obtain α_i by a one-dimensional search. 2) Calculate the value of $\bar{\psi}_{i+1}(t_f)$ using $\bar{u}_{i+1}$. If linearization holds, $\bar{\psi}_{i+1}(t_f)$ should be the same as $\psi_i(t_f)$. If not, reduce m_i so that $\|\bar{\psi}_{i+1} - \psi_i\| < \epsilon$, where ϵ is a small positive number. 3) Choose $d\psi_i$ and calculate the corresponding δu_i. Add this to $\bar{u}_{i+1}$: $u_{i+1} = \bar{u}_{i+1} + \delta u_i$. Note that $d\psi_i$ should not be so large that the linearity assumption is violated.

If this algorithm is used on a linear-quadratic problem with linear terminal constraints, the directions of search $\bar{p}_i$, $i = 0, n - 1$ will be conjugate and convergence would be obtained at a geometrically fast rate.[5] For a nonlinear problem, however, the directions p_i, $i = 0, n - 1$ will not be conjugate in general, because of the dependence of λ_ψ on u and the addition of δu from Eq. (24) at each step. To bypass this difficulty partially, one may try to satisfy the terminal constraints first and then hold them constant using the gradient projection scheme. This method may work well if the constraints are linear, but if the constraints are highly nonlinear, m_i will have to be chosen small enough so that linearization holds. In such a case, it might be better to approach near the optimum using the penalty function method and then refine the solution using the gradient projection method. Typically, in most of the optimization problems, the step size δu_i gets smaller and smaller as one approaches the optimum; so the linearization assumption would not be violated and the gradient projection method would generate nearly conjugate directions near the minimum.

IV. Flight Path for Minimum-Time Climb of a V/STOL Aircraft

Compared to conventional aircraft, V/STOL aircraft have an extra control variable, namely the angle between the thrust direction and a reference axis in the aircraft. It is of interest to know how this extra control variable may be used to improve the performance of the aircraft.

If a flight is long enough, it can be divided into three paths: 1) climb phase starting from the ground and going up to some cruise condition, 2) cruise at some constant altitude and velocity, and 3) landing phase. Depending on the particular use to which the V/STOL aircraft is put, there may be flight-path constraints on paths 1 and 3. If the cruise conditions are known, the optimization problem reduces to optimization of the two arcs 1 and 3 separately, because the cruise conditions specify the state completely at the end of path 1 and at the beginning of path 3.

Here, we shall consider the hypothetical jet-lift aircraft of Ref. 12. Gallant[13] has considered a tilt-wing V/STOL aircraft and obtained minimum-direct-cost flight paths for a 50-mile flight starting from the end of the transition to the beginning of the retransition.

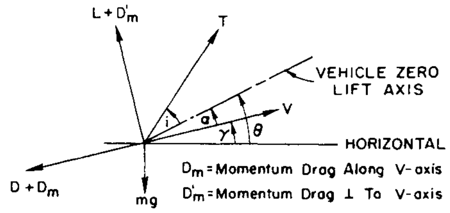

Fig. 1 Vector diagram of a V/STOL aircraft.

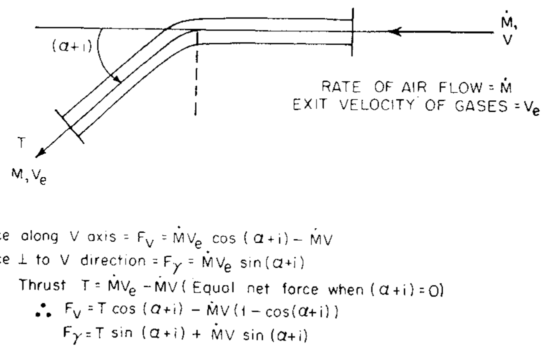

Fig. 2 Thrust diagram of tilt-jet.

Problem Formulation

The aircraft will be approximated as a mass point. Figure 1 shows the forces acting on the aircraft. Figure 2 shows the thrust force in greater detail. It is assumed that the jet inlets are always pointed in the direction of the relative wind velocity. This approximation is reasonable in view of the rather crude model assumed for the V/STOL aircraft and in view of the final results which show that the angle of attack is kept small during most of the flight.

$$\dot{V} = \frac{T}{m}\cos(\alpha + i) - \frac{D}{m} - g \sin\gamma - \frac{\dot{M}}{m} V[1 - \cos(\alpha + i)] \tag{26}$$

$$\dot{\gamma} = \frac{T}{mV}\sin(\alpha + i) + \frac{L}{mV} - \frac{g}{V}\cos\gamma + \frac{\dot{M}}{m}\sin(\alpha + i) \tag{27}$$

$$\dot{h} = V \sin\gamma \tag{28}$$

$$\dot{x} = V \cos\gamma \tag{29}$$

where

$$\text{Lift:} \quad L = \tfrac{1}{2}\rho V^2 C_L S \tag{30}$$

$$\text{Drag:} \quad D = \tfrac{1}{2}\rho V^2 C_D S \tag{31}$$

$$C_L = C_{L_\alpha}\alpha \tag{32}$$

$$C_D = C_{D_0} + C_{D_1}\alpha^2 \tag{33}$$

$$\text{Air Density:}\ \rho = 0.0023769(1 - 0.6875 \times 10^{-5}h)^{4.2561} \tag{34}$$

Equation (34) holds for $h \leqslant 36{,}000$ ft.

The characteristics of the hypothetical aircraft are

Thrust: $T = T_0(1 - 0.55h/30{,}000)$ where h is in ft

Mass: $m = 56{,}902/32.2$ slugs (taken as constant during climb)

Rate of Airflow: $\dot{M} = T_0/(65 \times 32.2)$ slugs/sec if T_0 is in lb

Wing Area: $S = 421$ ft^2

$$C_{L_\alpha} = 5.73, \quad C_{D_0} = 0.027, \quad C_{D_1} = 1.93$$

The maximum (L/D) ratio is 12.6 and occurs at $\alpha = 6.8°$.

There are three control variables in the problem: magnitude of thrust vector (T_0), $(0 \leqslant T_0 \leqslant T_{0_{max}})$; direction of thrust vector (i); angle of attack (α) or pitch angle (θ). It is preferable to use θ instead of α as the control variable. The use of θ as the control variable adds extra damping terms into the equations of motion which help in convergence of numerical computations of optimal flight paths.

We shall obtain minimum time paths under the following assumptions:

1) Thrust T_0 is kept constant at its maximum value. This is a reasonable assumption for the climb phase of the flight. In particular, we shall use $T_0 = 1.25$ mg.

2) Initial conditions for the problem are

$$V(0) = 0, \quad h(0) = 0, \quad x(0) = 0$$

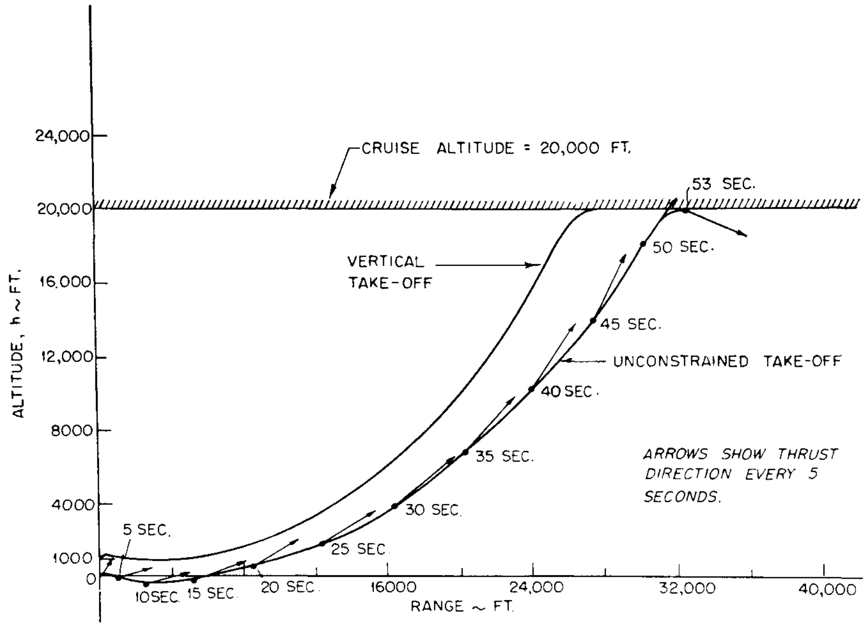

Fig. 3 Altitude vs range histories for 1) unconstrained takeoff and 2) vertical takeoff.

The $\dot{\gamma}$ equation has a singularity at $V = 0$. To integrate the equations of motion numerically, we must start with a finite V. The aircraft would attain this velocity after flying for some time, say t_1, in some particular manner. This part of the flight may be partially or completely determined by restrictions on the runway available for takeoff, e.g., if the aircraft must take off vertically, then $\gamma(t_1) = 90°$ where t_1 will be some time either during or at the end of the vertical takeoff period. We now consider a few specific cases.

Unconstrained Takeoff

At the beginning of our investigation, we did not know whether this V/STOL aircraft should take off like a conventional aircraft (by first picking up speed along the runway) or whether it should take off directly making some angle $\gamma(0+) > 0$ to the horizontal. Hence, we solved the unconstrained problem treating $\gamma(t_1)$ as a control parameter, allowing altitude to go negative if this was optimal.

The initial velocity for this case was taken as (treating t_1 as starting time denoted by 0) $V(0) = 50$ fps. Since $\gamma(0)$ was to be chosen optimally, the optimization process had to drive $\lambda_\gamma(0)$ to zero. Changing $h(0)$ from several hundred feet to zero did not change the results significantly, so $h(0) = 0$ was used. The final time t_f was minimized with the terminal conditions $\gamma(t_f) = 0$, $h(t_f) = 20{,}000$ ft, $V(t_f)$ free, $x(t_f)$ free.

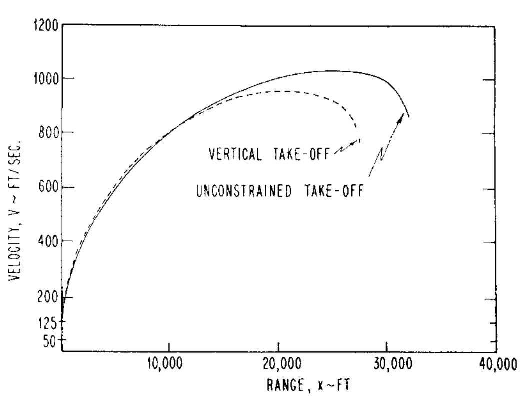

Fig. 4 Velocity vs range for 1) unconstrained takeoff and 2) vertical takeoff.

A constraint on $V(t_f)$ can be met fairly easily by changing the thrust magnitude and/or θ and i towards the end of the climb phase. The control variables used here were θ and i.

Figures 3, 4, 5a, and 6–8 show the results obtained for the case when there were no constraints on takeoff. The optimum

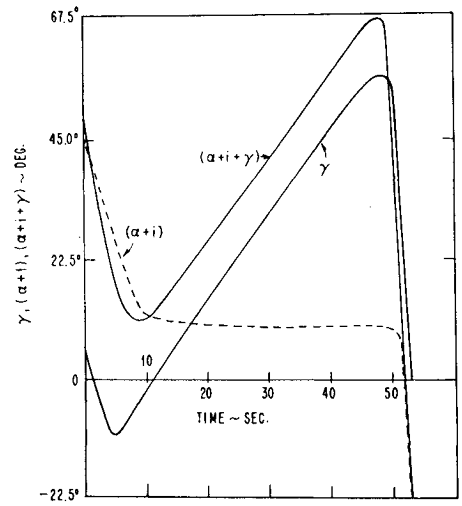

Fig. 5a) Flight-path angle γ, $(\alpha + i)$, and thrust direction $(\alpha + i + \gamma)$ vs time for unconstrained takeoff.

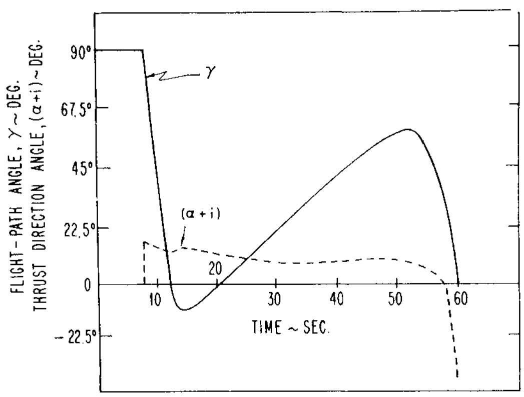

Fig. 5b) Flight-path angle and $(\alpha + i)$ vs time with vertical takeoff constraint.

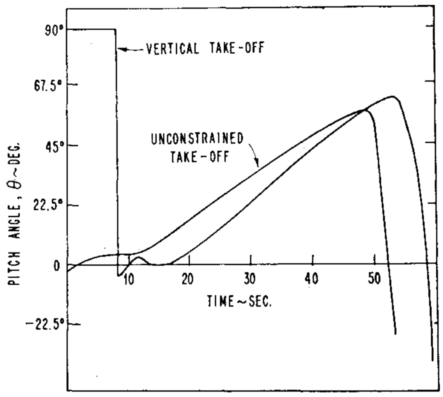

Fig. 6 Pitch angle θ vs time for 1) unconstrained takeoff and 2) vertical takeoff.

value of $\gamma(0)$ at $V = 50$ fps turns out to be about 7°. However, the interesting fact is that γ soon becomes negative and the aircraft goes about 300 ft underground. Reasons for this seem to be

1) the thrust is greater at lower altitudes;

2) the aircraft should pick up velocity as fast as possible in order to generate aerodynamic lift. Lift obtained from tilting the jet is not very efficient because it gives lots of momentum drag. The angle $(\alpha + i)$ is apparently kept low (about 11°) in order to keep this drag low (cf. Fig. 5a).

To keep the pressure drag low, angle of attack is also kept small (mostly about 2°), as shown in Fig. 8. The aircraft dives down because gravity helps it in picking up speed. Figure 4 shows velocity vs range. A maximum velocity of 1025 fps is attained towards the end of the climb. Figure 9 shows the time history of the total lift over drag ratio, viz., $(L + D'_m)/(D + D_m)$. This ratio is rapidly increased to a maximum of about 12 and is maintained in the vicinity of 6 for most of the flight.

After the aircraft has picked up velocity during the diving maneuver, γ increases quickly to a maximum value of 56.6° (Fig. 5a). Figure 3 shows how h increases during this phase. However, θ also increases at the same time, so that $\alpha = \theta - \gamma$ remains small. Jet-tilt angle i is also kept small. Thus, the total drag is kept low. The lift force reaches a maximum value of about 70,000 lb. This maneuver is followed by a rapid change in θ to a negative value of about −25°. This is necessary to meet the terminal condition on γ, viz., $\gamma(t_f) = 0$. The aircraft experiences a downward acceleration of 10 g. The total time taken by the aircraft is 53 sec. Calculations show that if the aircraft is made to climb vertically all the way up from the ground, it takes twice as much time. The velocity in that case never exceeds 300 fps.

Thus, the results show that a V/STOL aircraft without takeoff constraints should fly very much like a conventional aircraft. Aerodynamic lift is more efficient than jet-lift. On the other hand, the aircraft should keep angle of attack small to keep aerodynamic drag low.

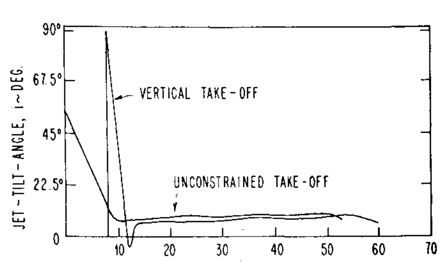

Fig. 7 Jet-tilt angle i vs time for 1) unconstrained takeoff and 2) vertical takeoff.

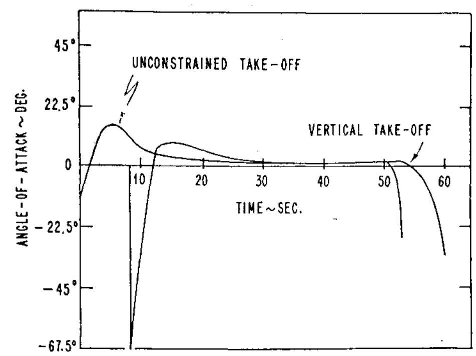

Fig. 8 Angle of attack vs time for 1) unconstrained takeoff and 2) vertical takeoff.

Horizontal Takeoff Constraint

Imposing the constraint that the aircraft cannot go underground, numerical results show that the optimal path is along $h = 0$ until the vehicle reaches $V \approx 300$ fps, when the path starts up. Thus, for the fastest climb-out this STOL aircraft should fly parallel to the ground at low altitude for a considerable distance!

Vertical Takeoff Constraint

Next we consider the restriction that the aircraft must fly vertically up to an altitude of 1000 ft. From the results obtained previously, it appears that the best way to do this would be to make $\theta = 90°$ so that $\alpha = 0$, $i = 0$, $\dot{\gamma} = 0$. Integration of the V equation gives $V = 125$ fps at $h = 1000$ ft. Time taken is 8 sec. The optimization problem is now solved with the following initial conditions: $V(8) = 125$ fps, $\gamma(8) = 90°$, $h(8) = 1000$ ft, $x(8) = 0$.

The results are shown in Figs. 3, 4, 5b, and 6–8. The total path (from takeoff) is 60 sec long and is similar to the unconstrained takeoff case. The aircraft goes up first due to positive γ, but soon dives down to a minimum altitude of about 980 ft. The control variables θ and i have discontinuities at $t = 8$ sec when the constraints are relaxed.

Similar behavior would be obtained if the aircraft were constrained to take off at some other constant value of flight-path angle γ_C. The equation $\dot{\gamma} = 0$ determines one of the control variables in terms of the other (say i in terms of θ). If θ is constrained by $\theta \leq \gamma_C$, as in the previous case, one would intuitively expect that θ would remain constant at γ_C.

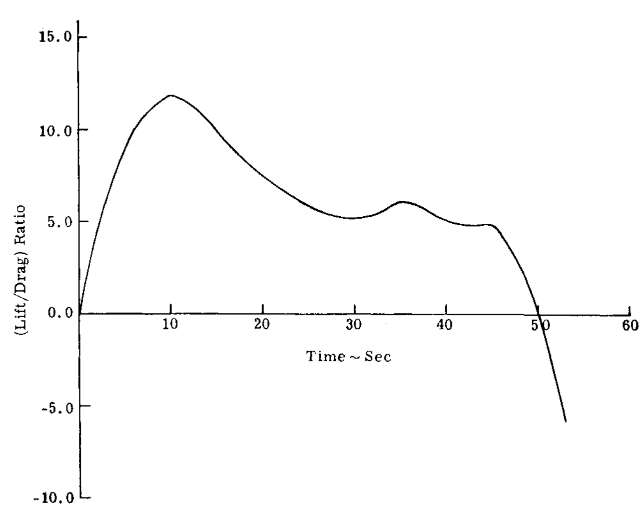

Fig. 9 Time history of (L/D) ratio for unconstrained takeoff.

Other Constraints

A number of other physical constraints might be imposed on the flight path. Passenger comfort dictates that the accelerations be kept less than about 1 g and the pitch angle be limited to about 30°. These constraints can readily be incorporated in the optimization problem. In most of the cases, the result of imposing a constraint is to put the constrained variable on the boundary for part of the flight if, in the unconstrained optimal path, it exceeds the constrained value. These constraints have been considered in Ref. 18.

V. Conclusion

Our computational experience has shown that the conjugate gradient method, though very efficient for simple optimization problems, may run into difficulties when applied to more complicated problems. Some of the difficulties that may be encountered are

1) Gradient of the objective function with respect to the step size may not become zero or small enough during one-dimensional search. Accumulation of errors due to this source can produce directions of search which increase rather than decrease the performance index.**

In such cases, it was found useful to revert back to the local gradient direction and start the process over again. This procedure is similar to the one suggested by Beckman[14] and also used by Fletcher and Reeves[2] for nonlinear programming problems.

2) Use of penalty functions may create narrow valleys in the state space and make convergence extremely slow. The use of the gradient projection method, though more complicated, may help in this case, particularly near the optimum.

Conjugate Gradient Methods vs Steepest-Descent Methods

1) For optimal control problems having either no or few constraints, conjugate gradient methods, though requiring more programming, seem to be faster than the steepest-descent method of Ref. 16.

2) For optimal control problems with a large number of constraints, it becomes necessary to limit the step size of conjugate gradient methods during the one-dimensional search, and the directions of search are then no longer conjugate to each other. In nonlinear problems with nonlinear constraints, our limited experience indicates that conjugate gradient methods are not as effective as steepest-descent methods when starting far from the optimum.

Conjugate Gradient Methods vs Second Variation Methods

1) Conjugate gradient methods require less programming and less computation per iteration than second variation methods.

2) Second variation methods require the matrix of second variations of the Hamiltonian with respect to the control (H_{uu}) to be nonsingular. Conjugate gradient methods do not require this.

3) Conjugate gradient methods do not converge to extremals containing conjugate points, whereas second variation methods try to converge towards these extremals.

4) Second variation methods lead to more accurate solutions than conjugate gradient methods, particularly to better control histories.

**Refer to Eq. (7). During the one-dimension search, $\partial J/\partial c_i(x_i + c_i p_i) = g_{i+1}{}^T p_i$ should be made zero. But if $g_{i+1}{}^T p_i > 0$, then during the next iteration, the quantity $g_{i+1}{}^T p_{i+1} = -\|g_{i+1}\|^2 + \beta_i g_{i+1}{}^T p_i$ can become positive if the second term is larger than the first.

References

[1] Hestenes, M. R. and Stiefel, E., "Method of Conjugate Gradient for Solving Linear Systems," *Journal of Research of the National Bureau of Standards*, Vol. 49, 1952, p. 409.

[2] Fletcher, R. and Reeves, C. M., "Function Minimization by Conjugate Gradients," *Computer Journal*, July 1964.

[3] Hayes, R. M., "Iterative Methods of Solving Linear Problems on Hilbert Spaces," *Applied Mathematics*, Series 39, 1954, National Bureau of Standards.

[4] Antosiewicz, H. A. and Rheinboldt, W. C., "Numerical Analysis and Functional Analysis," *Survey of Numerical Analysis*, edited by J. Todd, McGraw-Hill, New York, 1962.

[5] Daniel, J. W., "The Conjugate Gradient Method for Linear and Nonlinear Operator Equations," *Society of Industrial and Applied Mathematics Journal of Numerical Analysis*, Vol. 4, No. 1, 1967.

[6] Lasdon, L. S., Mitter, S., and Warren, A. D., "The Method of Conjugate Gradients for Optimal Control Problems," *Institute of Electrical and Electronics Engineers, Transactions on Automatic Control*, April 1967.

[7] Sinnott, J. F. and Luenberger, D. G., "Solution of Optimal Control Problems by the Method of Conjugate Gradients," Joint Automatic Control Conference, 1967, *Preprints*, pp. 566-575.

[8] Rutishauser, H., "Theory of Gradient Methods," Chap. II in Stiefel, E, "Uber Einige Methoden der Relaxationrechnung," *Zeitschrift Fuer Angewandte Mathematik und Physik*, 1952.

[9] Moyer, H. G. and Pinkham, G., "Several Trajectory Optimization Techniques, Part II: Application," *Computing Methods in Optimization Problems*, edited by A. V. Balakrishnan and L. W. Neustadt, Academic Press, New York, 1964.

[10] Bryson, A. E. and Ho, Y. C., *Applied Optimal Control*, Blaisdell, Waltham, Mass., 1969.

[11] Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming Part I: Linear Constraints," *Journal of the Society of Industrial and Applied Mathematics*, Vol. 8, No. 1, 1960, pp. 181–217.

[12] Miller, R. H. et al., "A Systems Analysis of Short Haul Air Transportation," Pt. III, TR-65-1, Aug. 1965, prepared for the U. S. Dept. of Commerce by the Flight Transportation Lab., Massachusetts Institute of Technology, Cambridge, Mass., p. II–30.

[13] Gallant, R. A., "Application of the Calculus of Variations in Determining Optimum Flight Profiles For Commercial Short Haul Aircraft," M.S. thesis, Feb. 1967, Massachusetts Institute of Technology, Cambridge, Mass.

[14] Beckman, F. S., "The Solution of Linear Equations by the Conjugate Gradient Method," *Mathematical Methods for Digital Computers*, Vol. I, edited by A. Ralston and H. S. Wilf, Wiley, New York, 1960.

[15] Ho, Y. C., "Computational Procedure for Optimal Control Problem with State Variable Constraint," *Journal of Mathematical Analysis and Applications*, Vol. 5, 1962, pp. 216–224.

[16] Bryson, A. E. and Denham, W. F., "A Steepest Ascent Method for Solving Optimum Programming Problems," *Journal of Applied Mechanics*, June 1962, pp. 247–257.

[17] Mehra, R. K., "Studies in Smoothing and in Conjugate Gradient Methods Applied to Optimal Control Problems," Ph.D. thesis, Dec. 1967, Div. of Engineering and Applied Physics, Harvard Univ., Cambridge, Mass.

[18] Leet, J. R., "Optimum Takeoff and Landing of a V/STOL Aircraft—Hybrid Computer Simulation," M.S. thesis, Sept. 1968, Massachusetts Institute of Technology, Cambridge, Mass.