Hilbert's Metric and Positive Contraction Mappings in a Banach Space

P. J. BUSHELL

Communicated by J. SERRIN

1. Introduction

The metric to be considered was introduced by DAVID HILBERT [13] in 1895 in an early paper on the foundations of geometry. Special cases of Hilbert's metric occur in the earlier work of Cayley, Beltrami and KLEIN [15]. HILBERT [13] constructs a model for a metric hyperbolic geometry in which there are three noncollinear points forming a triangle with the length of one side equal to the sum of the lengths of the other two sides.

The usefulness of HILBERT'S metric in algebra and analysis was made clear by GARRETT BIRKHOFF [3] in 1957. Birkhoff showed that the Perron-Frobenius theorem for non-negative matrices and Jentzch's theorem for integral operators with positive kernel could both be proved by an application of the Banach contraction mapping theorem in suitable metric spaces. BIRKHOFF'S papers [3–6] relied heavily on arguments from differential projective geometry. The present paper is partly expository, supplying simple proofs for BIRKHOFF'S main theorems.

Consideration of positive homogeneous mappings of degree p, 0 , by these methods seems to be new, although some results were given by THOMPSON [19] using a different but related metric. Theorems 3.4 and 3.5 are new. The simple proofs of the applications in Section 5(i), Section 6(ii) and Section 7 also appear to be new.

The metric is defined in Section 2 in a cone of positive elements in a Banach space. Properties of positive homogeneous mappings are studied in Section 3; the necessary completeness criteria for the metric spaces to be considered are given in Section 4. Applications of the theory to non-negative matrices, positive integral operators, positive-definite symmetric matrices and ordinary differential equations are given in Sections 5–8.

The main aim of the paper is to give an elementary introduction to a tool which has been somewhat neglected, perhaps because of the lack of a readily available account of its simple properties.

2. Definition of the Projective Metric

HILBERT'S original definition of the projective metric involved the logarithm of the cross-ratio of certain points in the interior of a convex cone in R^n . We begin with the definition of the HILBERT metric in a general setting.

Hilbert's Metric

Let X be a real Banach space and let K be a closed solid cone in X, that is, a closed subset K with the properties: (i) \hat{K} , the interior of K, is not empty, (ii) $K+K\subset K$, (iii) $\lambda K\subset K$ for all $\lambda \ge 0$ and (iv) $K\cap -K=\{0\}$. The relations \le and < are defined in X in the usual way by saying that $x \le y$ and x < y if and only if $(y-x)\in K$ and $(y-x)\in \check{K}$, respectively. Since the cone is closed, the partial ordering \le is Archimedean, that is, if $nx \le y$ for n=1, 2, 3, ..., then $x \le 0$. The standard example of a bad cone which is neither closed nor Archimedean (nor normal) is $K=\{(x_1, x_2): x_1>0 \text{ or } x_1=0 \text{ and } x_2 \ge 0\} \subset \mathbb{R}^2$, in which $n(0, 1) \le (1, 1)$, and hence $(0, 1) \le (1/n, 1/n)$, for n=1, 2, 3, ...

Definition 2.1. If $x, y \in K^+ = K \setminus \{0\}$, we define $M(x|y) = \inf\{\lambda : x \leq \lambda y\}$, or $M(x|y) = \infty$ if the set is empty, and $m(x|y) = \sup\{\mu : \mu y \leq x\}$.

Lemma 2.1. If $x, y \in K^+$, then

$$m(x|y) \ y \le x \le M(x|y) \ y \tag{2.1}$$

where the right-hand inequality has meaning only if M(x|y) is finite.

Proof. $x \leq \{M(x/y) + 1/n\} y$ for n = 1, 2, 3, ...; hence $n\{x - M(x/y) y\} \leq y$ for n = 1, 2, 3, ..., and by the Archimedean property $x - M(x/y) y \leq 0$. Similarly, $m(x/y) y - x \leq 0$.

Corollary. If $x, y \in K^+$, then

$$0 \le m(x|y) \le M(x|y) \le \infty.$$
(2.2)

Proof. Suppose that M(x|y) < m(x|y); then

$$\{M(x|y) - m(x|y)\} \ y = \{M(x|y) \ y - x\} + \{x - m(x|y) \ y\} \ge 0$$

and

$$-\{M(x/y)-m(x/y)\} y \ge 0.$$

Therefore, $\{M(x/y) - m(x/y)\} = 0$, and hence y=0, contrary to the choice of $y \in K^+$.

Definition 2.2. Hilbert's projective metric $d(\cdot, \cdot)$ is defined in K^+ by

$$d(x, y) = \log\{M(x/y)/m(x/y)\}.$$

Theorem 2.1. $\{K, d\}$ is a pseudo-metric space and $E = \{K \cap U, d\}$ is a metric space, where U denotes the unit sphere in X.

Proof. It is clear from equation (2.2) that $d(x, y) \ge 0$ and from equation (2.1) that d(x, y)=0 if and only if $x = \lambda y$ for some positive λ .

We now show that d(x, y) is finite for all x, y in K. Choose $\varepsilon > 0$ such that $S(x; \varepsilon) \subset K$ and $S(y; \varepsilon) \subset K$, where $S(x; \varepsilon)$ is the ball centre x and radius ε . Then $x - (\varepsilon/||y||) y \ge 0$ and $y - (\varepsilon/||x||) x \ge 0$, and hence

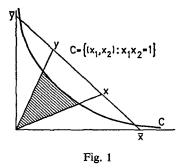
$$0 < \{\varepsilon \mid \mid y \mid \} \le m(x|y) \le M(x|y) \le \{ \mid x \mid | \varepsilon \} < \infty.$$

$$(2.3)$$

The triangle inequality follows easily from the inequalities

$$x \leq M(x|y) \ y \leq M(x|y) \ M(y|z) \ z,$$

23*



which gives $M(x|z) \leq M(x|y)M(y|z)$, and from the companion inequality $m(x|z) \geq m(x|y)m(y|z)$.

The notation used in defining the functions M and m is explained by the following example.

Example. Let $X = R^n$ and $K = \{(x_1, x_2, ..., x_n) : x_i \ge 0 \ (1 \le i \le n)\}$. Then

$$M(x/y) = \max_{i} \{x_i/y_i\}$$
 and $m(x/y) = \min_{i} \{x_i/y_i\}$

and hence

$$d(x, y) = \log \max_{i, j} \{x_i y_j / y_i x_j\}.$$

It is worth remarking that when $X = R^2$,

 $d(x, y) = 2 \times \{\text{Shaded area in sketch}\}$

and

 $d(x, y) = \log |\operatorname{cross-ratio}(x\bar{x}, y\bar{y})|.$

Lemma 2.2. If $x, y \in \breve{K}$, then $d(\lambda x, \mu y) = d(x, y)$ for all $\lambda, \mu > 0$.

The proof of this fundamental property is obvious.

Lemma 2.3. If $x, y \in K$ and $\alpha, \beta \ge 0$, then

- (i) $M(\alpha x + \beta y/y) = \alpha M(x/y) + \beta$,
- (ii) $m(\alpha x + \beta y/y) = \alpha m(x/y) + \beta$,
- (iii) M(x|y) m(y|x) = 1.

Proof. See BAUER [1, 2] for details of the elementary proof.

3. Positive Mappings

In this section we consider properties of mappings in K.

Definition 3.1. If $A: K \to K$ we say that A is non-negative, and if $A: \mathring{K} \to \mathring{K}$ we say that A is positive.

Definition 3.2. If A is positive and if $A(\lambda x) = \lambda^p A(x)$ for all $x \in \mathring{K}$ and $\lambda > 0$, we say that A is positive homogeneous of degree p in \mathring{K} .

Hilbert's Metric

Definition 3.3. If A is positive, the projective diameter $\Delta(A)$ of A is defined by

$$\Delta(A) = \sup \{ d(Ax, Ay) \colon x, y \in \check{K} \}.$$

Definition 3.4. If A is positive, the contraction ratio k(A) of A is defined by

$$k(A) = \inf \{ \lambda \colon d(Ax, Ay) \leq \lambda d(x, y) \ \forall x, y \in \check{K} \}$$

Definition 3.5. If $x, y \in \mathbf{K}$, the oscillation of x and y is defined by

$$\operatorname{osc}(x|y) = M(x|y) - m(x|y)$$

and if A is positive, the oscillation ratio N(A) of A is defined by

$$N(A) = \inf \{ \lambda : \operatorname{osc}(A x / A y) \leq \lambda \operatorname{osc}(x / y) \ \forall x, y \in \check{K} \}.$$

Definition 3.6. If $A: X \to X$, A is said to be monotone increasing if $x \le y$ implies $Ax \le Ay$.

Theorem 3.1. Let A be a monotone increasing positive mapping which is positive homogeneous of degree p in \hat{K} . Then the contraction ratio k(A) does not exceed p.

Corollary. Let A be a positive linear mapping; then $k(A) \leq 1$.

Proof. It follows from

that

$$m(x|y) \ y \le x \le M(x|y) \ y,$$
$$[m(x|y)]^{p} A \ y \le A \ x \le [M(x|y)]^{p} A \ y,$$

and hence

$$[m(x|y)]^{p} \leq m(Ax|Ay) \leq M(Ax|Ay) \leq [M(x|y)]^{p}.$$
(3.1)

If A is a positive linear mapping in X, its contraction ratio is related to its projective diameter.

Theorem 3.2 [BIRKHOFF, 1957]. If A is a positive linear mapping in X, then

$$k(A) = \tanh \frac{1}{4} \Delta(A).$$

Proof [BUSHELL, 1973]. We begin with a result of independent interest; namely,

$$k(A) = N(A). \tag{3.2}$$

OSTROWSKI [17] used a simple limiting argument based on Lemma 2.3, (i) and (ii), to show that $N(A) \leq k(A)$. Using Lemma 2.3, (i), (ii) and (iii), BUSHELL [10] reversed the argument to show that $k(A) \leq N(A)$.

Finally, BAUER [1] showed that $N(A) = \tanh \frac{1}{4} \Delta(A)$, using nothing more complicated than the arithmetic-geometric mean inequality.

Most of our subsequent applications of the theory stem from the next theorem.

Theorem 3.3. Let A be either

(a) a monotone increasing positive mapping which is positive homogeneous of degree p(0 in <math>K, or

(b) a positive linear mapping with finite projective diameter. If the metric space $E = \{ \vec{K} \cap U, d \}$ is complete, then in case (a) there exists a unique $x \in \vec{K}$ such that Ax = x, and in case (b) there exists a unique positive eigenvector of A in E.

P. J. BUSHELL

Proof. Let F(x) = Ax/||Ax||; then F is a map from E into E and is the composition of a strict contraction (from Theorems 3.1 and 3.2) and a normalising isometry (from Lemma 2.2). By the Banach contraction mapping theorem there exists a unique z in E such that F(z)=z, and if we set $x = ||Ax||^{1/(1-p)}z$, the result follows at once.

Figure 1 suggests the existence of many other projective metrics in \hat{K} . In view of the importance of the contraction ratio, the next theorem is of some interest. It asserts that among a wide class of possible projective metrics, HILBERT'S gives the best contraction ratio.

Theorem 3.4. Let F be a differentiable real valued function in $[0, \infty)$ such that

- (i) $F(t) \ge 0$ for $0 \le t < \infty$ and F(0) = 0,
- (ii) $F'(t) \ge 0$ for $0 \le t < \infty$ and F'(0) > 0, and
- (iii) $F(s+t) \leq F(s) + F(t)$.

Then $d_F(x, y) = F(\log[M(x|y)|m(x|y)])$ is a pseudo-metric in \mathring{K} , and if k_F is the associated contraction ratio, then $k_F(A) \ge k(A)$.

Proof. It is easy to check that d_F is a pseudo-metric in \vec{K} . Moreover, if $x, y \in \vec{K}$ and $\varepsilon > 0$, then

$$k_F(A) \ge F\left(\log\left[\frac{\varepsilon M(Ax/Ay)+1}{\varepsilon m(Ax/Ay)+1}\right]\right) / F\left(\log\left[\frac{\varepsilon M(x/y)+1}{\varepsilon m(x/y)+1}\right]\right).$$

Letting $\varepsilon \downarrow 0$ and using l'Hospital's rule to evaluate the limit, we obtain

$$k_F(A) \ge [\operatorname{osc}(A x | A y) | \operatorname{osc}(x | y)].$$

Therefore, $k_F(A) \ge N(A) = k(A)$.

Example. If $X \in \mathbb{R}^2$, K is the first quadrant, and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with a, b, c, d > 0 and $ad - bc \neq 0$, then

$$k(A) = \left| \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right|.$$

If $F(t) = \tanh \frac{1}{2} t$, then elementary calculations show that

$$k_F(A) = \left| \frac{a d - b c}{a d + b c} \right|.$$

We remark also that HILBERT's pseudo-metric always puts the boundary of the cone at an infinite distance from any interior point, but in this example, if $x \in \mathbf{K}$ and $b \in \partial K$, then d(x, b) = 1.

Theorem 3.5. Let $\{X, \leq\}$ be such that if $0 < x \leq y$ then $||x|| \leq ||y||$. If G is a positive mapping from E into E and if $F(x) = G(x) + z_0$, where $z_0 \in \mathcal{K}$, then $k(F) \leq k(G)$.

Proof. If $x, y \in E$ then

$$m(x|y) \le 1 \le M(x|y). \tag{3.3}$$

For, if M(x/y) < 1, then $x \le M(x/y) y$ implies $1 = ||x|| \le M(x/y) ||y|| < 1$, which is absurd; similarly $m(x/y) \ge 1$.

Since $G(x), G(y) \in E$ for all $x, y \in E$, we have

$$F(x) \leq M(G(x)/G(y))F(y) + \{1 - M(G(x)/G(y))\} z_0 \leq M(G(x)/G(y))F(y)$$

and similarly

$$F(x) \ge m(G(x)/G(y)) F(y)$$

Therefore, for all $x, y \in E$,

$$d(F(x), F(y)) \leq d(G(x), G(y)) \leq k(G) d(x, y),$$

and hence $k(F) \leq k(G)$.

4. Completeness Criteria

Next we consider criteria for the completeness of the metric space E. Following our policy of giving elementary arguments when possible, we first consider three important special cases.

Theorem 4.1. If $X = R^n$ and if $K = \{(x_1, x_2, ..., x_n) : x_i \ge 0, 1 \le i \le n\}$, then E is complete.

Proof. First we show that for $x, y \in E$

$$||x-y|| \le \exp\{d(x, y)\} - 1.$$
(4.1)

For, if $x, y \in E$, it is plain that

$$m(x|y) \le 1 \le M(x|y) \tag{4.2}$$

and hence that

$$\|x - y\| = \left\{\sum_{1}^{n} (x_i - y_i)^2\right\}^{1/2} \leq \left\{\sum_{1}^{n} [M(x/y) - m(x/y)]^2 y_i^2\right\}^{1/2}$$
$$\leq \left\{M(x/y) - m(x/y)\right\} \leq \left[\exp\left\{d(x, y)\right\} - 1\right] m(x/y).$$

Moreover,

$$M(x|y) = \max\{1 + [x_i - y_i]/y_i: 1 \le i \le n\} \le 1 + ||x - y||/m(y|u)\}$$

where u = (1, 1, ..., 1). Similarly, if ||x - y|| < m(y/u), then

$$m(x|y) \ge 1 - ||x - y||/m(y|u).$$

It follows that if ||x-y|| < m(y/u), then

$$\|x - y\| \ge m(y/u) \tanh\{\frac{1}{2} d(x, y)\}.$$
(4.3)

The completeness of E follows easily from inequalities (4.1) and (4.3).

Theorem 4.2. If X = C[0, 1] and $K = \{x(\cdot) \in X : x(t) \ge 0 \text{ in } 0 \le t \le 1\}$, then E is complete.

Proof. If u(t) = 1 for $0 \le t \le 1$, then inequalities (4.1)–(4.3) are established easily for elements in E and the result follows.

Theorem 4.3. If $X = B[\mathbb{R}^n]$, the space of real $n \times n$ matrices with norm $||A|| = \sup\{Ax: ||x|| = 1\}$, and if K is the cone of real positive semi-definite symmetric matrices in X, then \hat{K} is the cone of real positive definite symmetric matrices in X and E is complete.

P. J. BUSHELL

Proof. The analogues of inequalities (4.1)–(4.3) are proved in BUSHELL [11].

We state now a general criterion due to BIRKHOFF [1] which covers the examples of Theorems 4.1 and 4.2 but not that of Theorem 4.3. (See also Theorem 5 of BIRKHOFF [4].)

Theorem 4.4. If $\{X, \leq\}$ is a Banach lattice, then E is complete.

5. Applications to Non-Negative Matrices

(i) Let $A = (a_{i,j})$ be an $n \times n$ matrix with non-negative entries and suppose that A is indecomposable, that is, there does not exist a permutation matrix P such that $PAP' = \begin{pmatrix} B & O \\ C & D \end{pmatrix}$, where B and D are square submatrices. It follows that, for some $m \ge 1$, $A^m = B = (b_{i,j})$ has positive entries. Following BIRKHOFF [3], we observe that A^m is a positive linear mapping in the interior of the positive orthant in R^n and that

$$\Delta(A^m) = \max\{\log[b_{ij}b_{pq}/b_{iq}b_{pj}]: 1 \leq i, j, p, q \leq n\} < \infty.$$

Therefore, from Theorem 3.2 and 4.1 and the well-known extension of the Banach contraction mapping theorem, it follows that A has a unique positive eigenvector (with an associated positive eigenvalue) (cf. SAMUELSON [18]).

For the remainder of the Perron-Frobenius theorem proved by these methods, see BIRKHOFF [4].

(ii) Let $A = (a_{i,j})$ be an $n \times n$ matrix with non-negative entries and at least one positive element in each row. Then if $0 , there exists a unique positive solution <math>(x_1, x_2, ..., x_n)$ of the equations

$$x_i = \sum_{j=1}^n a_{ij} x_j^p \quad (1 \le i \le n).$$

Proof. Let $F(x) = \left(\sum_{j=1}^{n} a_{ij} x_j^p\right)$ and use Theorems 3.3(a) and 4.1.

Such systems of equations occur in the non-linear Leontief model of a closed exchange economy. The existence of a unique positive solution corresponds to the existence of a unique state of equilibrium in the economy (KARLIN [14] Chapter 8.7).

6. Applications to Positive Integral Operators

(i) Let $Ax(t) = \int_{0}^{1} K(t, s) x(s) ds$, where $K(\cdot, \cdot)$ is positive and continuous in the unit square $0 \le t, s \le 1$. Then

$$0 < \alpha \leq K(t,s) \leq \beta < \infty \qquad (0 \leq t, s \leq 1)$$

and hence

$$\alpha_{0}^{1} x(s) ds \leq A x \leq \beta_{0}^{1} x(s) ds$$
(6.1)

Hilbert's Metric

for every positive $x(\cdot)$ continuous in [0, 1]. If u(t)=1 in $0 \le t \le 1$, it follows from (6.1) that

$$d(Ax, u) \leq \log(\beta/\alpha)$$

and hence that $\Delta(A) \leq 2 \log(\beta/\alpha) < \infty$. Therefore, from Theorems 3.3(b) and 4.2, A has a unique positive continuous eigenfunction (with associated positive eigenvalue).

Inequalities such as (6.1) were used independently by BIRKHOFF in his theory of uniformly semi-primitive operators and by KRASNOSELSKII in his theory of u_0 -positive operators.

For the remainder of Jentzch's theorem proved by these methods see BIRKHOFF [4] or KRASNOSELSKII [16].

(ii) Let $Ax(t) = \int_{0}^{1} K(t, s) \{x(s)\}^{p} ds$, where $K(\cdot, \cdot)$ is non-negative and contin-

uous in the unit square $0 \le t$, $s \le 1$. Suppose that 0 and that for each fixed t in [0, 1], <math>K(t, s) is positive on a set of positive measure in $0 \le s \le 1$. Then, there exists a unique positive solution continuous in [0, 1] of the equation Ax = x.

Proof. Use Theorems 3.3(a) and 4.2.

Corollary 1. Under the same hypothesis on $K(\cdot, \cdot)$ and for q > 1, there exists a unique positive continuous solution of

$${x(t)}^q = \int_0^1 K(t, s) x(s) ds.$$

Corollary 2. Let A be a positive mapping in the cone of non-negative functions in C[0, 1]. Then the positive spectrum of A, that is, the set of eigenvalues with positive eigenvectors, is the interval $(0, \infty)$. To each λ in the positive spectrum there is a unique positive eigenfunction $x(\cdot, \lambda)$ and if $\lambda_1 < \lambda_2$ then $x(\cdot, \lambda_1) \ge x(\cdot, \lambda_2)$.

Proof. Let $x(t, \lambda) = \{x(t)/\lambda^{1/(1-p)}\}$, where $x(\cdot)$ is the unique fixed point of A in K.

Corollary 2 can be derived from KRASNOSELSKII's theory of u_0 -concave operators (see [16] Chapter 7).

7. Application to Positive Definite Symmetric Matrices

Let X be the space of real $n \times n$ matrices and K the cone of positive semidefinite matrices in X. Then \hat{K} is the set of positive definite matrices in X.

If $\mathscr{L}: \mathring{K} \to \mathring{K}$ is linear and q > 1, then there exist a unique $A \in \mathring{K}$ such that $\mathscr{L}(A) = A^{q}$.

Proof. Apply Theorems 3.3(a) and 4.3 to $F(A) = \{\mathscr{L}(A)\}^{1/4}$.

Corollary. If T is a real non-singular $n \times n$ matrix, there exists a unique $A \in K$ such that $T^*AT = A^2$. (See BUSHELL [11].)

8. Applications to Systems of Ordinary Differential Equations

The interested reader is referred to BIRKHOFF and KOTIN [6-8] and to BUSHELL [9, 12].

The author's work was supported in part by the United Kingdom Science Research Council Grant No. B/RG/28436.

References

- BAUER, F. L., An elementary proof of the Hopf inequality for positive operators. Numerische Math. 7, 331-337 (1965)
- 2. BAUER, F. L., E. DEUTSCH, & J. STOER, Abschätzungen für die Eigenwerte positiver linearer Operatoren. Linear Algebra and its Applications 2, 275–301 (1969)
- 3. BIRKHOFF, G., Extensions of Jentzch's Theorem. Trans. Amer. Math. Soc. 85, 219–227 (1957)
- BIRKHOFF, G., Uniformly semi-primitive multiplicative processes. Trans. Amer. Math. Soc. 104, 37-51 (1962)
- 5. BIRKHOFF, G., Uniformly semi-primitive multiplicative processes II. J. Math. Mech. 14, 507-512 (1965)
- 6. BIRKHOFF, G., & L. KOTIN, Essentially positive systems of linear differential equations. Bull. Amer. Math. Soc. 71, 771-772 (1965)
- 7. BIRKHOFF, G., & L. KOTIN, Linear second order differential equations of positive type. J. d'Analyse Math. 18, 43-52 (1967)
- BIRKHOFF, G., & L. KOTIN, Third order positive cyclic systems of linear differential equations. J. Diff. Eq. 5, 182–196 (1969)
- 9. BUSHELL, P. J., On systems of linear ordinary differential equations and the uniqueness of monotone solutions. J. London Math. Soc. (2) 5, 235-239 (1972)
- BUSHELL, P. J., On the projective contraction ratio for positive linear mappings. J. London Math. Soc. (2) 6, 256-258 (1963)
- 11. BUSHELL, P. J., On solutions of the matrix equation $T'AT = A^2$. Linear Algebra and its Applications (to appear)
- 12. BUSHELL, P. J., On solutions in a cone of ordinary differential equations in an ordered Banach space (to appear)
- HILBERT, D., Über die gerade Linie als kürzeste Verbindung zweier Punkte. Math. Ann. 46, 91-96 (1895)
- KARLIN, S., Mathematical Methods and Theory in Games, Programming and Economics, Vol. I. Reading, Mass.: Addison-Wesley 1959
- 15. KLEIN, F., Ueber die sogenannte Nicht-Euklidische Geometrie. Math. Ann. 4, 573-625 (1871)
- KRASNOSELSKII, M. A., Positive Solutions of Operator Equations. Groningen, The Netherlands: Noordhoff 1964
- 17. OSTROWSKI, A. M., Positive matrices and functional analysis. Recent Advances in Matrix Theory. Madison: University of Wisconsin Press, 1964
- 18. SAMUELSON, H., On the Perron-Frobenius Theorem. Mich. Math. J. 4, 57-59 (1957)
- THOMPSON, A. C., On certain contraction mappings in a partially ordered vector space. Proc. Amer. Math. Soc. 14, 438-443 (1963)

Mathematics Division University of Sussex Falmer, Brighton

(Received July 25, 1973)