THREE THEOREMS ON THE ENVELOPE OF EXTREMALS

BY MARSTON MORSE

1. *Introduction.* We are concerned with the envelope in the small, more specifically with the conjugate locus in the small. In the large, noteworthy papers have recently been written by Whitehead [**4**] and Myers [**3**], and the reader may also refer to the work of the author [**6**].

In the analytic case in the plane the theory is relatively complete. For a brief account and references see Bolza [**1**], pages 357–369. In 3-space Mason and Bliss [**2**] have treated the envelope in the case where the envelope is ordinary. Hahn [**5**] has reduced the minimum problem in 3-space in the non-parametric form to the study of an analytic function of two variables whose Hessian vanishes at the point in question. This transformation of the problem does not however clear up the difficulties inherent in the envelope theory.

There are three theorems on the envelope which go considerably further than the above theory. Of these theorems the first is a topological characterization of a conjugate point, and has been proved by Morse and Littauer [**7**]. The second theorem is a basic result on the analytic representation of the envelope neighboring one of its points. It is an immediate consequence of two theorems proved by the author, one [**9**] on the order of a conjugate point, and the other, Morse ([**6**], page 235), on the continuation of conjugate points. It is similar to a theorem independently derived from the author's results by Whitehead ([**4**], page 690).

The two preceding theorems refer to the analytic case. The theorem to which most of this paper is devoted is not so restricted. It gives sufficient conditions for a relative minimum in the problems in parametric form when the end points A and B of the given extremal g are conjugate.

2. *The Functional.* Let R be an open region in the space of the variables $(x_1, \cdots, x_m) = (x)$. Let

$$F(x_1, \cdots, x_m, r_1, \cdots, r_m) = F(x, r)$$

be a function of class C''' for x in R and $(r)\neq(0)$. We suppose that $F(x, r)$ is positively homogeneous in the variables r of order one. We shall start with the integral in the usual parametric form

$$J = \int_{t_1}^{t_2} F(x, \dot{x})dt,$$

where $(\dot{x})$ stands for the set of derivatives of (x) with respect to t.

Let g be an extremal lying in R with end points A and B. We assume that the Legendre S-condition holds along g([**6**], p. 120). If F is analytic, g will be an analytic curve.

3. *A Local Representation of the Envelope.* Referring to the extremal g of the preceding section, let (λ) be the set of direction cosines of the tangent to g at A. Let (λ) represent a point P on a unit $(m-1)$-sphere S with center at the origin. Let (α) be a set of parameters in a regular representation of S neighboring P on S. Suppose that $(\alpha)=(\alpha_0)$ represents P. Let s be the arc length of the extremals issuing from A, measuring s from A. It is well known that the extremals which issue from A with directions determined by (α) can be represented in the form ([**6**], p. 117) $x_i=x_i(s, \alpha)$, $(i=1, \cdots, m)$, where the functions $x_i(s, \alpha)$ and their first partial derivatives as to s are of class C'' (analytic if F is analytic) for (α) near (α_0) and values of s near those on g. The real zeros (s, α) of the Jacobian

$$\Delta(s, \alpha) = \frac{D(x_1, \cdots, x_m)}{D(s, \alpha_1, \cdots, \alpha_n)}, \qquad (n = m - 1),$$

when substituted in $x_i(s, \alpha)$ define the *envelope* of the family. Let s_0 be the length of g. We note that $\Delta(s_0, \alpha_0)=0$ since B is assumed conjugate to A. From the result on page 119 of [**6**], we infer that $\Delta(s, \alpha_0)$ admits a representation neighboring (s_0) of the form

$$\Delta(s, \alpha_0) = (s - s_0)^r A(s), \qquad (A(s_0) \neq 0),$$

in which the integer r is at most $m-1$ and $A(s)$ is analytic at s_0.

If F is analytic, $\Delta(s, \alpha)$ can be represented neighboring (s_0, α_0) in the form (see Osgood [**10**])

$$\Delta(s, \alpha) = \phi(s, \alpha)B(s, \alpha), \qquad (B(s_0, \alpha_0) \neq 0),$$

where

$$\phi \equiv (s - s_0)^r + A_1(\alpha)(s - s_0)^{r-1} + \cdots + A_r(\alpha),$$

where $A_i(\alpha)$ and $B(s, \alpha)$ are analytic at (s_0, α_0) and A_i vanishes there. The equation $\phi = 0$ is accordingly satisfied by r roots s, real or complex, corresponding to each set (α) sufficiently near (α_0). But according to Lemma 13.3 of Morse ([**6**], p. 235), these r roots will all be real if (α) is real and sufficiently near (α_0). The first theorem is then as follows.

THEOREM 1. *The sets* (s, α) *neighboring* (s_0, α_0) *which determine points on the envelope can be grouped into* r *real single-valued, continuous functions,* $s_i = s_i(\alpha)$, $(i = 1, \cdots, r)$, *analytic except at most on analytic loci* M_p *of dimension* $p < n$. *Any two of these functions which are not identical will be distinct except at most on loci similar to* M_p.

4. *Sufficient Conditions Involving the Envelope.* Let A_e denote the set of extremals which run from A and make angles at most e with the ray positively tangent to g at A. Let e' and e'' be two positive constants, and let $H(g, e', e'')$ be the set of points at a distance less than e' from B, excluding B, and lying on rays issuing from B making angles less than e'' with the ray negatively tangent to g at B. With this understood we state the following theorem.

THEOREM 2. *If* B *is the first conjugate point of* A, *sufficient conditions that* g *afford a proper, strong, relative minimum in the problem in parametric form are as follows*: (I) *that the Weierstrass and Legendre S-conditions hold along* g; (II) *that there exist positive constants* e, e', *and* e'', *such that* $H(g\ e', e'')$ *contains no conjugate point of* A *on the respective extremals of* A_e; (III) *that no extremal of* A_e *pass through* B *save* g.

In this theorem it is understood that the domain of the points (x) consists of any sufficiently small neighborhood of g. To prove the theorem we suppose that a transformation has been made to coordinates $(z_1, \cdots, z_m) = (x, y_1, \cdots, y_n)$, $(n = m - 1)$, as in [**8**], (p. 381), and that g is carried into the segment γ of the x axis with $a \leqq x \leqq 0$. One thereby obtains a new integrand $G(z, \dot{z})$ in parametric form. If the hypotheses of the theorem are satisfied by g and $F(x, \dot{x})$ and the corresponding envelope, it is clear that the hypotheses of a similar theorem will be satisfied by the extremal γ by $G(z, \dot{z})$ and the transformed envelope. To avoid undue complexity we understand from this point on that Theorem 2 refers to the transformed extremal function G, and envelope. We shall refer to the extremal γ as the extremal g. It is clear that no generality will be lost in the proof by virtue of these conventions. The region $H(g, e', e'')$ taken in its new sense will be bounded in part by a cone $K(e'')$ whose vertex angle is $2e''$ and whose axis is the negative axis of x. We continue with the following lemma.

LEMMA 1. *If η is a sufficiently small positive constant, no extremal of A_η will meet the cone $K(e'')$ in more than one point.*

The proof of this lemma can be left to the reader.

The discs $D(\sigma, c)$. Let c be a negative constant such that $0 < -c < e' \cos e''$. For such values of c the conical region $H(g, e', e'')$ will intersect the n-plane $x = c$ in an n-dimensional open spherical disc D_c with center on the x axis. We denote the radius of this disc by $\rho(c)$. We now describe a set of discs concentric with D_c. These new discs shall have radii $\sigma \leqq \rho(c)$ and be denoted by $D(\sigma, c)$. Like D_c they shall lie on the n-plane $x = c$ and have centers on the x axis. They shall not include their spherical boundaries. Let α be a constant less than the constants e and η of Theorem 2 and Lemma 1, respectively. If σ is sufficiently small, the disc $D(\sigma, c)$ will have the following property:

(i) There will exist a subset of the extremals of A_α which includes the extremal g, which is such that one and only one of its extremals passes through each point of $D(\sigma, c)$, and which forms a proper field, neighboring each point of $D(\sigma, c)$.

Of the values of σ for which $\sigma \leqq \rho(c)$ let $\sigma(c)$ be the maximum for which $D(\sigma, c)$ possesses the property (i). Inasmuch as $D(\sigma, c)$ is open, such a maximum clearly exists. We continue with the following lemma.

LEMMA 2. *If β is a sufficiently small positive constant and $0 > c \geqq -\beta$, then $\sigma(c) = \rho(c)$.*

We shall begin the proof by establishing the following statement.

(a) *If* $\sigma(c) < \rho(c)$, *there exists an extremal which issues from A making the angle* α *with the x axis and which passes through a point on the boundary of the disc* $D[\sigma(c), c]$.

Let (v) denote the set of coordinates of a point on the n-plane $x=c$. When the point (c, v) is on $D[\sigma(c), c]$, there is a unique extremal E in the set described in (i) which passes through the point (c, v). Let $p_i(v)$ denote the slope functions of this extremal at A. Suppose (a) false. If ϵ is a sufficiently small positive constant, there will then exist no extremals which make the angle α with the x axis and which pass through points on the domain

$$(1) \qquad D[\sigma_1, c] - D[\sigma_0, c], \quad \sigma_1 = \sigma(c) + \epsilon, \quad \sigma_0 = \sigma(c).$$

If, moreover, ϵ is chosen so small that $\sigma(c)+\epsilon \leqq \rho(c)$, there will be no conjugate points of A belonging to the extremals of A_α and lying on the disc $D(\sigma_1, c)$.

It follows that the functions $p_i(v)$ as defined for (v) on $D(\sigma_0, c)$ can be continued along any path that lies on $D(\sigma_1, c)$, yielding thereby only slopes $p_i(v)$ which define extremals of A_α. The functions $p_i(v)$ so extended conceivably might be multiple-valued. But the domain $D(\sigma_1, c)$ is simply connected and it follows from a well known argument due to Schoenflies (see Bliss, [**11**], pp. 37–41) that the extended functions $p_i(v)$ are single valued.

The number $\sigma(c)$ is accordingly not the largest number for which property (i) holds. From this contradiction we infer the truth of (a).

Suppose that the lemma is false. There will then exist a sequence c_m of values of c tending to 0 as a limit as m becomes infinite and such that $\sigma(c_m) < \rho(c_m)$, $(m=1, \cdots)$. Set $\sigma(c_m) = \sigma_m$. Corresponding to c_m, as we have just seen, there will exist an extremal E_m which issues from A making an angle α with the x axis and which passes through a point on the boundary of $D(\sigma_m, c_m)$. Let h_m represent the direction of E_m at A. The directions h_m will have a cluster direction h. Let E^* be the extremal issuing from A with the direction h. The extremal E^* will make an angle α with the x axis at A. Moreover E^* will pass through the point B, since σ_m tends to zero as m becomes infinite. This, however, is contrary to hypothesis III of the theorem. We infer the truth of the lemma.

The field Σ. *We shall define a field* Σ *of extremals which includes as an inner point each point of the x axis for which* $a<x<0$, *which covers the points of* $H(g, e', e'')$ *for which* $x>-\beta$, *where* β *is the constant of Lemma* 2, *and neighboring A includes all extremals which issue from A and make sufficiently small angles with the positive x axis.*

Let $U(c)$ denote the set of extremals in (i) for $\sigma=\rho(c)$ and $0>c>-\beta$. It follows from Lemmas 1 and 2 that the extremals of $U(c)$ are continuations of a subset of extremals of $U(c')$, provided $c>c'$. The field Σ will now be defined as follows. For $a<x\leqq-\beta$, Σ shall consist of the extremals of A_d, where d is so small a positive constant that A_d forms a proper field in the domain of points covered by A_d for $a<x\leqq-\beta$. For $-\beta<x$, Σ shall consist of the continuation of the extremals of $U(-\beta)$ up to the points Q where these extremals meet the conical boundary of $H(g, e', e'')$, not including the points Q. It follows from Lemma 2 that Σ has the properties required.

Let $p_i(x, y)$ be the slope functions in non-parametric form of the extremals of Σ. With the aid of these slope functions the Hilbert integral corresponding to the field Σ and the non-parametric problem will take the form

$$I=\int M(x, y)dx+N_i(x, y)dy_i.$$

We extend the definition of I so as to include the points A and B, taking M and N_i at A and B as zero. On Σ there will exist a function $H(x, y)$ of class C' such that

$$dH(x, y)\equiv Mdx+N_idy_i.$$

Moreover, on Σ the partial derivatives M and N_i of H are bounded in absolute value. It follows that $H(x, y)$ tends to unique limiting values at A and B, respectively, and will be defined at A and B as equal to these values. Upon evaluating I along the x axis one sees that these limiting values are zero. We continue with the following lemma.

LEMMA 3. *The Hilbert integral I is zero along any path h of the form $y_i=y_i(x)$, $(a\leqq x\leqq 0)$, where $y_i(a)=y_i(0)=0$, $y_i(x)$ is of class D', and the path h lies in the field Σ except for its end points A and B.*

The lemma follows at once upon noting that along h

$$\frac{dI}{dx}=\frac{d}{dx}H[x,\,y(x)]$$

except at most at the end points and corners of h. It follows that

$$I_h=H(0,0)-H(a,0)=0,$$

and the proof is complete.

Let c^* be a value of x such that $a<c^*<0$. If parameters $(v_1,\cdots,v_n)=(v)$ are taken sufficiently near zero, the end points A and B of g can be joined to the point $(x,y)=(c^*,v)$ by unique extremals. Let the resulting broken extremal be denoted by $E(v)$. Let the value of J along $E(v)$ be denoted by $J(v)$. If the set (v) is sufficiently near (0), $E(v)$ will be a curve of the type admitted in the preceding lemma. For such a curve we can use the preceding lemma to establish the Weierstrass formula

$$J(v)-J(0)=\int_a^0 E[x,\,y,\,y',\,p(x,\,y)]dx$$

where E is the Weierstrass E-function in non-parametric form with slope functions belonging to the field Σ and with

$$p_i(0,0)=p_i(a,0)=0.$$

It follows that

$$J(v)>J(0),\qquad (v)\neq(0),\tag{2}$$

provided (v) is sufficiently near (0). We continue with the following lemma.

LEMMA 4. *If U is a sufficiently small neighborhood of g, J assumes on g a proper, strong, minimum relative to curves*

$$y_i=y_i(x),\qquad (a\leqq x\leqq 0),\tag{3}$$

of class D' which join the end points of g on U.

To prove the lemma we divide the curve (3) into two segments h and k by the n-plane $x=c^*$ and compare h and k with the respective segments of the broken extremal $E(v)$ which join the same end points. As is well known the component extremals of $E(v)$ will give at least as small a value to J as do h and k respectively, provided U is sufficiently small. For such a U we infer from (2) that g is a minimizing arc relative to curves of the type (3).

Observe that the curve (3) cannot be an extremal if U is sufficiently small by virtue of condition III of the theorem. It follows that the minimum is proper. Our final lemma is as follows.

LEMMA 5. *If N is a sufficiently small neighborhood of g, g affords a proper, strong minimum to J relative to curves λ of class D' in parametric form which lie on N and join the end points of g.*

Let g be divided into four segments by points with x coordinates a_1, a_2, a_3 such that $a<a_1<a_2<a_3<0$. Set $a_0=a$ and $a_4=0$. Let V be a subneighborhood of U. (See Lemma 4.) Let p_i be an arbitrary point on V and on the n-plane $x=a_i$, and let

$$p_ip_{i+1}, \qquad (i = 0, 1, 2, 3), \tag{4}$$

denote the extremal joining p_i to p_{i+1}. If V is sufficiently small, each extremal (4) will afford a proper strong minimum to J relative to curves of class D' in parametric form which join its end points on V and do not cross the n-planes $x=a_{i-1}$, $(i=1, 2, 3)$, $x=a_{i+2}$, $(i=0, 1, 2)$. Let $N\subset V$ be so small a neighborhood of g that the extremals (4) lie on V when the points p_i lie on N. With this choice of the neighborhood N, we shall establish the lemma. To that end we shall replace the given curve λ by a curve which is admissible in Lemma 4, without however increasing the value of J.

Let k be a segment of λ whose end points p_i and p_{i+1} lie on the n-planes $x=a_i$ and $x=a_{i+1}$, respectively, and which does not cross the n-planes $x=a_{i-1}$ and $x=a_{i+2}$. If k is a proper subarc of no arc of λ with these properties, we shall admit a replacement of k by the corresponding extremal (4). A finite number of such replacements suitably chosen and successively performed will yield a curve admissible in Lemma 4. Lemma 5 and our theorem follow.

Bibliography

1. O. Bolza, *Variationsrechnung*, Teubner.

2. G. A. Bliss and M. Mason, *The properties of curves in space which minimize a definite integral*, Transactions of this Society, vol. 9 (1908), pp. 440–466.

3. S. B. Myers, *Riemannian manifolds in the large*, Duke Mathematical Journal, vol. 1 (1935), pp. 39–49.

4. J. H. C. Whitehead, *On the covering of a complete space by the geodesics through a point*, Annals of Mathematics, vol. 36 (1935), pp. 679–705.

5. H. Hahn, *Über räumliche Variationsprobleme*, Mathematische Annalen, vol. 70 (1911), pp. 110–142.

6. M. Morse, *Calculus of Variations in the Large*, Colloquium Publications of this Society, vol. 18, 1934.

7. M. Morse and S. B. Littauer, *A characterization of fields in the calculus of variations*, Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 724–730.

8. M. Morse, *The foundations of the calculus of variations in the large in m-space*, Transactions of this Society, vol. 31 (1929), pp. 379–404.

9. M. Morse, *The order of vanishing of the determinant of a conjugate base*, Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 319–320.

10. W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. 2, 1929. Teubner.

The Institute for Advanced Study