A DEGENERATE PROBLEM OF BOLZA

FRANK FAULKNER

1. **Introduction.** The problem of Bolza [1, §69] with one total differential equation in three variables is discussed. One variable is required to be monotonic for a solution to exist; the extremals then generate a family of surfaces. This allows a simple geometric interpretation in the case of separated end conditions, and emphasis is on that case. Conditions corresponding to the conditions of transversality are developed.

2. **Statement of the problem.** The problem solved is that of finding a curve C^* with the following properties.

(a) A total differential equation

$$Pdx + Qdy + Rdz = 0 \tag{1}$$

is satisfied along C^* except at corners; the functions P, Q, and R are each analytic and R is bounded away from zero in the region of interest A.

(b) The variable x is monotonic nondecreasing along C^*; that is, $dx/ds \geqq 0$ along C^*.

(c) The beginning of C^*, the point (x_1, y_1, z_1), lies on the manifold defined by a system of equations

$$G_i(x, y, z) = 0; \tag{2}$$

there may be one, two, or three equations in this system.

(d) The end of C^*, (x_2, y_2, z_2), may be required to satisfy one or two equations of the form

$$H_i(x, y, z) = 0; \tag{3}$$

there may be no equation of this form.

Curves satisfying these conditions will be called *admissible curves*.

(e) A function $f(x, y, z)$ or $f(x_2, y_2, z_2)$ is required to have a minimum at (x_2, y_2, z_2), compared with its value at the end of other admissible curves.

An asterisk (*) will be used to denote values and conditions associated with C^* and the minimum.

Two features distinguish this from the usual problem of Bolza with separated end-conditions: the linearity of the differential equa-

Presented to the Society, December 28, 1952; received by the editors August 2, 1954 and, in revised form, January 4, 1955.

tion, and the inequality (b); also, the function to be minimized has the form $f(x_2, y_2, z_2)$ instead of $f(x_2, y_2, z_2) - f(x_1, y_1, z_1)$ [1, p. 191].

Under the conditions given, the differential equation can always be reduced to the form

$$dz = K(x, y, z)dx, \tag{4}$$

by quadrature and renaming the variables. In the remainder of the paper, it will be assumed that equation (1) is in this canonical form.

Assume that K has the following properties.

(f) The equation

$$K_y = 0, \tag{5}$$

where the subscript denotes the partial derivative, defines y as a function

$$y = y^*(x, z) \tag{6}$$

for all values of x, z in A.

(g) The second partial derivative is positive,

$$K_{yy} > 0 \tag{7}$$

in A. These two conditions ensure that $K(x, y, z)$ has a minimum as a function of y for $y = y^*(x, z)$.

(h) Finally,

$$K_z \geqq 0 \tag{8}$$

in A.

3. **Derivation of conditions for a minimum.** The conditions for a minimum for f may be derived in the usual way. Assume that a solution has been found in the form $y = y^*(x)$, $z = z^*(x)$; $y^*(x)$ is found to be equal to $y^*(x, z)$ on C^*. If $y^*(x)$ is determined, z^* is given by equation (4), which may be rewritten as

$$z^* = z_1 + \int_{x_1}^{x} K(t, y^*, z^*)dt; \tag{9}$$

$y^*(x)$ generally has a discontinuity at x_1 and at x_2, but z^* is continuous.

Before considering variations of f, let us investigate the variations of z. This is done in two steps, finding the variations due to variations of $y(x)$ for $x > x_1$, and then augmenting those with the variations due to variations in the initial values (x_1, y_1, z_1).

Consider a new function $Y(x, a) = y^* + av$, where a is a parameter

and v is an arbitrary piecewise-continuous function of x. This generates a new function

$$Z(x, a) = z_1 + \int_{x_1}^{x} K(t, Y, Z)dt, \tag{10}$$

which may be expanded in powers of a

$$\begin{aligned} z^* + aw + \frac{a^2 w_2(x)}{2} + \cdots \\ = z_1 + \int_{x_1}^{x} \Big[K^* + a(K_y^* v + K_z^* w) + \frac{a^2}{2}(K_{yy}^* v^2 \\ + 2K_{yz}^* vw + K_{zz}^* w^2 + K_z^* w_2) + \cdots \Big] dt; \end{aligned} \tag{11}$$

the asterisk (*) on K denotes that the arguments are those associated with C^*. By equating the coefficients of a on the two sides of this equation, we get

$$\begin{aligned} w &= \int_{x_1}^{x} (K_y^* v + K_z^* w)dt, \\ w &= \int_{x_1}^{x} K_y^* v dt + \int_{x_1}^{x} K_z^* \int_{x_1}^{t} K_y v dt_1 dt \\ &\quad + \int_{x_1}^{x} K_z^* \int_{x_1}^{t} K_z^* \int_{x_1}^{t_1} K_y^* v dt_2 dt_1 dt + \cdots, \end{aligned} \tag{12}$$

or

$$w(x) = I\left(\int_{x_1}^{x} K_y^* v dt \right). \tag{13}$$

This defines the linear homogeneous operator I

$$I[F(x)] = F(x) + \int_{x_1}^{x} K_z^* F(t)dt + \int_{x_1}^{x} K_z^* \int_{x_1}^{t} K_z^* F(t_1)dt_1 dt + \cdots.$$

Let $I_2(F) = I(F)_{x=x_2}$ and $I = I(1)$; these will be used later.

It is seen that if $w(x)$ is to be zero for an arbitrary choice of $v(x)$ in equation (12), then condition (5), $K_y = 0$, must hold on C^*.

For if we choose $v = K_y^*$, then the first integral is positive whenever x is larger than the value where K_y^* is first different from zero; the other integrals are either positive or zero, due to condition (8). Hence w becomes positive for all x sufficiently large. Likewise, if $v = -K_y^*$,

w becomes negative. Hence $y=y^*(x, z)$ is a necessary condition for the variation of z to vanish.

Then there is the further variation of z due to the variation of the endpoints. Assume that $x_1(a)$, $y_1(a)$, $z_1(a)$; $x_2(a)$, $y_2(a)$, $z_2(a)$ are functions with continuous second derivatives, satisfying conditions (2) and (3). If we differentiate equation (10), we see that, for $a=0$,

$$\frac{dz_2}{da} = \frac{dx_2}{da}K(x_2, y^*[x_2, z_2], z_2) + I_2\left\{\int_{x_1}^{x} K_y^* v dt + \frac{dz_1}{da} - \frac{dx_1}{da}K(x_1, y^*[x_1, z_1], z_1)\right\}.$$

Finally, for f to be a minimum,

$$0 = \frac{df}{da} = f_{x_2}\frac{dx_2}{da} + f_{y_2}\frac{dy_2}{da} + f_{z_2}\frac{dz_2}{da};$$

this expands to

$$0 = \{f_{x_2} + f_{z_2}K(x_2, y^*[x_2, z_2], z_2)\}\frac{dx_2}{da} + f_{y_2}\frac{dy_2}{da} + f_{z_2}I_2\left\{\int_{x_1}^{x} K_y^* v dt + \frac{dz_1}{da} - \frac{dx_1}{da}K(x_1, y^*[x_1, z_1], z_1)\right\}. \tag{14}$$

Except for the singular case where $f_{z_2}=0$, the solution to the problem is given by the system

$$dz_1 - K(x_1, y^*[x_1 z_1], z_1)dx_1 = 0, \quad G_i(x_1, y_1, z_1) = 0; \tag{15}$$

$$dz = K(x, y, z)dx, \quad y = y^*(x, z) \qquad (\text{for } x_1 < x < x_2); \tag{16}$$

$$\{f_{x_2} + f_{z_2}K(x_2, y^*[x_2, z_2], z_2)\}dx_2 + f_{y_2}dy_2 = 0, \quad H_i(x_2, y_2, z_2) = 0. \tag{17}$$

These are necessary conditions for an extreme value for f, excluding singular cases. The curve C^* furnishing this minimum consists of three segments. The first is the straight line parallel to the y axis, running from the point (x_1, y_1, z_1) to $(x_1, y^*[x_1, z_1], z_1)$. The second runs from $(x_1, y^*[x_1, z_1], z_1)$ to $(x_2, y^*[x_2, z_2], z_2)$; the differential equa-

tion (4) and condition (5) are satisfied along it. The third segment is the straight line parallel to the y axis from this last point to (x_2, y_2, z_2).

4. Geometrical interpretation. The geometric significance of these equations may be seen with the aid of a lemma which follows.

If y is eliminated between equations (4) and (6), the resulting differential equation

$$dz = K(x, y^*[x, z], z)dx = L(x, z)dx \tag{18}$$

is the equation of a family of cylinders $\{S\}$ of the form $\phi(x, z) = \text{const.}$ with generators parallel to the y axis. If extremals for the given problem are defined as the curves on which equation (6) is satisfied except where $dx/ds = 0$, this family of cylinders is generated by the extremals of equation (4).

These cylinders have the following minimizing property. Let (x_1, y_1, z_1) be any point of A and S_1 the corresponding cylinder defined by $\phi(x, z) = \phi(x_1, z_1)$.

LEMMA. *Any point (x_2, y_2, z_2) of S_1, with $x_2 > x_1$, may be attained by an admissible curve from (x_1, y_1, z_1) and the value z_2 is smaller than the value associated with x_2 on any admissible curve which is not an extremal.*

PROOF. The point (x_2, y_2, z_2) is attained by an extremal C^* as described in the next to last paragraph of §3.

Let $y = y^*(x)$ and $z = z^*(x)$ denote the functional values on C^*; let C be any other curve, defined by $Y(x)$ and $Z(x)$; Y may be discontinuous. Let x_3 be the point where $Y(x)$ is first different from $y^*(x)$. Then for $x > x_3$

$$Z(x) - z(x) = \int_{x_3}^{x} [K(x, Y, Z) - K(x, y^*, z^*)]dx.$$

Now

$$K(x, Y, Z) \geqq K(x, Y, z^*)$$

(since $K_z \geqq 0$) unless $Z < z^*$, and

$$K(x, Y, z^*) > K(x, y^*, z^*)$$

in some neighborhood of x_3, with $x > x_3$. It follows by contrapositive argument that the integrand is positive or zero for all $x > x_3$ and is positive in some neighborhood of x_3. Hence $Z(x) > z^*(x)$ for all $x > x_3$.

Hence the Euler equation leads to a minimum value of z on an admissible curve.

The results of §3 may be interpreted now. The Euler equation and the differential equation define directrices for the family of cylinders $\{S\}$. If there are three equations $G_i(x_1, y_1, z_1)=0$, these define a point; if there is one or two of these, equations (14) are the condition that the manifold be tangent to the corresponding cylinder of the family $\{S\}$ at (x_1, y_1, z_1). In either case, equations (15) and (16) define a minimum value of $z(x)$ for admissible curves, for $x>x_1$. The surface S_1 furnishes a boundary to the region of points which may be attained by an admissible curve from the manifold $G_i(x_1, y_1, z_1)=0$.

If there are two equations $H_i(x_2, y_2, z_2)=0$, then the adjunction of the equation

$$\phi(x_2, z_2) = \phi(x_1, z_1) \tag{19}$$

determines the point (x_2, y_2, z_2); f has an extreme value by virtue of the fact that z has an extreme value. If there is but one equation $H_i=0$, or no equation of this form, then this together with (18) defines a manifold; equations (17) define the condition that this manifold is tangent to the surface $f(x, y, z)=\text{const.}$ at (x_2, y_2, z_2).

If we think of z as positive upward, for an extreme value for f equations (2) must either define a point or be tangent to the family $\{S\}$ from above so that a lowest cylinder S_1 is determined. The corresponding member of the family $f=\text{const.}$ must be tangent to the manifold defined by S_1 and equations (3) if f is to have an extreme value.

The problem is thus reduced to the solution of the differential equation (18) and two minimization problems involving accessory conditions, problems of a type commonly studied in advanced calculus.

5. Remarks, singular cases, mixed end conditions. If $\nabla f=0$ at some point inside the region which may be attained by admissible curves, then f may have a minimum at this point, and this point may be attained by a set of admissible curves, not extremals, which end at this point. Another singular case occurs if $x_1=x_2$; there are various others.

Conditions (15) and (17) correspond to the conditions of transversality in the nondegenerate problem since they arise from the variation of the end points. Equation (5) is the degenerate Euler equation corresponding to the linear differential equation (4).

In the case of mixed end conditions, a function $F(x_1, y_1, z_1; x_2, y_2, z_2)$ is to be minimized, subject to conditions (a) and (b) and a set of end conditions

$$G_i(x_1, y_1, z_1; x_2, y_2, z_2) = 0; \tag{20}$$

there may be any number from none to five in the set of end conditions. Equations (15) and (17) are replaced by (20) and

$$\begin{aligned}[F_{x_1} - F_{z_2}L(x_1, z_1)I_2]dx_1 + F_{y_1}dy_1 + [F_{z_1} + F_{z_2}I_2]dz_1 \\ + [F_{x_2} + F_{z_2}L(x_2, z_2)]dx_2 + F_{y_2}dy_2 = 0.\end{aligned} \tag{21}$$

One case which allows an elementary interpretation is the one where F has the form

$$F = f(x_2, y_2, z_2) - g(x_1, y_1, z_1) \tag{22}$$

and there are no end conditions. If ϕ is an integral of equation (18) with continuous derivatives, the condition (21) may be expressed as

$$(\nabla g)_1 \parallel (\nabla \phi)_1,$$

$$(\nabla f)_2 \parallel (\nabla \phi)_2,$$

$$\frac{|\nabla g|_1}{|\nabla \phi|_1} = \frac{|\nabla f|_2}{|\nabla \phi|_2},$$

where the subscripts 1 and 2 denote the points (x_1, y_1, z_1) and (x_2, y_2, z_2) respectively.

If the restriction that x be nondecreasing is removed, then in general there is no minimum or maximum for f. A familiar example is the case of a force field which is not derivable from a potential. Suppose that the force vector has components P, Q which are functions of x, y. The work W along a specified path is given by the differential equation

$$dW = Pdx + Qdy.$$

If two points (x_1, y_1) and (x_2, y_2) are selected and we try to choose a path to minimize the work done in going from one to the other, it is well known that no such minimum exists (for example, see [2; pp. 7, 8]); paths can be selected which give any desired finite value for the work done. Except for the restriction that one variable be monotonic this is a problem of the type investigated in this paper, with $z = W$, $R \equiv -1$, with three equations (2) (x_1 and y_1 are given, and $z_1 = 0$), with two equations (3) (x_2 and y_2 are specified, and with $f(x_2, y_2, z_2) \equiv z_2$).

If the differential equation (1) is not reduced to the canonical form (4), the Euler equation (5) has the form

$$P(R_y - Q_z) + Q(P_z - R_x) + R(Q_x - P_y) = 0.$$

If the expression on the left in this equation vanishes identically, equation (1) is integrable and equation (5) is satisfied identically after the reduction to canonical form. In this case the problem is conceptually simpler, but the work of solution is similar. The surfaces $\{S\}$ of §4 are replaced by the integral surfaces of equation (1); then the extreme values of f are determined by the contacts of these with the surfaces defined by $G_i=0$, $H_i=0$, and $f=\text{const.}$, almost exactly as in §4. There is not generally a single curve connecting (x_1, y_1, z_1) and (x_2, y_2, z_2) but any curve serves which lies in the corresponding integral surface of equation (1) and connects the two points. The restriction that x be monotonic has no particular significance for this case. In the work problem used as an example above, this is the case where the force function is derived from a potential and the work done is the same along all paths connecting the two selected points.

Bibliography

1. G. A. Bliss, *Lectures on the calculus of variations*, Chicago, 1946.

2. B. Baule, *Die Mathematik des Naturforschers und Ingenieurs*, vol. 5, *Variationsrechnung*, Zurich, 1947.

University of Michigan and
USN Postgraduate School