INFORMATION SCIENCES **36**, 231–247 (1985)

Fuzzy and Probability Vectors as Elements of a Vector Space*

HAIM SHVAYTSER

and

SHMUEL PELEG

Department of Computer Science, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

ABSTRACT

The problem of defining vector space operations on fuzzy and probability vectors is discussed. It is shown that such a definition is equivalent to choosing a 1-1 and onto mapping from the unit interval into the real axis. Although such a mapping cannot be continuous, it is suggested that under certain approximations a continuous mapping can be chosen. A characterization of some useful mappings with applications to image processing is also given.

1. INTRODUCTION

The concept of fuzzy sets as a generalization of crisp sets was introduced by Zadeh in [1]. Operations on fuzzy sets were defined as a generalization of operations on crisp sets. The definition of a fuzzy event as a fuzzy set in R^n was given by Zadeh in [2]. If one assumes a finite sample space of dimension n, a fuzzy event can be represented by a real valued vector of n coordinates (the membership function) in the unit interval. We refer to such a vector as a fuzzy vector.

DEFINITION. A fuzzy vector is an n valued real vector $x=(x_i)_{i=1,\dots,n}$ where

$$0\leqslant x_i\leqslant 1 \qquad \forall i.$$

It is sometimes convenient to normalize the coordinates of the vector so that their sum is 1. This enables in many cases a probabilistic interpretation of the fuzzy event. We refer to such vectors as probability vectors.

*This research has been supported by the Israel National Council for Research and Development, grant No. 2450.

 0020-0255/85/$03.30

DEFINITION. A probability vector is an n valued real vector $x=(x_i)_{i=1,\ldots,n}$ where

$$0 \leqslant x_i \leqslant 1 \quad \forall i \quad \text{and} \quad \sum_{i=1}^{n} x_i = 1.$$

Notice that while a fuzzy vector of n elements is from an n dimensional space, a similar probability vector is from a space of dimension $n-1$.

Several operations on fuzzy sets, which are generalizations of operations on crisp sets, were defined in [1]. These definitions enable the generalization of classical probability theory to fuzzy probability theory [2]. One possible disadvantage of the classical definitions is that some numerical analysis algorithms assume elements to be members of a vector space, with addition and scalar multiplication that obey several axioms. Classical definitions of fuzzy set operations do not obey these axioms. For example, if fuzzy addition is taken as the min or max operation, or the algebraic product or sum, one cannot define $-x$ for every fuzzy vector x, such that $x + -x = 0$.

We present in this paper a new definition of operations on fuzzy and probability vectors. Some properties of the suggested operations are discussed, and applications to image processing, where a picture is viewed as a fuzzy vector, are also given.

2. OPERATIONS ON FUZZY VECTORS

Instead of defining the operations of addition and scalar multiplication such that the fuzzy vectors are elements of a vector space, we make use of the fact that all finite vector spaces with the same dimension are isomorphic. This means that a definition of these operations on fuzzy vectors is equivalent to a definition of a *one to one* and *onto* transformation from the space of fuzzy vectors into R^n, where n is the dimension of the space of fuzzy vectors. Such transformations exist, because the domain and the range have the same cardinal number. However, they cannot be continuous, since the domain is not an open set, while the range is open. (The domain is not open because its coordinates are from closed intervals [0,1].) Without continuity some algorithms might converge in the R^n vector space induced by the mapping, but not in the $[0,1]^n$ domain of the fuzzy vectors.

To overcome the continuous mapping problem, a smaller set of fuzzy events, consisting only of those that do not include any crisp knowledge, are considered. We define them as strictly fuzzy.

DEFINITION. A *strictly fuzzy event* (*set*) is a fuzzy event (set) whose membership function is in the open interval $(0,1)$.

A *strict fuzzy vector* is a fuzzy vector whose coordinates are in the open interval $(0,1)$.

A *strict probability vector* is a probability vector whose coordinates are in the open interval $(0,1)$.

Although crisp events cannot be described by strict fuzzy vectors, they can still be described as a limit of strict fuzzy vectors. One can always have a strictly fuzzy event which is as close as needed to any crisp event.

REMARK. For simplicity, we omit the word "strict" when referring to fuzzy events, fuzzy vectors, and probability vectors. Unless otherwise stated, we always mean the strict versions of the above terms.

We return to the problem of finding a continuous mapping from the set of fuzzy vectors into R^n. Although any 1-1 and onto transformation from $(0,1)^n$ into R^n will suffice, we try to characterize some simple transformations with "meaning." We shall look for transformations that transform each coordinate separately, i.e. transformations such as

$$\Phi:(0,1)\to(-\infty,\infty).$$

We require the following properties:

(1) Φ must be 1-1, onto, and continuous.

(2) Property (1) implies that Φ must be monotonic. We require Φ to be monotonic increasing to ensure that the order between grades of membership is preserved.

(3) Φ is antisymmetric around 0.5, i.e., $\Phi(0.5-x)=-\Phi(x)$.

From (1), (2), and (3) it follows that

$$\Phi(0)=-\infty,\qquad \Phi(0.5)=0,\qquad \Phi(1)=\infty.$$

We introduce a new variable t related to the value of the membership function x by

$$t=\frac{x-0.5}{0.5}. \tag{1}$$

Thus t satisfies

$$-1<t<1.$$

The transformation Φ written in terms of t is

$$\Phi(x) = \Psi(t) = \Psi\left(\frac{x-0.5}{0.5}\right),$$

and we have

$$\Psi:(-1,1) \to (-\infty,\infty).$$

Ψ is 1-1, onto, continuous, monotonic increasing, and antisymmetric around 0, satisfying

$$\Psi(-1) = -\infty, \qquad \Psi(0) = 0, \qquad \Psi(1) = \infty.$$

From its antisymmetry around 0 it follows that Ψ is an odd function and its Taylor expansion is

$$\Psi(t) = \sum_{j=0}^{\infty} a_j t^{2j+1}.$$

Because $\Psi(1) = \infty$, there are infinitely many $a_j \neq 0$. To ensure convergence for $-1 < t < 1$ we take a_j to be a rational function of j. To ensure the monotonically increasing property we take $a_j \geqslant 0\ \forall j$. A typical graph of such a function is shown in Figure 1.

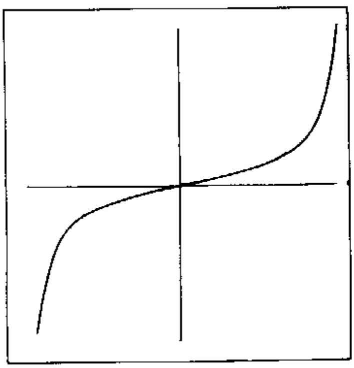

Fig. 1. Graph of a typical Ψ function.

EXAMPLES.

(a)

$$a_j = 1 \qquad \forall j,$$

$$\Psi(t) = t \sum_{j=0}^{\infty} (t^2)^j = \frac{t}{1-t^2},$$

$$\Psi^{-1}(s) = \frac{\sqrt{4s^2+1}-1}{2s}.$$

We call this transformation the *ratio transformation*.

(b)

$$a_j = \frac{1}{2j+1} \qquad \forall j,$$

$$\Psi(t) = \sum_{j=0}^{\infty} \frac{1}{2j+1} \cdot t^{2j+1} = \log\left(\frac{1+t}{1-t}\right) = 2\tanh^{-1}(t),$$

$$\Psi^{-1}(s) = \frac{e^s-1}{e^s+1} = \tanh(\tfrac{1}{2}s).$$

We call this transformation the *log-ratio transformation*.

(c) The tangent function is also such a transformation, but the expression for the a_j is more complicated.

Throughout the remainder of the paper we use the following convention: If x is a fuzzy vector, we define $\Psi(x)$ as the transformation Ψ applied to each coordinate of x [after applying the normalization substitution of Equation (1)].

2.1. *A DEFINITION OF OPERATIONS ON FUZZY VECTORS*

Let x, y be fuzzy vectors, and α a scalar. We define

$$x + y \equiv \Psi^{-1}(\Psi(x) + \Psi(y)), \tag{2}$$

$$\alpha \cdot x \equiv \Psi^{-1}(\alpha \cdot \Psi(x)). \tag{3}$$

We also have a definition for the inner product between two fuzzy vectors:

$$\langle x, y \rangle \equiv \sum_{i=1}^{n} \Psi(x_i) \cdot \Psi(y_i). \tag{4}$$

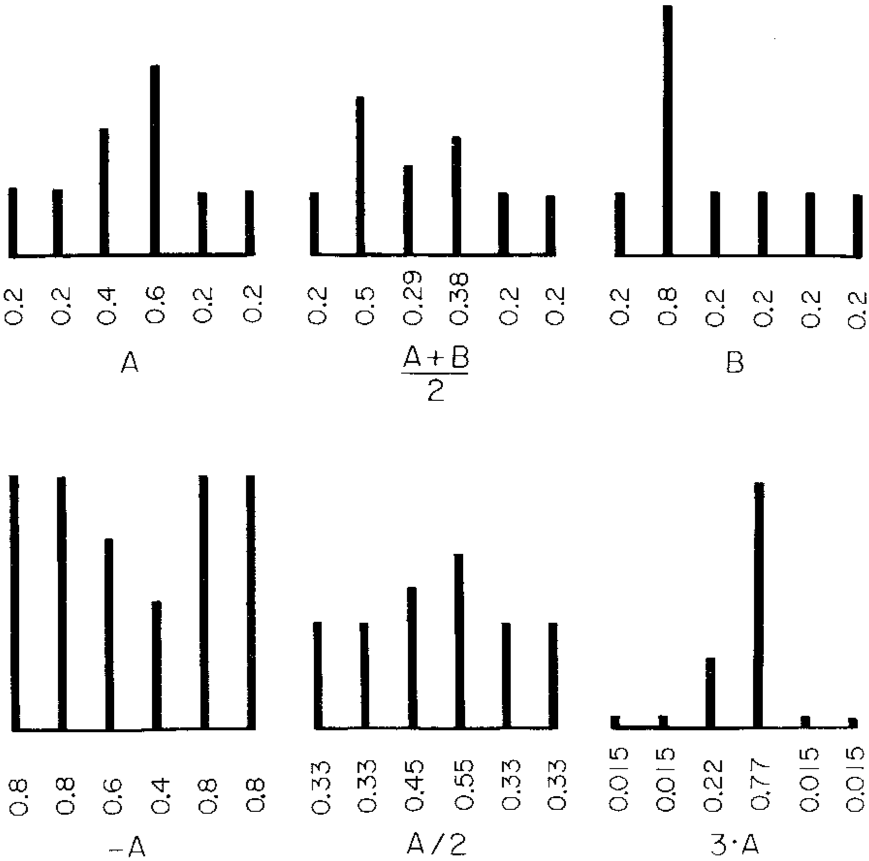

Fig. 2. Addition and multiplication in the log-ratio space for fuzzy vectors.

Examples of these operations for the log-ratio transformation are given in Figure 2.

The following properties are of interest:

(a) $-x = (-1)\cdot x = \Psi^{-1}(-\Psi(x))$. From the property (3) it follows that every value x_i of the membership function is transformed into $1 - x_i$, which is the classical definition of negation.

(b) The zero vector is the vector with all values equal to 0.5.

(c) The sum $A + B$ preserves information from both A and B.

(d) Intuitively, the operation $\alpha \cdot A$ makes the "information" conveyed in the vector "sharper" when $\alpha > 1$, and "blurs" this information when $0 < \alpha < 1$.

When $\alpha \to \infty$ every value less than 0.5 becomes 0, and every value greater than 0.5 becomes 1; the value 0.5 does not change.

(e) From the Taylor expansion it follows that Ψ is linear up to $O(t^3)$. This means that for values near 0.5 the new definition of addition is equivalent to classical algebraic addition.

3. OPERATIONS ON PROBABILITY VECTORS

Since an n valued probability vector is an element of an $n-1$ dimensional space, we have to look for a transformation from $(0,1)^n$ into R^{n-1} (for strict probability vectors). Although we could apply techniques similar to those of the previous section, we prefer to consider here one specific mapping.

3.1. *THE LOG-RATIO TRANSFORMATION FOR PROBABILITY VECTORS*

Given a probability vector

$$P = (p_1, p_2, \ldots, p_n), \qquad 0 < p_i < 1, \quad \sum_{i=1}^{n} p_i = 1,$$

we suggest the following transformation: For a certain $1 \leqslant r \leqslant n$ we define

$$q_i = \Psi(p_i) \equiv \log\left(\frac{p_i}{p_r}\right) \tag{5}$$

This is the log-ratio transformation for the probability vector case. Notice that different transformations are obtained for different values of r. However, we shall show that the resulting operations will be independent of r.

The inverse transformation is

$$p_i = \Psi^{-1}(q_i) = \frac{e^{q_i}}{\sum_{j=1}^{n} e^{q_j}}.$$

The log-ratio transformation can be viewed as a composition of two transformations. The first transfers the probability vector P into a vector of the corresponding ratios

$$\tilde{P} = \left(\frac{p_1}{p_r}, \frac{p_2}{p_r}, \ldots, \frac{p_n}{p_r}\right).$$

($p_r > 0$, since we deal with strict probability vectors.) The ratios in $\tilde{P}$ are in the range $(0, \infty)$, and a simple way of mapping them into $(-\infty, \infty)$ is to take the logarithm.

3.2. DEFINITIONS OF OPERATIONS

The operations of addition and scalar multiplication for probability vectors can now be defined in an analogous way to the definitions of these operations for fuzzy vectors in Equations (2) and (3) of Section 2.1:

Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, and $z = (z_1, \ldots, z_n)$ be probability vectors, and α a scalar. We define

$$z = x + y \quad \Leftrightarrow \quad z = \Psi^{-1}(\Psi(x) + \Psi(y)), \quad \text{i.e.,} \quad z_i = \frac{x_i \cdot y_i}{\sum_{j=1}^{n} x_j \cdot y_j}, \tag{6}$$

$$z = \alpha \cdot x \quad \Leftrightarrow \quad z = \Psi^{-1}(\alpha \cdot \Psi(x)), \qquad \text{i.e.,} \quad z_i = \frac{x_i^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}}. \tag{7}$$

The definition of inner product is a little more complicated. The inner product induced by the transformation (5) is

$$\langle x, y \rangle \equiv \sum_{i=1}^{n} \Psi(x_i) \cdot \Psi(y_i) = \sum_{i=1}^{n} \log\left(\frac{x_i}{x_r}\right) \cdot \log\left(\frac{y_i}{y_r}\right)$$

This has the disadvantage of being dependent on r. For different choices of r, different values of $\langle x, y \rangle$ can be obtained. This asymmetry can be corrected by using the fact that the sum of inner products is also an inner product. Using this, we have the following definition for inner product:

$$\langle x, y \rangle \equiv \sum_{r=1}^{n} \sum_{i=1}^{n} \log\left(\frac{x_i}{x_r}\right) \log\left(\frac{y_i}{y_r}\right) \tag{8}$$

Examples of these operations are given in Figure 3.

Some observations about the operations on probability vectors and their entropy are of interest:

(a) The zero vector is the vector with all values equal to $1/n$. This is the maximum entropy vector.

(b) The scalar multiplication $\alpha \cdot x$, where x is not the zero probability vector, has the property of decreasing entropy when $\alpha > 1$ and increasing entropy when $0 < \alpha < 1$. This property will be proved in the appendix.

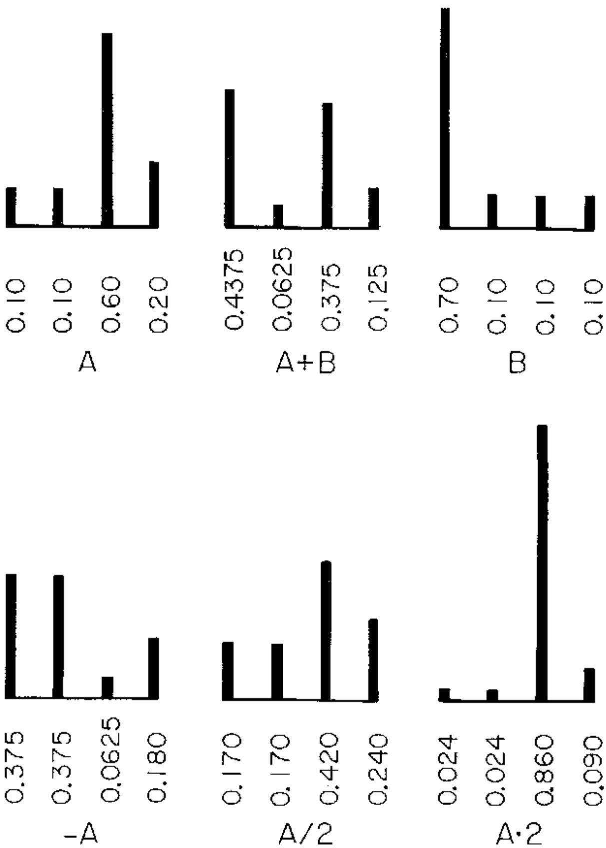

Fig. 3. Addition and multiplication in the log-ratio space for probability vectors.

3.3. *THE BAYESIAN FORMULA*

Using the operations defined in the previous section considerably simplifies some formulas from probability theory. As an example we consider the Bayesian formula.

Consider an event V and the set of mutually exclusive events $\{W_i\}$ whose union is the certain event. We introduce the following notation:

$$w_i = \text{Prob}(W_i),$$

$$q_i = \text{Prob}(W_i \mid V),$$

$$r_i = \text{Prob}(V \mid W_i).$$

Notice that $w = (w_1, w_2, \dots, w_n)$ and $q = (q_1, q_2, \dots, q_n)$ are probability vectors, but $r = (r_1, r_2, \dots, r_n)$ is not a probability vector, since $\sum_{i=1}^{n} r_i = \text{Prob}(V) \neq 1$.

The Bayesian formula for the above expression is

$$\text{Prob}(W_i \mid V) = \frac{\text{Prob}(V \mid W_i)\,\text{Prob}(W_i)}{\text{Prob}(V)} \qquad \forall i, \tag{9}$$

or

$$q_i = \frac{r_i w_i}{\sum_{i=1}^{n} r_i}.$$

We shall show that it reduces to simply

$$q = r + w, \tag{10}$$

using the addition operation of Section 3.2. Equation (10) is therefore to be understood as

$$\Psi(q_i) = \Psi(r_i) + \Psi(w_i) \qquad \forall i$$

Proof. From (9) we obtain

$$\begin{aligned}\frac{\text{Prob}(W_i \mid V)}{\text{Prob}(W_1 \mid V)} &= \frac{\text{Prob}(W_i)}{\text{Prob}(W_1)} \cdot \frac{\text{Prob}(V \mid W_i)}{\text{Prob}(V)} \cdot \frac{\text{Prob}(W_1)}{\text{Prob}(W_1 \mid V)} \\ &= \frac{\text{Prob}(W_i)}{\text{Prob}(W_1)} \cdot \frac{\text{Prob}(V \mid W_i)}{\text{Prob}(V)} \cdot \frac{\text{Prob}(V)}{\text{Prob}(V \mid W_1)} \\ &= \frac{\text{Prob}(W_i)}{\text{Prob}(W_1)} \cdot \frac{\text{Prob}(V \mid W_i)}{\text{Prob}(V \mid W_1)},\end{aligned}$$

i.e.,

$$\frac{q_i}{q_1} = \frac{w_i}{w_1} \cdot \frac{r_i}{r_1},$$

and hence

$$\log \frac{q_i}{q_1} = \log \frac{w_i}{w_1} + \log \frac{r_i}{r_1}. \qquad \blacksquare$$

4. APPLICATIONS TO IMAGE PROCESSING

In this section we briefly describe how the ideas of the previous sections can help to solve equations involving digital pictures. A broader treatment of this topic can be found in [3].

4.1. PICTURES AS FUZZY VECTORS

A digital picture is usually represented by a matrix of discrete values called grey levels, representing the brightness of each point. Brightness is always positive with no physical limit to its magnitude, but in computer representation grey levels are usually normalized to some finite range. With a range of [0,1] a picture can be viewed as a fuzzy vector, where the grey level represents the grade of membership in the set of black (or white) points.

4.2. PICTURES AS ELEMENTS OF A VECTOR SPACE

Defining operations on the set of pictures such that this set is a linear vector space, and possible applications to image processing, were discussed by the authors in [4]. One of the methods suggested there is the method described in Section 2. Examples of the resulting operations on pictures in the log-ratio space are shown in Figure 4.

4.3. EQUATIONS WITH PICTURES AS UNKNOWNS

In many image processing problems it is necessary to solve equations with pictures as unknowns. An important example is removing blur when a picture is blurred by a known operator. For example, if a picture is taken while the camera moves with a horizontal uniform motion, the blurring operation replaces each value in the matrix representation of the picture by the average of some of its horizontal neighbors.

The general blurring problem can be described by the following formula:

$$y = f(x),$$

where x is the original picture, f is the blurring operator, and y is the blurred picture. Given the blurred picture y and the blurring operator f, we try to find the original picture x. To do so, the operator f has to be inverted. This can be done using Fourier techniques, but the results using algebraic numerical algorithms like the conjugate-gradient algorithm [5] are usually better. These

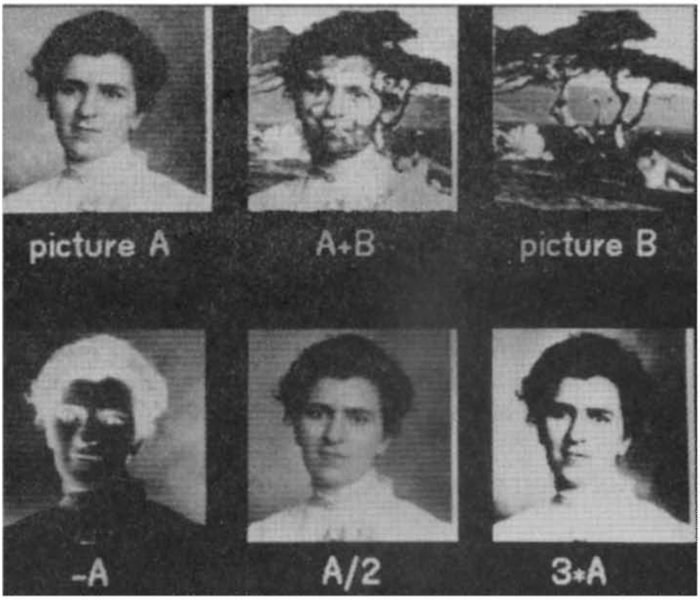

Fig. 4. Addition and multiplication in the log-ratio space for pictures.

algorithms, as well as the Fourier techniques, assume that the pictures are elements of a vector space, and hence produce a picture matrix with values not necessarily in the range [0,1]. Using the definitions for vector space operations of Section 2, these algorithms do not result in out of range values, and the picture solutions are better [3]. An example of a picture blurred by uniform motion blur, and the restoration of the picture using the iterative conjugate-gradient algorithm, is shown in Figure 5.

APPENDIX

In this appendix we investigate the relationship between α and the entropy of $\alpha \cdot x$, where x is a probability vector and the multiplication is as defined by Equation (7).

Let x be a probability vector, i.e.

$$x = (x_1, \ldots, x_n) \qquad 0 < x_i < 1 \quad \sum_{i=1}^{n} x_i = 1.$$

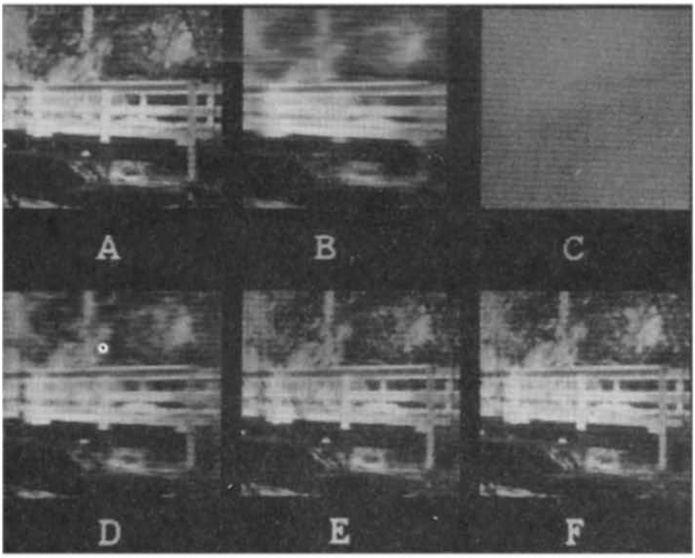

Fig. 5. Restoration of motion blurred picture in the log-ratio space: A, original picture; B, blurred picture; C, initial guess; D, after 5 iterations; E, after 10 iterations; F, after 20 iterations.

Then its entropy is

$$H(x) \equiv -\sum_{i=1}^{n} x_i \log x_i .$$

We are interested in

$$f(\alpha) = H(\alpha \cdot x).$$

From (7) we know that the ith coordinate of $\alpha \cdot x$ is

$$\frac{x_i^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}}$$

and hence

$$f(\alpha) = H(\alpha \cdot x) = -\sum_{i=1}^{n} \frac{x_i^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}} \log \frac{x_i^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}} \qquad (11)$$

The following observations are immediate:

0. If $x \equiv 0$, i.e. $x_i = 1/n\ \forall i$, then $f(\alpha) \equiv f(0)$.

1. $f(\alpha) \geqslant 0$.

2. $f(\alpha)$ has a global maximum at $\alpha = 0$.

This follows because by substituting $\alpha = 0$ in Equation (11) we obtain $f(0) = \log n$, and it is known that this is a maximum for an entropy function [6].

3. Although $f(-\alpha) \neq f(\alpha)$ in general, it is enough to investigate for $\alpha \geqslant 0$, since

$$f(-\alpha) = H(-\alpha \cdot x) = H(\alpha \cdot (-x)) = H(\alpha \cdot y),$$

where $y = -x$ is also a probability vector.

We claim the following:

4.

$$\lim_{\alpha \to \infty} f(\alpha) = \log(\#\max), \qquad \lim_{\alpha \to -\infty} f(\alpha) = \log(\#\min),$$

where #max(# min) is the number of coordinates whose probability is maximal (minimal).

5. $\alpha = 0$ is the only logical extremum of $f(\alpha)$.

If x is not the zero vector, then from observations 1–5 it follows that $f(\alpha)$ has a plot such as appears in Figure 6.

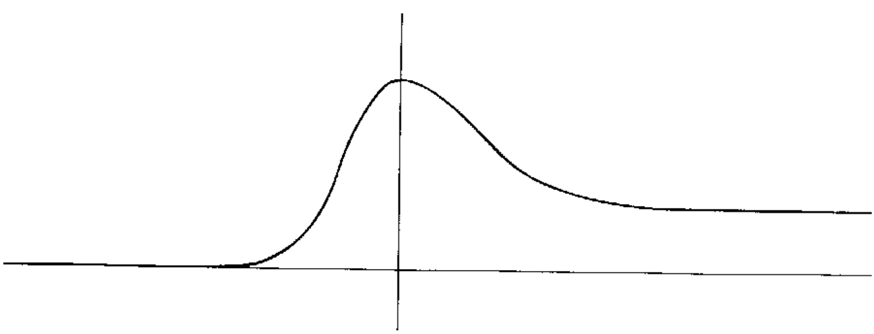

Fig. 6. A plot of $f(\alpha)$.

Proof of observation 4. Assume x_r is maximal. Then

$$Q(x_i) = \frac{x_i^\alpha}{\sum_{j=1}^n x_j^\alpha} = \frac{(x_i/x_r)^\alpha}{\sum_{j=1}^n (x_j/x_r)^\alpha}$$

and

$$\lim_{\alpha\to\infty} Q(x_i) = \begin{cases} \dfrac{1}{\#\max}, & x_i \text{ maximal},\\ 0 & \text{otherwise}, \end{cases}$$

and since

$$f(\alpha) = -\sum_{i=1}^n Q(x_i)\log Q(x_i),$$

it follows that

$$\lim_{n\to\infty} f(\alpha) = -\sum_{x_i = x_r} \frac{1}{\#\max}\log\frac{1}{\#\max} = -\log\frac{1}{\#\max} = \log(\#\max). \quad \blacksquare$$

The proof of the other part of observation 4 is similar and will be omitted.

Proof of observation 5. We prove that $f(\alpha)$ has no extremum points other than $\alpha = 0$. We have

$$f(\alpha) = -\sum_{i=1}^n \frac{x_i^\alpha}{\sum_{j=1}^n x_j^\alpha}\log\frac{x_i^\alpha}{\sum_{j=1}^n x_j^\alpha}$$

Let

$$y(\alpha) = \sum_{i=1}^n x_i^\alpha, \qquad y'(\alpha) = \sum_{i=1}^n x_i^\alpha \log x_i, \qquad y''(\alpha) = \sum_{i=1}^n x_i^\alpha\cdot(\log x_i)^2,$$

$$z(\alpha) = \log y, \qquad z'(\alpha) = \frac{y'}{y}, \qquad z''(\alpha) = \frac{y''y - (y')^2}{y^2}.$$

Then

$$f(\alpha) = \frac{1}{y}(\alpha y' - y\log y) = \alpha z' - z,$$

$$f'(\alpha) = \alpha z''.$$

If $\alpha = 0$ then $f'(\alpha) = 0$, so that $\alpha = 0$ is an extremum point. To prove that there are no other extremum points it is enough to show that $z'' > 0$, and since $y^2 > 0$ it is enough to show that

$$y''y > (y')^2,$$

or

$$\sum_{i=1}^{n} x_i^\alpha (\log x_i)^2 \sum_{j=1}^{n} x_j^\alpha > \left(\sum_{i=1}^{n} x_i^\alpha \log x_i \right)^2. \tag{12}$$

We shall show that this inequality follows from the Cauchy-Schwartz inequality:

$$\left(\sum_{i=1}^{n} a_i b_i \right)^2 \leqslant \sum_{i=1}^{n} a_i^2 \sum_{j=1}^{n} b_j^2 \tag{13}$$

with equality if and only if

$$\frac{a_i}{b_i} = \frac{a_j}{b_j} \qquad \forall i, j.$$

Let

$$a_i = \sqrt{x_i^\alpha}, \qquad b_i = (\log x_i)\sqrt{x_i^\alpha}$$

Then the inequality (12) is reduced to the inequality (13). Since $b_i / a_i = \log x_i$, it follows that the case of equality in (12) is when $x_i = x_j$ $\forall i, j$, and then x is the zero vector. ■

The authors would like to thank J. Gonczarowski and A. Rosenfeld for going over the manuscript, and N. Linial for helpful suggestions, especially concerning the proof in the appendix.

REFERENCES

1. L. A. Zadeh, Fuzzy sets, *Inform. and Control* 8:338–353 (1965).
2. L. A. Zadeh, Probability measures of fuzzy events, *J. Math. Anal. Appl.* 23:421–427 (1968).
3. H. Shvaytser and S. Peleg, Inversion of picture operators, *proceedings* at 1983 IEEE International Conference on Systems Man and Cybernetics, Bombay and New Delhi, Dec. 1983, pp. 202–206.

4. H. Shvaytser and S. Peleg, Pictures as elements in a vector space in, *IEEE Conference on Computer Vision and Pattern Recognition*, Arlington, Va., June 1983, pp. 222–230.
5. M. Hestenes, *Conjugate Direction Methods in Optimization*, Springer, 1980.
6. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed., McGraw-Hill, New York, 1984.

Received 5 March 1985; revised 2 April 1985