PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 76, Number 1, August 1979

NEAR-DERIVATIONS AND INFORMATION FUNCTIONS[1]

JOHN LAWRENCE[2], GEOFF MESS AND FRANK ZORZITTO[3]

ABSTRACT. Near-derivations γ satisfy the conditions $\gamma(xy) = x\gamma(y) + y\gamma(x)$ for $x,y \in R$, $\gamma(x+y) \geqslant \gamma(x) + \gamma(y)$ for $x,y \geqslant 0$, $\gamma(x) = 0$ for $x \in Q$. Existence of near-derivations other than derivations is tied in with that of nonnegative information functions and an example of Daróczy and Maksa. Conditions for near-derivations to be derivations are discussed.

1. Introduction. A *near-derivation* is a self-mapping γ of the reals R such that:

$$\gamma(xy) = x\gamma(y) + y\gamma(x) \quad \text{for all } x,y \text{ in } R, \tag{1}$$

$$\gamma(x+y) \geqslant \gamma(x) + \gamma(y) \quad \text{for all } x,y \geqslant 0, \tag{2}$$

$$\gamma(x) = 0 \quad \text{for all rational } x. \tag{3}$$

The question immediately arises as to whether there are any near-derivations that are not already derivations.

An *information function* is a mapping f: $[0, 1] \to R$ satisfying the equation

$$f(x) + (1-x)f(y/(1-x)) = f(y) + (1-y)f(x/(1-y)),$$

and the normalizing conditions

$$f(0) = f(1) = 0, \qquad f\left(\tfrac{1}{2}\right) = 1.$$

The most important example of an information function is the Shannon function s given by

$$s(x) = -x\log x - (1-x)\log(1-x) \quad \text{for } 0 \leqslant x \leqslant 1.$$

(Here and throughout we adopt the conventions that $\log \equiv \log_2$ and $0 \log 0 = 0$.) The motivation for the above definition of an information function is developed by Aczél and Daróczy in [1], where such functions are studied extensively as part of an analysis of measures of information.

The present research was motivated by a problem cited in [1]. Namely, is the Shannon function s the only nonnegative information function? This was recently solved by Daróczy and Maksa [2]. They showed that for any

Received by the editors October 2, 1978.

AMS (MOS) subject classifications (1970). Primary 39A15; Secondary 94A15.

[1]This research was carried out partly in the NRC Summer Research workshops held at the University of Waterloo in 1978.

[2]Research supported by NRC grant A4540 and University of Waterloo research grant 131-7052.

[3]Research supported by NRC grant A4655 and University of Waterloo research grant 131-7067.

derivation δ on R, the formula

$$f(x) = s(x) + \delta(x)^2/x(1-x)$$

provides a nonnegative information function. In our paper we show that f is a nonnegative information function if and only if

$$f(x) = s(x) - \gamma(x) - \gamma(1-x),$$

where γ is a near-derivation. It follows from this that s is the minimal nonnegative information function. The latter fact is also proved in [2] by a different approach. In addition the Daróczy-Maksa example leads to examples of near-derivations which are not derivations. We conclude with some properties of near-derivations.

2. Connection of near-derivations to information functions. A theorem in [1] leads to the connection between information functions and near-derivations.

REPRESENTATION THEOREM (ACZÉL-DARÓCZY). *A function f: $[0, 1] \to R$ is an information function if and only if there is a self-mapping δ of R such that*

$$f(x) = -\delta(x) - \delta(1-x) \quad \textit{for all } x \textit{ in } [0, 1], \tag{4}$$

$$\delta(xy) = x\delta(y) + y\delta(x) \quad \textit{for all } x,y \textit{ in } R, \tag{5}$$

$$\delta(2) = 2. \tag{6}$$

In addition it can be readily checked that f is nonnegative if and only if the representing function δ satisfies

$$\delta(x+y) \geqslant \delta(x) + \delta(y) \quad \text{for all } x,y \geqslant 0. \tag{7}$$

Our first result is a refinement of the representation theorem for nonnegative f.

THEOREM 1. *A function f: $[0, 1] \to R$ is a nonnegative information function if and only if*

$$f(x) = s(x) - \gamma(x) - \gamma(1-x), \tag{8}$$

where γ is a near-derivation.

PROOF. Let γ be a near-derivation. Let σ: $R \to R$ be defined by $\sigma(x) = x \log|x|$, and let $\delta = \sigma + \gamma$. Clearly σ satisfies conditions (5)–(7). Since γ is a near-derivation, δ satisfies (5)–(7) also. By the representation theorem, the formula $f(x) = -\delta(x) - \delta(1-x)$, $0 \leqslant x \leqslant 1$, provides a nonnegative information function f.

Conversely, let f be a nonnegative information function. By the representation theorem let δ: $R \to R$ satisfy (4)–(7) for f. Let $\sigma(x) = x \log|x|$ as before, and let $\gamma = \delta - \sigma$. We shall prove γ is a near-derivation. It will then follow that

$$f(x) = -(\sigma+\gamma)(x) - (\sigma+\gamma)(1-x) = s(x) - \gamma(x) - \gamma(1-x)$$

as desired.

Now, property (1) of near-derivations holds for γ because δ and σ have it. Property (3) for γ follows from [3, Theorem 2], where it is shown that $\delta = \sigma$ on the rationals. Thus property (2) remains, namely that

$$\delta(x+y) - (x+y)\log(x+y) \geqslant (\delta(x) - x\log x) + (\delta(y) - y\log y) \tag{9}$$

for all $x,y \geqslant 0$. This inequality will follow from the subsequent lemmas.

LEMMA 2. *Suppose* $\delta\colon R \to R$ *satisfies conditions* (5)–(7). *Then*

$$\delta(1+x) \geqslant (1+x)\log(1+x) + \delta(x) - x\log x$$

for any $x \geqslant 0$.

PROOF. For any positive integer n and $x \geqslant 0$ we apply (5) and (7) to get

$$n(1+x)^{n-1}\delta(1+x) = \delta((1+x)^n) = \delta\left(\sum_{k=0}^{n}\binom{n}{k}x^k\right) \geqslant \sum_{k=0}^{n}\delta\left(\binom{n}{k}x^k\right).$$

Thus

$$\delta(1+x) \geqslant \frac{1}{n(1+x)^{n-1}}\sum_{k=0}^{n}\left(\binom{n}{k}\delta(x^k) + x^k\delta\left(\binom{n}{k}\right)\right).$$

In [3] it is proved that $\delta(r) = r\log r$ for any positive rational r. Hence

$$\begin{aligned}\delta(1+x) &\geqslant \frac{1}{n(1+x)^{n-1}}\sum_{k=0}^{n}\binom{n}{k}kx^{k-1}\delta(x) \\ &\quad + \frac{1}{n(1+x)^{n-1}}\sum_{k=0}^{n}x^k\binom{n}{k}\log\binom{n}{k} \\ &= \delta(x) + \frac{1+x}{n}\sum_{k=0}^{n}\binom{n}{k}\frac{x^k}{(1+x)^n}\log\binom{n}{k} \\ &= \delta(x) + a_n - b_n + c_n,\end{aligned}$$

where

$$a_n = \frac{1+x}{n}\sum_{k=0}^{n}\binom{n}{k}\frac{x^k}{(1+x)^n}\log\left(\binom{n}{k}\frac{x^k}{(1+x)^n}\right),$$

$$b_n = \frac{1+x}{n}\sum_{k=0}^{n}\binom{n}{k}\frac{x^k}{(1+x)^n}\log(x^k) = x\log x,$$

$$c_n = \frac{1+x}{n}\sum_{k=0}^{n}\binom{n}{k}\frac{x^k}{(1+x)^n}\log((1+x)^n) = (1+x)\log(1+x).$$

Since

$$\sum_{k=0}^{n}\binom{n}{k}\frac{x^k}{(1+x)^n} = 1$$

it follows from [**1**, Theorem 1.3.7] that

$$|a_n| \leqslant \frac{1+x}{n}\log(n+1).$$

Thus $a_n \to 0$ as $n \to \infty$. We therefore get in the limit

$$\delta(1+x) \geqslant \delta(x) - x\log x + (1+x)\log(1+x).$$

LEMMA 3. *For δ satisfying* (5)–(7) *and any $x,y \geqslant 0$, inequality* (9) *must hold.*

PROOF. By using property (5) and Lemma 2 we get

$$\begin{aligned}\delta(x+y) = \delta(x(1+y/x)) &= x\delta(1+y/x) + (1+y/x)\delta(x)\\ &\geqslant x((1+y/x)\log(1+y/x)) + \delta(y/x)\\ &\quad - (y/x)\log(y/x) + (1+y/x)\delta(x).\end{aligned}$$

Use of (5) to expand $\delta(y/x)$ and some logarithmic properties reduces the latter expression to

$$(x+y)\log(x+y) + \delta(x) - x\log x + \delta(y) - y\log y,$$

whence (9) follows.

THEOREM 4. *If f is any nonnegative information function and s is the Shannon function, then $f \geqslant s$.*

PROOF. Let δ represent f according to the representation theorem. By Lemma 3, inequality (9) holds for δ. With $0 \leqslant x \leqslant 1$ and $y = 1 - x$, (9) specializes to

$$\delta(x) + \delta(1-x) \leqslant x\log x + (1-x)\log(1-x),$$

which says that $f \geqslant s$.

THEOREM 5. *Not all near-derivations must be derivations.*

PROOF. Formula (8) yields an example of such near-derivations for every nonnegative information function f different from s.

3. Some properties of near-derivations. We are interested in conditions for the additivity of near-derivations. For example, it can be seen from [**3**] that if F, K are fields of real numbers and K is quadratic over F and γ is a near-derivation vanishing on F, then γ is additive on K. By Theorem 5 this need not happen if K is not algebraic over F. If, for a transcendental x, a near-derivation γ vanishing on F were linear on the field $F(x)$, then γ would be constant on the coset $x + F$. We are therefore led to examine γ restricted to $x + F$.

THEOREM 6. *Let F be a real field, x a real number and γ a near-derivation vanishing on F. Then for a,b in F:*

(a) $0 \leqslant x + a \leqslant x + b$ *implies that* $\gamma(x+a) \leqslant \gamma(x+b)$;

(b) $x + a \leqslant x + b \leqslant 0$ *implies that* $\gamma(x+a) \leqslant \gamma(x+b)$;

(c) $x + a \leqslant 0 \leqslant x + b$ *implies that* $\gamma(x+b) \leqslant \gamma(x+a)$;

(d) $\gamma(x + a)$ *and* $\gamma(x - a)$ *converge to the same finite value as* a *tends to* $+\infty$ *inside* F.

PROOF. To prove (a) we use (2) and the vanishing of γ on F. This yields:

$$\gamma(x + b) = \gamma(x + a + b - a) \geqslant \gamma(x + a) + \gamma(b - a) = \gamma(x + a).$$

Items (b) and (c) similarly use (2) and the fact that $\gamma = 0$ on F, plus that near-derivations are odd functions. The existence of the finite limits

$$\beta_l(x) = \lim_{\substack{a \to +\infty \\ a \in F}} \gamma(x - a), \qquad \beta_r(x) = \lim_{\substack{a \to +\infty \\ a \in F}} \gamma(x + a)$$

follows from the monotonicity conditions (a) and (b), and the bounding condition (c). It also follows from (c) that $\beta_r(x) \leqslant \beta_l(x)$.

To get equality for β_r and β_l consider the finite quantity $\beta_l(x^2)$ and apply (1).

$$\begin{aligned}
\beta_l(x^2) &= \lim_{\substack{a \to +\infty \\ a \in F}} \gamma(x^2 - a) = \lim_{\substack{a \to +\infty \\ a \in F}} \gamma(x^2 - a^2) \\
&= \lim_{\substack{a \to +\infty \\ a \in F}} ((x + a)\gamma(x - a) + (x - a)\gamma(x + a)) \\
&= x\beta_l(x) + x\beta_r(x) + \lim_{\substack{a \to +\infty \\ a \in F}} a(\gamma(x - a) - \gamma(x + a)).
\end{aligned}$$

The limit of $a(\gamma(x - a) - \gamma(x + a))$ is now finite. This means $\gamma(x - a) - \gamma(x + a)$ tends to 0, and hence $\beta_l(x) = \beta_r(x)$.

The next result gives a condition for our γ to be a derivation.

THEOREM 7. *Let* γ *be a near-derivation, and let the function* β *be defined by* $\beta(x) = \lim_{a \to +\infty; a \in Q} \gamma(x + a)$, x *real. Then* β *is* Q*-linear, and, for all real* x, $\beta(x^2) \geqslant 2x\beta(x)$. *Furthermore,* β *is a derivation if and only if* $\beta = \gamma$.

PROOF. The superadditivity (2) of γ implies the same for β as follows:

$$\begin{aligned}
\beta(x + y) &= \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(x + y + a) = \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(x + a/2 + y + a/2) \\
&\geqslant \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(x + a/2) + \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(y + a/2) = \beta(x) + \beta(y).
\end{aligned}$$

The subadditivity of β comes from the equality of β and β_l in Theorem 6(d), and the fact that, for $u > v > 0$, $\gamma(u - v) \leqslant \gamma(u) + \gamma(-v)$. Indeed,

$$\begin{aligned}
\beta(x + y) &= \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma((x + 2a) - (-y + a)) \\
&\leqslant \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(x + 2a) + \lim_{\substack{a \to +\infty \\ a \in Q}} \gamma(y - a) = \beta(x) + \beta(y).
\end{aligned}$$

Next, for x real use the equality in the proof of Theorem 6(d) to obtain:

$$\beta(x^2) = 2x\beta(x) + \lim_{\substack{a \to +\infty \\ a \in Q}} a(\gamma(x - a) - \gamma(x + a)). \tag{10}$$

Theorem 6(c) implies that $\gamma(x - a) \geqslant \gamma(x + a)$ when a is large. Thus $\beta(x^2) \geqslant 2x\beta(x)$.

Now if $\beta = \gamma$, then both are derivations because β is additive and γ satisfies (1). Conversely suppose β is a derivation. The difference $\gamma_1 = \gamma - \beta$ is then a near-derivation for which Theorem 6 applies. In particular, for any $x \geqslant 0$, $\gamma_1(x + a)$ increases to a finite value $\beta_1(x)$ as $a \to +\infty$ through the rationals. By the definition of γ_1 it follows that $\beta_1(x) = 0$. Thus $\gamma_1(x) \leqslant 0$ for all $x \geqslant 0$. However, one can check that any function satisfying (1) and (3) must have a dense graph in R^2 or else vanish everywhere. We must conclude that $\gamma_1 = 0$, or $\beta = \gamma$.

A corollary of Theorem 7 seems worth noting.

THEOREM 8. A near-derivation γ is a derivation if and only if, for all x, $a(\gamma(x - a) - \gamma(x + a)) \to 0$ as $a \to +\infty$ through the rationals.

PROOF. By formula (10), $\beta(x^2) = 2x\beta(x)$. This makes β a derivation, and hence γ.

REFERENCES

1. J. Aczél and Z. Daróczy, *On measures of information and their characterizations*, Academic Press, New York, 1975.

2. Z. Daróczy and Gy. Maksa, Nonnegative information functions (to appear).

3. J. Lawrence, *The uniqueness of the nonnegative information function on algebraic extensions*, (to appear).

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA N2L 3G1