A METHOD FOR MINIMIZING THE SUM OF ABSOLUTE VALUES OF DEVIATIONS

By Robert R. Singleton

1. Introduction. In the *Philosophical Magazine*, 7th series, May 1930, E. C. Rhodes described a method of computation for the estimation of parameters by minimizing the sum of absolute values of deviations. His is an iterative and recursive method, in the following sense. There is a direct method for minimization with one parameter. Assuming a method for minimization with $n - 1$ parameters, Rhodes imposes a relation between the n parameters (in an n-parameter problem) and finds a restricted minimum by the method for $n - 1$ parameters. In this sense his method is recursive. He then repeats the process, by imposing on the n parameters a new relation determined by the restricted minimum. In this sense his method is iterative. The process is finite, ending when a restricted minimum immediately succeeds itself, indicating a true minimum.

Rhodes' paper presents the method without proof. The purpose of the present paper is to analyze the situation in detail sufficient to indicate proofs for various methods, and to present a new method which reduces the labor of solution by eliminating the recursive feature. The iterative approach is retained. The solution of Rhodes' illustrative problem will be given for comparison between the two methods.

The paper uses geometric terminology and develops to quite an extent the geometry of a surface representing the summed absolute deviations. This seems the clearest means of presenting the relationships. Further analysis of the properties of this surface should lead to an even more direct method for attaining the minimum than the one here presented.

In the writing of the paper, no attention has been given to sets of observations or equations among which a linear dependence may exist. In practice, such a situation almost never occurs. If the need arises, the adjustments which must be made to take care of dependence are in each case fairly obvious.

2. Geometric Analogue of Summed Absolute Deviations. Let n observations on $\nu + 1$ variates be represented by x^i_α, y^i where $i = 1, \cdots, n$; $\alpha = 1, \cdots, \nu$. Unless otherwise noted, latin indices have range 1 to n, greek indices, 1 to ν. The summation convention of tensor analysis is used.

The variates are to be statistically related by the linear function[1]

$$\hat{y}^i = x^i_\alpha u^\alpha,$$

[1] This includes the linear function with a constant, since a variate $x^i = 1$ may be used.

$\hat{y}^i$ being an estimate of y^i. u^α are to be determined so that $v = \Sigma_i \mid \hat{y}^i - y^i \mid$ is a minimum. Set

$$v^i = x^i_\alpha u^\alpha - y^i \tag{1}$$

and determine functions $e^i(u^\alpha)$ so that $e^i v^i \geq 0$, and $\mid e^i \mid = 1$. It is immaterial that e^i is not uniquely determined when u^α satisfies $v^i = 0$. Then $v = \Sigma_i e^i v^i$ is to be minimized. Using (1),

$$v = x_\alpha u^\alpha - y \tag{2}$$

where

$$x_\alpha = \Sigma_i e^i x^i_\alpha, \qquad y = \Sigma_i e^i y^i.$$

Consider a Euclidean $(\nu + 1)$-space, $E_{\nu+1}$, with coordinates $u^1, \cdots, u^\nu, v$. The coordinate hyperplane perpendicular to the v-axis will be called E_ν. In $E_{\nu+1}$ each of equations (1) for a particular i represents a ν-plane which intersects E_ν in a $(\nu - 1)$-plane when $v^i = 0$. Each of the equations

$$v^i = e^i(x^i_\alpha u^\alpha - y^i) \tag{3}$$

represents two half-planes which touch E_ν and each other along the $(\nu - 1)$-plane given in E_ν by the equation

$$x^i_\alpha u^\alpha - y^i = 0. \tag{4}$$

The functions on the right-hand side of (3) are thus continuous everywhere, and linear in any neighborhood of E_ν none of whose points satisfies (4). Since a sum of functions continuous and linear in a neighborhood is also continuous and linear in that neighborhood, it follows that the function on the right in (2) is continuous for all u, and linear for every neighborhood of E_ν containing no points which satisfy (4) for any i. Hence

OBSERVATION I: *The surface (S) given in $E_{\nu+1}$ by (2) consists of portions of ν-planes joined together. The projection of these joins on E_ν forms a network of $(\nu - 1)$-planes determined in E_ν by equations (4).*

3. Existence of a Minimum. Define a "bend of degree r on S" to be the locus of all points on S whose u-coordinates satisfy a set of r independent equations of (4). To each set of r independent equations corresponds a unique bend of degree r.

If a linear relation $u^\alpha = a^\alpha_\sigma \lambda^\sigma + b^\alpha$; $\sigma = 1, \cdots, \mu < \nu$, rank $(a^\alpha_\sigma) = \mu$, is imposed on u^α, all the preceding development, reduced in dimension, applies to the new variates $x^i_\alpha a^\alpha_\sigma$, $y^i - x^i_\alpha b^\alpha$.

OBSERVATION II: *A section of S by a plane of any dimension $d < \nu$ has all the properties of an S-surface of dimension d.*

Since any set of consistent equations selected from (4) determines such a linear relation for u^α, the application of Observation I to any of the bends of S shows that each r-bend consists of linear elements of dimension $\nu - r$, joined at points which lie on linear elements of lesser dimension. Thus S is a polyhedron. Its faces we term complexes of dimension ν, C_ν, and the linear elements of its edges which lie wholly in bends of degree r, but not of degree $r + 1$ are complexes $C_{\nu-r}$ of dimension $\nu - r$. The boundary of any C_α, $\alpha > 0$, consists of complexes of lesser dimension. The term complex is not restricted to either open or closed complexes.

Since the function $v(u^\alpha)$ of (2) is non-negative, it possesses a greatest lower bound (g.l.b.) g. Since for some number $h > g$, there exists an N such that for all $| u^\alpha | > N$, $v(u^\alpha) > h$, it follows that for some closed neighborhood of E_ν the g.l.b. of v is g. Since v is continuous everywhere it attains its g.l.b., and so S has minimum points. Since the minimum of any complex not parallel to E_ν, lies on its boundary, and the boundary consists of complexes, it follows that the minimum points of S consist of C_0's and/or entire complexes of dimension > 0 which are parallel to E_ν. The next section will show that S has a unique minimum complex (including of course its boundary complexes) and furthermore is cup-shaped.

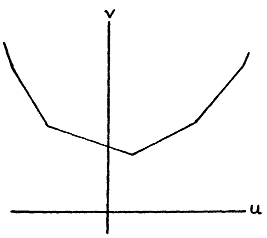

FIG. 1

4. Convexity Property; Uniqueness of the Minimum. Consider $\nu = 1$ in the preceding treatment (and for convenience not written). S looks generally like Fig. 1. The slope changes only where an equation of (4) has a root. Suppose the point is u_0, and $x^1u_0 - y^1 = 0$. From (3), since $v^1 \geq 0$, it follows that $e^1x^1 < 0$ for $u < u_0$, $e^1x^1 > 0$ for $u > u_0$. Since in (2) $x = \Sigma_i e^i x^i$, and since for h sufficiently small and $u_0 - h < u < u_0 + h$ the only e to change value[2] is e^1, we have that

$$x(u_1) + 2 \, | e^1x^1 | = x(u_2)$$

where

$$u_0 - h < u_1 < u_0 < u_2 < u_0 + h.$$

Hence the slope is a monotonic increasing step function. Since for u sufficiently small all $e^ix^i < 0$, and for u sufficiently large all $e^ix^i > 0$, at some intermediate point or points either the slope is zero or it changes from negative to

[2] The e's corresponding to equations proportional to equation (1) also change value at x_0. This does not destroy the argument.

positive without becoming zero. In the first case a single closed C_1 is the minimum complex; in the second, a C_0. In either case the curve given by (2) when $\nu = 1$ is concave upward and has just one minimum complex, except for complexes of lesser dimension constituting the boundary of this complex. An obvious consequence is

LEMMA I. *The set of points u for which v is less than some number N form a convex point set.*

This result is easily extended to the general dimension ν. If for any two points u_1, u_2 of E_ν, $v(u_1) < N$ and $v(u_2) < N$, the plane in $E_{\nu+1}$ given by $u^\alpha = u_1^\alpha + \lambda(u_2^\alpha - u_1^\alpha)$ makes a one-dimensional section of S. By Observation II, the points u lying on the projection of this section on E_ν have the property of Lemma I and of course lie on the straight line joining u_1 and u_2. This is the property required for a convex point set. Hence

THEOREM I. *The set of points u^α of E_ν for which $v(u^\alpha)$ as given by (2) is less than a fixed quantity form a convex point set.*

From this it follows immediately that there is a unique minimum complex. It is appropriate here to point out that no two complexes can be contained in a single plane of the same dimension. This follows from the equation giving monotonicity of slope in one dimension, and Observation II.

5. Gradient Directions. From here on the treatment will be of v as a function defined on E_ν, and the equations will represent objects in E_ν, unless otherwise stated. Complex and Bend also will refer to the projections on E_ν of the complexes and bends of S. For a single-valued function defined on E_ν the gradient at a point is the projection of a normal to the surface representing the function in $E_{\nu+1}$. If the function is defined only over a subspace of E_ν possessing derivatives, the gradient will be required also to be tangent to the subspace. This is sufficient to determine a unique direction, and preserves the property that for an infinitesimal displacement in any direction the value of the function decreases most rapidly in the direction of the gradient. Here gradient is taken negative to its usual sense.

A point u lying on a C_r but not on a C_{r-1} will have a gradient in C_r and also in each higher-dimensional complex on whose boundary C_r lies. If the gradient for u as a point of C_{r+k} points into C_{r+k} (remembering that u lies on the boundary) this will be called a usable gradient. In the case of the greatest k for which there exists a usable gradient, there exists but one C_{r+k} providing such a gradient, and that gradient is the "best" gradient; that is, of all directions in E_ν it provides the direction of most rapid decrease of the function v. This follows from Theorem I. Furthermore, all complexes of lesser dimension providing usable gradients lie on the boundary of this C_{r+k}. In fact

THEOREM II. *If for a point u on C_r, two complexes C_s and C_s', $s > r$, lying in different bends of degree $\nu - s$ but incident at C_r, both provide usable gradients for u, then the complex C_{s+1} on whose boundary lie both C_s and C_s' also provides a usable gradient for u.*

This follows from Theorem I. Select u_1 on the gradient in C_s, u_2 on the gradient in C'_s, for which $v(u_1) = v(u_2)$. The join of u_1 and u_2 lies in C_{s+1}, and for some point, u_3 on this join, $v(u_3)$ is less than $v(u_1) = v(u_2)$. Also, the distance $\overline{uu_3}$ is less than at least one of $\overline{uu_1}$, $\overline{uu_2}$. Hence C_{s+1} must contain a usable gradient.

6. Selection of Best Gradient at Bends. The direction of the gradient for a point u_0 considered as lying on a C_ν is given by

$$g^\alpha = -x_\alpha(u_0) = -\Sigma_i e^i(u_0)x^i_\alpha. \tag{5}$$

If u_0 lies in the interior of a face, this is unique. If u_0 lies in a bend, so that some e^i are not determined, the g^α for each face is found by selecting the indeterminate e's as $+1$ or -1, according to the face being considered.

For a point u_0 considered as lying on a bend of degree r, given by r independent equations of (4):

$$x^\lambda_\alpha u^\alpha - y^\lambda = 0, \qquad (\lambda = 1, \cdots, r), \tag{6}$$

the gradient for a particular $C_{\nu-r}$, determined by the conditions at the beginning of section 5, is

$$g^\alpha = x^\lambda_\alpha k_\lambda - x_\alpha \tag{7}$$

where k_λ satisfies

$$\Sigma_\alpha x^\mu_\alpha x^\lambda_\alpha k_\lambda = \Sigma_\alpha x^\mu_\alpha x_\alpha, \qquad (\mu = 1, \cdots, r)$$

and x_α is as given in (2), the choice of sign for the indeterminate e^λ $(\lambda = 1, \ldots, r)$ being immaterial. They may, in fact, be taken as 0 in this instance.

For a point u_0^α lying on an r-bend given by (6), to determine which complex contains the best gradient, each $(r - 1)$-bend incident on the r-bend at u_0 is tested for a usable gradient. Theorem II then determines the complex containing the best gradient.

There are $2r$ such complexes incident at u_0, given by the r sets of equations selected from (6):

$$(\lambda): x^\sigma_\alpha u^\alpha - y^\sigma = 0 \qquad \begin{matrix}(\sigma = 1, \cdots, \lambda - 1, \lambda + 1, \cdots, r) \\ (\lambda = 1, \cdots, r).\end{matrix} \tag{8}$$

The two complexes lying in the same $(r - 1)$-bend have the same equations in (8), but are distinguished later by $e^\lambda(u_0)$ for the omitted equation being taken first $+1$, then -1.

The gradient for the λth pair of complexes is

$$g^\alpha_\lambda = x^\sigma_\alpha k_\sigma - x_\alpha$$

similar to (7), but not identical. For $e^\lambda = +1$ in determining x_α, we have $g^\alpha_{\lambda+}$, and for $e^\lambda = -1$, $g^\alpha_{\lambda-}$. We restrict the consideration to $e^\lambda = +1$.

The line in the direction of greatest slope is then

$$u^\alpha = u_0^\alpha + {}'g_{\lambda+}^\alpha t.$$

Now u_0 is here considered lying on the complex given by (8λ) with $e^\lambda = +1$. In order that $g_{\lambda+}^\alpha$ point into this face, the deviation for the λth observation must exceed 0 when $t > 0$; otherwise, for a displacement in the direction of $g_{\lambda+}^\alpha$, e^λ changes sign immediately and the course is in the other complex. This deviation is

$$v^\lambda = x_\alpha^\lambda u^\alpha - y^\lambda = x_\alpha^\lambda u_0^\alpha - y^\lambda + x_\alpha^\lambda g_{\lambda+}^\alpha t = x_\alpha^\lambda g_{\lambda+}^\alpha t.$$

Had $g_{\lambda-}^\alpha$ been used, this deviation must be less than 0. Hence a necessary and sufficient condition that a complex given by (8) with either choice of e^λ possess a usable gradient is

$$\Phi_\lambda = e^\lambda[\Sigma_\alpha x_\alpha^\lambda x_\alpha^\sigma k_\sigma - \Sigma_\alpha x_\alpha^\lambda x_\alpha] > 0. \tag{9}$$

For $r = 1$ the condition is given by (9) with the first sum merely omitted. $\Phi_{\lambda+}$ and $\Phi_{\lambda-}$ cannot both exceed 0.

When all sets of equations (8λ) are tested by (9) the equations common to all sets possessing a usable gradient determine the complex with the best gradient, retaining the values of e for which (9) was satisfied.

7. Property of the Minimum Point. For a minimum point, given by (6) with $r = \nu$, all Φ_λ must be negative. Define $X^{\beta\gamma} = \Sigma_\alpha x_\alpha^\beta x_\alpha^\gamma$ and $X^{\beta 0} = \Sigma_\alpha x_\alpha^\beta x_\alpha$ for convenience. Then in (9), the numbers k_σ, -1 are seen from their definition in (7) to be proportional to the cofactors of the λth row of the matrix $(X^{\mu\sigma}, X^{\mu 0})$, μ having the same range as λ. Thus $\Phi_{\lambda+} = c \text{ Det } (X^{\mu\sigma}, X_+^{\mu 0})$, and $\Phi_{\lambda-} = -c \text{ Det } (X^{\mu\sigma}, X_-^{\mu 0})$, where in the first case $X^{\mu 0}$ is determined with $e^\lambda = +1$, in the second with $e^\lambda = -1$. The factor of proportionality, c, must be the same since $X^{\mu\sigma}$ is unaffected by change of e^λ. Now let $X^\mu = \Sigma_\alpha x_\alpha^\mu \overset{*}{x}_\alpha$ where $\overset{*}{x}_\alpha = \Sigma_k e^k \overset{k}{x}_\alpha$, the range of k omitting the range of λ. Then

$$\Phi_{\lambda+} = c\,[\text{Det } (X^{\mu\sigma}, X^\mu) + \text{Det } (X^{\mu\sigma}, X^{\mu\lambda})]$$

and

$$\Phi_{\lambda-} = -c\,[\text{Det } (X^{\mu\sigma}, X^\mu) - \text{Det } (X^{\mu\sigma}, X^{\mu\lambda})].$$

Hence

$$\Phi_{\lambda+}\Phi_{\lambda-} = -c^2\,\{[\text{Det } (X^{\mu\sigma}, X^\mu)]^2 - [\text{Det } (X^{\mu\sigma}, X^{\mu\lambda})]^2\}.$$

Now let A represent the square matrix (x_λ^α), α giving the rows and λ the columns. Let B_λ represent the matrix formed from A by replacing the λth column by $\overset{*}{x}_\alpha$. Then

$$\begin{aligned}\Phi_{\lambda+}\Phi_{\lambda-} &= -c^2\,[\text{Det}^2\,(A'B_\lambda) - \text{Det}^2\,(A'A)]\\ &= -c^2\,\text{Det}^2\,A\,(\text{Det}^2\,B_\lambda - \text{Det}^2\,A)\end{aligned}$$

and this will have the same sign as

$$\Psi_\lambda = |\,\text{Det}\,(A)\,| - |\,\text{Det}\,(B_\lambda)\,|.$$

Since $\Phi_{\lambda+}$ and $\Phi_{\lambda-}$ are never both positive, and at the minimum are both negative for all λ, at the minimum all $\Psi_\lambda > 0$. To determine all Ψ_λ together, let, in matrix notation, $z' = (z_1, \cdots, z_\nu)$ and $x^{*\prime} = (x_1^*, \cdots, x_\nu^*)$ where x_α^* were defined previously. Determine z as the solution of $Az = x^*$. Then $|\,\text{Det}\,(B_\lambda)\,|$ are equal to $|\,z_\lambda|\,|\,\text{Det}\,(A)\,|$. Hence a necessary and sufficient condition that $\Psi_\lambda > 0$ for all λ is that all $|\,z_\lambda\,|$ be less than one. Hence

THEOREM III: *If a zero-complex is given by a set of equations whose matrix is M, a necessary and sufficient condition that the complex be a unique minimum is that the solutions of $M'z = x^*$ be all less than one in absolute value. If k of the solutions are equal to one in absolute value, and the rest are less than one, the minimum is a complex of dimension k with the zero-complex as one of its corners.*

The last statement follows since if one solution is 1 in absolute value, a corresponding $\Phi_\lambda = 0$, and hence no gradient, usable or not, exists. Thus the corresponding complex is parallel to E_ν.

8. Minimization for One Dimension. A method for minimization of (2) when there is just one parameter evolves from the monotonicity of slope in that case. Suppose the variates are w^i and z^i, and (1) is

$$v^i = w^i t - z^i. \tag{10}$$

Suppose the variates are arranged in order of z^i/w^i, starting with the smallest. The slope of the rth segment (Fig. 1) from the left is

$$\sum_{i=1}^{r} |\,w^i\,| - \sum_{i=r+1}^{n} |\,w^i\,|.$$

The minimum occurs when the slope is 0 or changes from negative to positive; that is, when the first sum equals or exceeds the second; or when the first sum equals or exceeds half the total. This is a standard computation. If the change takes place when $r = k$, then $t = z^k/w^k$ is the value of t giving the minimum.

9. **Mimimization Procedure for $\nu + 1$ Dimensions.** For any continuous function with unique minimum and having the property of Theorem I, the following holds. Let u_0 be any point of E_ν. Let $u_{i+1} = u_i + \lambda_i t_i$, where λ_i is any direction chosen at random and t_i is the value of t for which the function attains a minimum on the curve $u = u_i + \lambda_i t$. Then the probability is one that $\lim_{i\to\infty} u_i = u_1$, where u_1 is a minimum point for the function. If λ_i is taken always as the gradient of u_i, such a procedure is called the "method of steepest descent" for approaching the minimum point.

Usually the limit is never attained. In this case, however, the minimum is

TABLE I

Method of Steepest Descent Applied

	Data				First Restricted Minimum					Second Restricted Minimum					Third (absolute) Minimum					for test	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(1)
i	x_1^i	x_2^i	x_3^i	y^i	z_0^i	$e^i(u_0)$	w_0^i	z_0^i / w_0^i	rank of (9)	z_1^i	e_1^i	w_1^i	z_1^i / w_1^i	rank of (14)	z_2^i	e_2^i	w_2^i	z_2^i / w_2^i	rank of (19)	e_3^i	i
1	1	−8	64	28	−78	−	−5165	.015	16	−47.044	−	−1149.570	−.0409	6	−92.773	−	−285.180	.3253	15	−	1
2	1	−7	49	−39	14	+	−3917	−.004	8	37.476	+	−942.102	.0398	12	0		0				2
3	1	−6	36	−40	36	+	−2841	−.013	6	53.027	+	−753.324	.0704	14	23.060	+	247.156	.0933	13	+	3
4	1	−5	25	−1	14	+	−1937	−.007	7	25.609	+	−583.236	.0439	13	2.408	+	456.288	.0053	8	+	4
5	1	−4	16	23	3	+	−1205	−.002	9	10.222	+	−431.838	.0237	11	−6.956	−	627.396	−.0111	2	−	5
6	1	−3	9	34	1	+	−645	−.0015	10	4.866	+	−299.130	.0163	10	−7.033	−	760.480	−.0092	3	−	6
7	1	−2	4	25	15	+	−257	−.058	2	16.540	+	−185.112	.0894	15	9.177	+	855.540	.0107	9	+	7
8	1	−1	1	40	1	+	−41	−.024	3	1.246	+	−89.784	.0139	8	−2.326	−	912.576	−.0025	5	−	8
9	1	0	0	43	−5	−	3	−1.67	1	−5.018	−	−13.146	−.3817	2	−5.541	−	931.588	−.0059	4	−	9
10	1	1	1	29	2	+	−125	−.016	4	2.749	+	44.802	−.0614	5	4.531	+	912.576	.0050	7		10
11	1	2	4	14	6	+	−425	−.014	5	8.547	+	84.060	−.1017	4	11.891	+	855.540	.0139	10	+	11
12	1	3	9	12	−7	−	−897	.008	14	−1.624	−	104.628	.0155	9	2.538	+	760.480	.0033	6	−	12
13	1	4	16	−16	2	+	−1541	−.001	11	11.236	+	106.506	−.1055	3	15.472	+	627.396	.0247	12	+	13
14	1	5	25	−14	−23	−	−2357	.010	15	−8.874	−	89.694	.0989	16	−5.306	−	456.288	−.0116	1	−	14
15	1	6	36	−46	−18	−	−3345	.005	12	2.048	+	54.192	−.0378	7	4.203	+	247.156	.0170	11	+	15
16	1	7	49	−68	−27	−	−4505	.006	13	0		0			0		0				16
17	1	8	64	−24	−106	−	−5837	.018	17	−71.017	−	−72.882	−.9744	1	−73.916	−	−285.180	.2592	14	−	17
Σ	17	0	408	0	−170		−35037			39.990		−4036.242			−120.570		8080.100				Σ

attained. The minimum can be approached as closely as desired, hence a complex incident on the minimum is reached. But the convex point sets of Theorem I surrounding the minimum complex are all similar convex polyhedrons in E_ν, whose corresponding faces are parallel, and the gradients at points on a bend cannot point into a higher dimensional complex on the bend. Hence the sequence of points lie on bends of successively greater degree, and must eventually attain the minimum complex.

TABLE II

Points u_k

$u_{k+1}^\alpha = u_k^\alpha + g_k^\alpha t_k$
$u_0 = (38,\ -5,\ -2)$
$u_1 = (37.98202,\ -4.74828,\ -1.48457)$
$u_2 = (37.45908,\ -2.07142,\ -1.85631)$
$u_3 = (32.83333,\ -2.07142,\ -1.76191)$

TABLE III

Computation of $t_k = z_k/w_k$

$\Sigma \mid w_k \mid$	in order of col.	exceeds	at i =	hence t_k =
$\Sigma \mid w_0 \mid$	(10)	17521	16	.00599334
$\Sigma \mid w_1 \mid$	(15)	2502	2	.0397792
$\Sigma \mid w_2 \mid$	(20)	4610	10	.00496545

TABLE IV

Gradients g_k^α for column $(5k + 8)$

k	g_k^1	g_k^2	g_k^3
0	−3	42	86
1	−13146	67293	−9345
2	−931588	0	19012

The computational procedure is as follows:

1. Select a point u_0.
2. Determine the gradient g_0^α from (5).
3. Compute $w_0^i = x_\alpha^i g_0^\alpha$, $z_0^i = y^i - x_\alpha^i u_0^\alpha$.
4. Determine t_0 by the method of section 8.
5. Compute $u_1^\alpha = u_0^\alpha + g_0^\alpha t_0$.
6. Determine the complex containing the best gradient by (9), and the gradient g_1^α by (7).

and so proceed to the minimum. This may be finally tested by Theorem III.

Step 5 is unnecessary, since the only use for u_1^α is to determine $e^i(u_1)$. But $e^i(u_1) = e^i(t_0)$, the latter referring to the computation in step 4. Also, after the first step, it is easier to compute z^i by

$$z_{k+1}^i = z_k^i - w_k^i t_k .$$

10. **Example.** The computation for (9) is not so great as it would seem, since some of the work is duplication and some must be computed anyway for the gradient. Even so, for $r \geq 3$ it becomes, perhaps, more arduous than its contribution would seem to justify. For $\nu \geq 4$ it is recommended that the test of (9) be omitted for points on bends of third degree or greater, and the final test of Theorem III be applied at the end of the work. If this test shows the minimum has not been reached, the complex in which lies the best gradient will be indicated at the same time.

The minimum number of steps is 0. The maximum number is tremendous but finite. The expected number is probably a little greater than ν.

In Tables I to IV, the method is applied to the problem used by Rhodes to illustrate his method. The independent variates are shown in columns (2), (3), (4), Table I, the dependent variate in column (5). The only other original datum is the initial point, selected by guess, shown in line 1, Table II. Since slightly different formulas were used in the computation, the signs of cols. (6), (8), (11), (16), (18) are reversed, and the gradients in Table IV are multiplied by constants. As they are used only for directions, this does not matter.

PRINCETON UNIVERSITY,
PRINCETON, N. J.