JOURNAL OF COMPUTER AND SYSTEM SCIENCES: 5, 26–40 (1971)

Realizations by Stochastic Finite Automata

J. W. CARLYLE*

Department of System Science, University of California, Los Angeles, California 90024

AND

A. PAZ†

Department of Computer Science, Technion—Israel Institute of Technology, Haifa, Israel

Received August 5, 1970

It is shown that a real-valued function $f(x)$, defined for strings x over a finite alphabet, is of the form $(\beta g(x) + \gamma) \exp(\delta | x |)$ for constants β, γ, δ, and the acceptance probability function g for a probabilistic automaton, if and only if f is of finite rank, where the latter external criterion is equivalent to the internal realizability of f by a finite-state sequential system permitted to have arbitrary real initial, transition, and output weights. The development encompasses multiple numerical outputs (finite vectors of functions of strings) and the corresponding generalization of this theorem; as an intermediate step, a set of sufficient conditions is established for equivalence of sequential systems (ss) with multiple outputs, yielding procedures for conversion of ss to numerical-output probabilistic automata (npa). Additional instances are given of application of these ideas in constructing npa equivalent to certain ss.

1. INTRODUCTION: AUTOMATA, FUNCTIONS, RANK

Let Σ be a finite set (the input alphabet), and let Σ^* be the set of all Σ-strings (finite sequences of elements of Σ) including the empty string Λ. As usual, if x and y are strings, then xy denotes concatenation (x followed by y), and $| x |$ is the length of x (so $| xy | = | x | + | y |$ and $| \Lambda | = 0$). Let $Z = \{1, 2,..., b\}$ (the alphabet of output indices). In what follows, f denotes a real-valued function with domain $Z \times \Sigma^*$, the values of f are written as $f_i(x)$ for $i \in Z$ and $x \in \Sigma^*$, and $[f]_n$ is the *n-segment* of f, i.e., the restriction of f to $| x | \leqslant n$.

* Research supported in part by the National Science Foundation under Grant GJ-803.

† Research supported in part by the Applied Mathematics Division, Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant 68-1408 at the University of California, Los Angeles.

Let c be a positive integer. By a *c-state numerical-output probabilistic automaton* (npa_c) we mean a system

$$A = (\pi, \{A(\sigma)\}, \{\eta_i\}, \tag{1}$$

where π is a c-component stochastic row vector (called the *initial vector*), $A(\sigma)$ is a $c \times c$ stochastic matrix (a *transition matrix*) for each $\sigma \in \Sigma$, and η_i is a real column c-vector (*final vector*) for each $i \in Z$. If A is a npa, the *(external) function* f *of* A is defined by

$$f_i(x) = \pi A(x) \eta_i , \tag{2}$$

where $A(x) = A(\sigma_1) A(\sigma_2) \cdots A(\sigma_n)$ when $x = \sigma_1\sigma_2 \cdots \sigma_n$, and $A(\Lambda)$ is the $c \times c$ identity. If $b = 1$ and the components of $\eta = \eta_1$ are restricted to be 0 and 1 only, we have the usual probabilistic automaton (pa) [1] with initial distribution π on the state set and final vector η whose "1" components define the set of accepting final states; the external function $f(x) = f_1(x)$ is then the probability of acceptance of word x. When $b = 1$ and η has arbitrary real components, $f(x)$ is the expected final output if, for $k = 1, 2,..., c$, the k-th component of η is interpreted as a numerical-valued output, cost, or return at state k; these npa were introduced by Page [2]. When $b > 1$ and the η_i are any real vectors, our definition of npa covers the case of multiple patterns of numerical outputs and $f_i(x)$ is the expected value for the ith pattern after string x has been read by the automaton. For example, if $b = 2$, η_1 has 0 and 1 components only, and η_2 is a real vector, we might interpret $f_1(x)$ as the probability of acceptance of x and $f_2(x)$ as the expected payoff at the final state after reading x.

If the npa A is such that all vectors η_i have only 0 and 1 component values, we may then interpret the b vectors $\eta_i = (\eta_{ij} : 1 \leqslant j \leqslant c)$ to be indicator functions of b sets of accepting states—not necessarily disjoint, encompassing acceptance–rejection by b attributes which may overlap. Alternatively, in this situation we may imagine the system A to have b binary output lines, with $\eta_{ij} = 1$ if and only if line i is excited when the system is in state j; at a given state several lines may be simultaneously excited. Here we shall say that A is a true probabilistic automaton (pa) if the accepting sets are disjoint, or equivalently if at most one line is excited at each state (i.e., for each j at most one η_{ij} is 1); pa are actually coextensive with npa in an appropriate sense (see Appendix I).

If f is a function on $Z \times \Sigma^*$ such that there exists an npa_c A whose external function f^A is identical with f, we say that f is *realizable by npa*, and that f is npa-realizable with (at most) c states, and in particular that f is realized by A, or that A is a c-state realization of f. One formulation of a problem with which we are concerned here can be put in the following general way: Find and characterize classes of f realizable by npa, or pa, with effectively calculable realizations based upon f. The problem is stated for finite automata or "acceptors" but has its analog for "transducers" (stochastic sequential machines or finite-state channels) and functions on $(Z \times \Sigma)^*$ [3, 4] and methods introduced for the latter case are employed here as a point of departure. However, the two situations are distinct in several aspects [5] (in particular, it is clear that probabilistic acceptors are, in an appropriate sense, reductions or specializations of transducers); the results obtained here are more definitive than those known for transducers, and do not appear to be transferable immediately to the transducer case. In a previous report [8], we noted without details preliminary versions of some of the results given here.

The principal concept which we adapt from the transducer case is that of the rank of an external function [3]. For a function f on $Z \times \Sigma^*$, we define the *rank* of f to be the largest integer $r = r(f)$ such that at least one $r \times r$ matrix of the form

$$F = (f_{jk}) = (f_{i_k}(x_j y_k)) \tag{3}$$

is nonsingular (for some choice of strings $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_r$, and integers $i_1, i_2, \ldots, i_r$ in Z): If no such finite r exists, we set $r(f) = \infty$. Our approach to the realization problem is to show that f is of finite rank if an only if it is realizable by certain pseudo-npa, called sequential systems here, and then to convert the latter to npa in various cases. In what follows, for a given f, matrices of the form (3) and their determinants will be called f-matrices and f-determinants.

2. Realization by Sequential Systems

A c-state *sequential system* (ss_c) A of the form (1), and its external function (2), are defined as in the case of npa_c, but with entries in π and $A(\sigma)$, as well as η_i, permitted to be unrestricted real numbers. For $b = 1$, ss are the generalized automata of Turakainen [6]; in the transducer case we have the corresponding concept of a pseudomachine [3, 4], and some of the arguments below parallel those for the pseudomachine case.

Two ss, with possibly different numbers of states, are said to be *equivalent* ($\sim$) if they have the same external function, or *n-equivalent* ($\overset{n}{\sim}$) if they have the same n-segment. In these definitions of equivalence and in what follows, we tacitly assume that the alphabet Z of output indices of one of the ss may be imagined to have been permuted or relabeled whenever needed to bring the two external functions into component–component coincidence; thus, we do not explicitly mention trivial "nonequivalences" due solely to component permutations.

A sequential system A_2 is a *minimal form* of the ss A_1 if $A_2 \sim A_1$ and A_2 has the minimum number of states among ss equivalent to A_1; A_1 is *in minimal form* if there is no equivalent A_2 with fewer states.

By the *rank* $r(A)$ of a sequential system A, we mean the rank of the external function for A.

Given a ss_c $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$ with external function f, we introduce the c-component row vectors $\pi(x) = \pi A(x)$ and column vectors $\eta_i(x) = A(x)\,\eta_i$; then

$$f_i(xy) = \pi(x)\,\eta_i(y), \tag{4}$$

$$\pi(xy) = \pi(x)\,A(y), \tag{5}$$

$$\eta_i(xy) = A(x)\,\eta_i(y) \tag{6}$$

for all strings x and y, with $\pi(\Lambda) = \pi$ and $\eta_i(\Lambda) = \eta_i$. If F is any $n \times n$ f-matrix, then using (4) we see that F can be factored in the form $F = GH$, where the rows of G are $\pi(x_j)$ and the columns of H are $n_{i_k}(y_k)$; since G is $n \times c$ and H is $c \times n$, we conclude that the rank of A must be finite, and is in fact no greater than the number c of states of the ss. The converse also holds: If f is any function of finite rank, then f is ss-realizable, as the following argument shows.

Let f be a function on $Z \times \Sigma^*$ having finite rank r, and let F be an $r \times r$ nonsingular f-matrix as in (3). Let F be bordered with an $(r + 1)$-st row and column to form an $(r + 1) \times (r + 1)$ f-matrix; the determinant must vanish, yielding

$$f_i(xy) = \sum_{j=1}^{r} m_j(x) f_i(x_j y), \tag{7}$$

with $x = x_{r+1}$, $y = y_{r+1}$, $i = i_{r+1}$, and $m_j(x) = (-1)^{r+j}\,|F_j|/|F|$, where F_j is the f-matrix obtained from F by replacing x_j with x.

Let $F(x)$ and $M(x)$ be the $r \times r$ matrices whose (j, k) elements are $f_{i_k}(x_j x y_k)$ and $m_k(x_j x)$, respectively. Then $F(\Lambda) = F$, and using (7) we see that $M(\Lambda) = I$ ($r \times r$ identity) and

$$F(x) = M(x)F \qquad \text{or} \qquad M(x) = F(x)F^{-1}, \tag{8}$$

$$F(xy) = M(x)F(y), \qquad M(xy) = M(x)\,M(y). \tag{9}$$

Let m be the row vector whose jth component is $m_j(\Lambda)$, and let $\xi_i(x)$ be the column vector whose jth component is $f_i(x_j x)$; then, again using (7), we obtain

$$\xi_i(xy) = M(x)\,\xi_i(y), \tag{10}$$

$$\xi_i(x) = M(x)\,\xi_i\,, \qquad \xi_i = \xi_i(\Lambda) = (f_i(x_j)), \tag{11}$$

$$f_i(x) = m\xi_i(x). \tag{12}$$

Now define a ss_r $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$ by setting

$$\pi = m, \qquad A(\sigma) = M(\sigma), \qquad \eta_i = \xi_i\,. \tag{13}$$

Then $A(x) = M(x)$ and, using (11) and (12), we see that the external function of the ss A is

$$\pi A(x)\,\eta_i = mM(x)\,\xi_i = f_i(x), \tag{14}$$

so that f is realizable by A. In fact, a realization with a true stochastic initial vector can be obtained, as the following argument shows. Let Q be any nonsingular matrix and replace (13) with

$$\pi = mQ^{-1}, \qquad A(\sigma) = QM(\sigma)\,Q^{-1}, \qquad \eta_i = Q\xi_i \tag{13'}$$

and (14) again holds. Clearly, we may take Q to be such that mQ^{-1} is a probability distribution; in particular, we may require Q to have m as its first row, so that $\pi = (1, 0, 0,..., 0)$. (See Appendix II for a further remark on (13').)

The results of the preceding paragraphs can be stated as theorems in various ways; we choose the following overall summary, in which for clarity some separate but equivalent statements are made regarding functions and ss, respectively.

THEOREM 1. *A function is of finite rank if and only if it is* ss-*realizable. If f is a function of finite rank r, every* ss-*realization of f must have at least r states, and there exist realizations with exactly r states. The rank of any sequential system is finite and cannot exceed the number of states of the system. The minimal forms of any* ss *A have $r(A)$ states, and for a given* ss, *a minimal form can be constructed effectively. Realizations can be assumed to have true stochastic initial vectors, and indeed can be supposed to be concentrated on a single state.*

In Theorem 1, by the effective construction of realizations we mean that a finite terminating set of calculations is involved, based upon a given finite segment of the external function. In fact, if f is of finite rank r it is finitely specifiable externally in the sense that only its $(2r - 1)$-segment need be given, and then a realization can be constructed; further details bearing on this point appear in Appendix II.

3. TRANSFORMATION OF ss TO npa

At this point, we turn our attention to the problem of finding true npa equivalent to ss; i.e., for a given ss which may have arbitrary entries in its initial vector and transition matrices, we seek (when possible) an equivalent system with nonnegative entries such that the initial vector and all rows of each transition matrix sum to unity. Our approach is based on the following sufficient condition for equivalence of two ss. (A necessary and sufficient condition along the same lines is mentioned in Appendix II.)

THEOREM 2. *Let A and A' be* ss *with c and c' states, respectively. If there exists a $c' \times c$ matrix B and a real constant $K \neq 0$ such that*

(α) $\pi' B = K\pi$,

(β) $A'(\sigma)B = BA(\sigma)$ *for each* $\sigma \in \Sigma$,

(γ) $\eta_i' = B\eta_i/K$ *for each* $i \in Z$,

then A and A' are equivalent.

The proof of Theorem 2 is immediate. First we note that property (β) extends directly from single symbols σ to arbitrary strings x. Then we have the following: If $K = 1$ (otherwise repeat with B replaced by B/K):

$$f_i(x) = \pi A(x)\,\eta_i = \pi' BA(x)\,\eta_i = \pi' A'(x)\,B\eta_i = \pi' A'(x)\,\eta_i' = f_i'(x).$$

The application of Theorem 2, which is of primary significance here, is that of seeking npa equivalent to given ss. For such purposes, we regard the conditions (α), (β), and (γ) of the theorem as equations with one of the ss known and the other required to satisfy npa properties. With this in mind, we introduce a geometrical interpretation of the conditions of the theorem. Assume that the system A in the theorem is given and consider the conditions as equations to be solved, if possible, for an unknown system A' subject to the restriction that A' be a true npa. Since there is no restriction on the entries in the vectors η_i' (according to the definition of npa), condition (γ) can be taken to be the definition of the vectors η_i', after (α) and (β) have been solved. We rephrase (α) and (β) geometrically as follows: If X is a matrix, let $C(X)$ denote the convex hull of the rows of X, i.e., the set of all vectors expressible as convex combinations of the rows of X. Then the equation (β) can be solved for a stochastic matrix $A'(\sigma)$ with given B if an only if $C(BA(\sigma)) \subset C(B)$, for in this and only this case each row of $BA(\sigma)$ can be expressed as a convex combination of the rows of B and the stochastic vector whose entries are the coefficients of the combination will be the corresponding row of $A'(\sigma)$. Similarly, the equation (α) is equivalent to the condition $K\pi \in C(B)$. We have thus proved the following:

COROLLARY 1. *Let $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$ be a given c-state* ss. *A sufficient condition for the existence of a d-state* npa *A' equivalent to A is the existence of a $d \times c$ matrix B and a constant $K \neq 0$ such that*

(α) $K\pi \in C(B)$,

(β) $C(BA(\sigma)) \subset C(B)$ *for each* $\sigma \in \Sigma$.

Then $A' = (\pi', \{A'(\sigma)\}, \{\eta_i'\})$ can be constructed by setting $\eta_i' = B\eta_i/K$ and taking π' and $A'(\sigma)$ to be (any) solutions of the equations

$$\pi' B = K\pi, \qquad A'(\sigma)B = BA(\sigma),$$

subject to the constraints that π' is a probability vector and the $A'(\sigma)$ are Markov matrices.

Corollary 2. *Let A be a given c-state* ss, *and suppose that there exists a $d \times c$ matrix B such that*

(α') *$C(B)$ is c-dimensional and the origin is an interior point,*

(β) *$C(BA(\sigma)) \subset C(B)$ for each $\sigma \in \Sigma$.*

Then there exist d-state npa *equivalent to A, and such a* npa *A' can be constructed by selecting a constant $K \neq 0$ such that $K\pi \in C(B)$, setting $\eta_i' = B\eta_i/K$, and taking π' and $A'(\sigma)$ to be stochastic solutions of the equations $\pi' B = K\pi$, $A'(\sigma)B = BA(\sigma)$.*

Proof. According to (α'), $C(B)$ contains a c-sphere of positive radius about the origin; clearly, then, given any arbitrary initial vector π for A, there corresponds some constant $K \neq 0$ such that $K\pi$ lies in this sphere and hence is in $C(B)$, so condition (α) of Corollary 1 is satisfied. The construction then follows from Corollary 1, using this value of K.

In Corollary 2, condition (α') implies that $C(B)$ must be generated by at least $c + 1$ convexly independent c-vectors; i.e., the number d of states in the npa realization is at least $c + 1$. Corollary 1 and Theorem 2 do not explicitly require $d > c$; nevertheless, in passing from ss to npa (when possible) one expects that the state set size will typically be enlarged, and indeed that d may be much greater than c in applications.

By a direct application of Corollary 2, we show that a class of ss with certain absolute-value restrictions is npa realizable, as follows. Here, we let V^c be the set of points in c-dimensional space having the property that the sum of absolute values of coordinates is $\leqslant 1$.

Theorem 3. *Let A be a c-state sequential system in which π and $\{\eta_i\}$ are arbitrary and all rows of each matrix $A(\sigma)$ are in V^c. Then there are* npa *equivalent to A; in particular, an equivalent* npa *A' can be constructed with $2c$ states.*

Proof. We apply Corollary 2, taking B to be the $2c \times c$ matrix whose rows are the unit vectors and their negatives:

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

Each row of $BA(\sigma)$ coincides with a row of $A(\sigma)$ or its negative; thus the rows of $BA(\sigma)$ must also be in V^c, so $C(BA(\sigma)) \subset V^c$. Since V^c is in fact identical with $C(B)$ (the rows of B are obviously the vertices of the convex polyhedron V^c), we conclude that condition (β) of Corollary 2 is satisfied. Since $0 \in V^c$, condition (α') is also satisfied. Thus an equivalent npa A' exists, and can be taken to have $2c$ states.

With the aid of Theorem 3 we next establish our main result on the characterization of external functions, viz., that if a certain multiplicative factor in the function is permitted, every ss is "npa-realizable" with that multiplier present; that is, the condition of finiteness of the rank of a function coincides with a form of npa realizability in the following sense.

THEOREM 4. *A function f is of finite rank if and only if*

$$f_i(x) = \alpha^{|x|} f_i^A(x)$$

for some npa A *and positive constant* α.

Proof. If A is given, with c states, the choice $\alpha = 1$ will of course yield $f = f^A$ of finite rank $\leqslant c$. Conversely, let f be given with finite rank r. Then we know that f is realizable by some r-state ss $\bar{A} = (\bar{\pi}, \{\bar{A}(\sigma)\}, \{\bar{\eta}_i\})$. Let α be a positive constant such that each of the matrices $(1/\alpha)\, \bar{A}(\sigma)$ satisfies the conditions of Theorem 3 (clearly such an α can always be found; it suffices to take α to be the maximum, over all rows of all matrices $\bar{A}(\sigma)$, of the row sum of absolute values). For the ss $\hat{A} = (\bar{\pi}, \{\bar{A}(\sigma)/\alpha\}, \{\eta_i\})$, consider the equivalent $2r$-state npa A' obtained from Theorem 3; since

$$f_i^{A'}(x) = f_i^{\hat{A}}(x) = \bar{\pi}(\bar{A}(x)/\alpha^{|x|})\bar{\eta}_i = f_i(x)/\alpha^{|x|},$$

the desired result follows by choosing A' to be the realizing npa.

COROLLARY 3. *A function f is of finite rank if and only if*

$$f_i(x) = \alpha^{|x|}(\beta f_i^A(x) + \gamma)$$

for some constants α, β, γ, and a true pa A.

Proof. According to Theorem 4, we need only establish that to each npa there corresponds a pa whose external function is linearly related to that of the npa. This is easily demonstrated by construction; details have been placed in Appendix I.

Theorem 4 and its Corollary demonstrate the close relationship between rank and realization. Although there exist functions which are not literally pa-realizable (known examples for functions of Markov chains can be adapted directly here), we see that realizability can always be effected through the introduction of factors which are at most dependent upon string length; that is, the only explicit function of string symbols needed is the external function of the realizing pa. In this sense, then, we can say that our results show that finiteness of rank and pa-realizability are essentially equivalent concepts.

We conclude with remarks on two typical examples of additional instances of application of Theorem 2 and its corollaries in obtaining realizations; the first relates to the operation of string reversal, and the second to eigenvalue properties, with an example accompanying the latter.

If f is a scalar external function (i.e., $b = 1$), the *reversal* of f is the function Rf defined by $Rf(x) = f(R(x))$, where $R(x)$ is the reverse of the string x; that is, $R(\sigma_1\sigma_2 \cdots \sigma_n) = \sigma_n\sigma_{n-1} \cdots \sigma_1$ and $R(\Lambda) = \Lambda$. It is known that class of scalar pa-realizable functions is closed under reversal [7]; a compact proof of this fact can be given with the aid of Corollary 1 as follows: Let the scalar function p be realized by a probabilistic automaton $A = (\pi, \{A(\sigma)\}, \eta)$; recall that in this case the sole η vector has 0 and 1 components. When $x = \sigma_1\sigma_2 \cdots \sigma_n$, $Rp(x) = p(R(x)) = \pi A(\sigma_n)\, A(\sigma_{n-1}) \cdots A(\sigma_1)\eta = \eta^T A^T(x)\, \pi^T$ (T denotes transpose) which is the external function of the ss $A^T = (\eta^T, \{A^T(\sigma)\}, \pi^T)$. Let B be a $2^c \times c$ matrix whose rows are the vertices of the unit cube, i.e., all vectors with 0 and 1 components. Since the vector η^T is already in this form by assumption, condition (α) of Corollary 1 is satisfied for the initial vector η^T of $A^T (K = 1)$. Condition (β) is also satisfied, since the columns of $A^T(\sigma)$ are stochastic, so each row of $BA^T(\sigma)$ is in the unit cube. Let A' be a 2^c-state npa equivalent to A^T, constructed with the aid of B as in Corollary 1. Since π is a stochastic vector, $\eta' = B\pi^T$ is a vector whose components lie between 0 and 1. Therefore, by an obvious state-splitting procedure (a special case of the construction in Appendix I), we can obtain a true pa A'' (η'' having 0 and 1 components) which is equivalent to A' (and has at most twice as many states as A'). (*Added in proof:* The authors are informed that a result related to Corollary 1 and its application to reversal was reported by H. Matuura, Y. Inagaki, and T. Hukumura in the record of the 1968 Nationl Convention, IECE Japan (in Japanese).)

In Theorem 2, condition (β) suggests that some of the classical properties of the matrix equation $A_1X = XA_2$ may be of use in obtaining realizations in various cases; for example, $X = 0$ is the only choice possible unless A_1 and A_2 share at least one eigenvalue. Another general observation about eigenvalues is that since in npa-realization applications A_1 is intended to be stochastic, its eigenvalues must follow the well-known pattern for stochastic matrices; some of these features may be interpreted in terms of properties of A_2 and $X(=B)$ through $A_1X = XA_2$ (see also the remark at the end of Appendix II). A more specific result involving eigenvalues is the following:

THEOREM 5. *Let $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$ be a c-state* ss, *and let λ be the largest among the eigenvalues of all $c \times c$ matrices of the form $A(\sigma)\, A^T(\sigma)$ for all $\sigma \in \Sigma$. If $\lambda < 1$, there exists a* npa *equivalent to A.*

Proof. Matrices of the form AA^T, being real and symmetric, have only real nonnegative eigenvalues so that λ is well-defined and nonnegative; moreover, the AA^T are evidently normal matrices, for which the matrix norm coincides with the largest eigenvalue, so we have

$$\max_{\sigma} \max_{\|v\|\neq 0} \frac{(vA(\sigma)A^T(\sigma), v)}{\| v \|^2} = \lambda$$

(where $\| \|$ and $(,)$ denote the usual Euclidean length and inner product in c dimensions), or

$$\| vA(\sigma)\|^2 \leqslant \lambda \| v \|^2$$

for all c-vectors v and all $\sigma \in \Sigma$. We assume that λ is positive (if $\lambda = 0$ the theorem follows trivially), and let Δ be a convex polyhedron whose surface circumscribes the unit sphere about the origin in c dimensions such that $\| v \|^2 \leqslant 1/\lambda$ for all $v \in \Delta$; since $1 < 1/\lambda < \infty$, this is evidently always possible if Δ is allowed to have sufficiently many vertices. Let B be a $d \times c$ matrix such that $C(B) = \Delta$, e.g., take the rows of B to be the vertices of Δ. Then for all $v \in C(B)$ we have $\| vA(\sigma) \|^2 \leqslant 1$, so vA is in the unit sphere and therefore in $C(B)$; thus, for this B, condition (β) of Corollary 2 is satisfied. Condition (α') of course also holds, so Corollary 2 yields a d-state npa realization for the ss A.

According to the method of proof of Theorem 5, one might use various classical results in the n-dimensional geometry of convex polyhedra to establish relations among c, λ, and the number d of states used in the stochastic realization, attempting to minimize d for given c and λ; we shall not pursue these details here. We also note that, like Theorem 3, Theorem 5 is independent of the nature of π and η_i for the ss A; only properties of the matrices $A(\sigma)$ are involved in the hypotheses. The following simple example shows that the hypotheses of the two theorems are not equivalent, i.e., there are cases where Theorem 5 provides realizations when Theorem 3 does not apply.

EXAMPLE. Let $\Sigma = \{\sigma\}$; i.e., there is only one input symbol (or we consider autonomous clocked systems). Let the number of states be $c = 2$, with

$$A(\sigma) = A = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \qquad AA^T = \frac{2}{3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so $\lambda = 2/3 < 1$ and Theorem 5 can be used (on the other hand, the row sums of absolute values in A are $2/\sqrt{3} > 1$ so Theorem 3 is not applicable). In two dimensions, the figure Δ must contain the unit disc; if for this purpose we examine the circumscribing regular polygons for $d = 3, 4, 5, 6, \ldots$, (triangle, square, pentagon, hexagon,...), the maximum squared length of a vector in Δ is the squared distance to a vertex, equal to 4, 2, 1.528, 4/3,..., respectively, and the product of these numbers with λ first becomes <1 for the case $d = 6$, so we choose Δ to be a regular hexagon circumscribing the unit disk. The matrix B can then be chosen to be a listing of the vertices of Δ with respect to some convenient orientation of axes; for example, placing a pair of opposite vertices on the horizontal axis, we may write

$$B = \frac{1}{\sqrt{3}} \begin{bmatrix} 2 & 0 \\ 1 & \sqrt{3} \\ -1 & \sqrt{3} \\ -2 & 0 \\ -1 & -\sqrt{3} \\ 1 & -\sqrt{3} \end{bmatrix}.$$

The equation $A'B = BA$ then has as one of its stochastic solutions

$$A' = \begin{bmatrix} 1-2\mu & \mu & \mu & 0 & 0 & 0 \\ 0 & \nu & \nu & 1-2\nu & 0 & 0 \\ 0 & \xi & \xi & 1-2\xi & 0 & 0 \\ 0 & 0 & 0 & 1-2\mu & \mu & \mu \\ 1-2\nu & 0 & 0 & 0 & \nu & \nu \\ 1-2\xi & 0 & 0 & 0 & \xi & \xi \end{bmatrix};$$

$$\mu = \frac{1}{2+\sqrt{3}}, \qquad \nu = \frac{1+\sqrt{3}}{5+\sqrt{3}}, \qquad \xi = \frac{\sqrt{3}-1}{1+3\sqrt{3}}.$$

Given any initial vector π of A, we can find a corresponding stochastic initial vector π' for A' by selecting K such that $K\pi \in \Delta$ and then solving $\pi'B = K\pi$; for example, if

$$\pi = (-1 \quad 1),$$

the choice $K = \sqrt{3} - 1$ places $K\pi$ on the perimeter of Δ, where it is a convex combination of the two neighboring vertices, thus yielding

$$\pi' = (0 \;\; 0 \;\; \sqrt{3}-1 \;\; 2-\sqrt{3} \;\; 0 \;\; 0).$$

Any arbitrary set $\{\eta_i\}$ of final vectors for A can then be transformed to the system A' by setting $\eta_i' = B\eta_i/(\sqrt{3}-1)$.

Appendix I

In establishing Corollary 3, we used the following fact: The family of all npa-realizable functions coincides with the closure of the family of pa-realizable functions with respect to addition of real constants and multiplication by real constants; moreover, if desired, the additive constants can be restricted to be nonpositive and the multiplicative constants can be restricted to be greater than or equal to 1. Equivalently, given a npa A, it is possible to construct a pa $\bar{A}$ and constants $\beta \geqslant 1$, $\gamma \leqslant 0$ such that

$$f^A(x) = \beta f^{\bar{A}}(x) + \gamma.$$

We demonstrate the latter here (equivalence with the former assertion is immediate); our construction of $\bar{A}$ may not be the most economical in state-set size but suffices to exhibit the result. The general method of construction is suggested by that used by Turakainen [6] to show that languages defined by cut-points for scalar ($b = 1$) npa can also be so defined using true scalar pa.

First, given the c-state npa $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$, let $A' = (\pi, \{A(\sigma)\}, \{\eta_i'\})$, where the vectors $\eta_i' = (\eta_{ij}')$ are defined by

$$\eta_{ij}' = (\eta_{ij} - \gamma)/\beta,$$
$$\gamma = \min(0, \min_{i,j} \eta_{ij}),$$
$$\beta = \max\left(1, \max_j \sum_{i=1}^{b} (\eta_{ij} - \gamma)\right).$$

Then we have $f^A(x) = \beta f^{A'}(x) + \gamma$, and β and γ lie in the asserted regions; note that $\gamma = 0$ unless some η_{ij} is negative. Now A' is a c-state npa such that all vector components η_{ij}' are nonnegative and

$$\sum_{i=1}^{b} \eta_{ij}' \leqslant 1, \qquad j = 1, 2, \ldots, c.$$

In this situation, it is easy to see that a straightforward "state-splitting" construction will produce a true npa $\bar{A}$ with $(b + 1)c$ states such that $\bar{A} \sim A$; specifically, we may define $\bar{A}$ as follows. Let $a_{mn}(\sigma)$ denote a typical element of the $c \times c$ matrix $A(\sigma)$, and let the definition of η_{ij}' be extended to include $i = b + 1$ by setting

$$\eta_{b+1,j}' = 1 - \sum_{i=1}^{b} \eta_{ij}' .$$

Then for automaton $\bar{A}$, the probability $a_{mj}(\sigma)\, \eta_{ij}'$ is assigned to the transition from state $m + c(k - 1)$ to state $j + c(i - 1)$ for $1 \leqslant i \leqslant b + 1$, $1 \leqslant j \leqslant c$, $1 \leqslant k \leqslant b + 1$, and $1 \leqslant m \leqslant c$; for each σ, this indeed defines a Markov matrix of size $(b + 1)c$ partitioned into $(b + 1)^2$ blocks in the obvious manner, and it is easy to see that for each string x the matrix $\bar{A}(x)$ is partitioned in the same manner, with entries $a_{mj}(x)\, \eta_{ij}'$ for the same indexing. The new final vectors $\bar{\eta}_i$ for $i = 1, 2, \ldots, b$ are defined by setting $\bar{\eta}_{ij} = 1$ when $c(i - 1) < j \leqslant ci$, with $\bar{\eta}_{ij} = 0$ otherwise; the $\bar{\eta}_i$ vectors are then evidently indicator functions for disjoint state sets. Taking the $(b+1)c$ component stochastic row vector $\bar{\pi}$ to be π followed by bc zeros, we have $\bar{\pi}\bar{A}(x)\,\bar{\eta}_i = \pi A(x)\,\eta_i{}'$.

Appendix II

In this Appendix, we include (i) remarks on the sense in which a finite-rank function is finitely specifiable so that a realization is effectively calculable, and (ii) details related to necessary and sufficient conditions for equivalence of ss.

Given a ss_c $A = (\pi, \{A(\sigma)\}, \{\eta_i\})$, let Π and N be the linear spaces spanned by the vectors $\pi(x)$ and $\eta_i(x)$, respectively, for all strings x and all $i \in Z$; for $t = 0, 1, 2,\ldots$, let Π_t and N_t be defined in the same way for all strings such that $|x| \leqslant t$. Then $\Pi_t \uparrow \Pi$, $N_t \uparrow N$, and moreover the dimension of $\Pi_t(N_t)$ increases strictly until it coincides with $\Pi(N)$, which must occur for $t \leqslant c-1$ (we omit details here, which are entirely parallel to those already standard in studies of pa and probabilistic transducers [1, 5]). Let f be a function of finite rank r, let A be any ss_r realization, and let G and H denote matrices as defined following (4); for all sufficiently large n the rows of G and columns of H can be selected to span Π and N, respectively, so this condition can be achieved for string arguments x_j and y_k all of whose lengths are bounded above by $r-1$. Therefore, in view of the factorization $F = GH$, a nonsingular f-matrix F of maximal size must occur among those f-matrices with string arguments such that $|x_j y_k| \leqslant 2(r-1)$, and then $F(\sigma)$ can be constructed from knowledge of the $(2(r-1)+1)$-segment of f; since this suffices to construct a realization, e.g. (13), from which all of f can be computed, we see that any function of finite rank r is uniquely determined by its $(2r-1)$-sement, and from this finite specification realizations are effectively constructible.

In the absence of any information about the value of $r = r(f)$ except its finiteness, this constructive interpretation is not possible, since an examination of the factorization $F = GH$ shows that it may happen that F is singular for all string arguments of a given length, while a nonsingular F is obtainable for some strings of greater length; therefore, r cannot be deduced effectively solely by examining $\det F$ for sets of strings of increasing length; an upper bound on r must be known. However, we have the following result (useful, e.g., in connection with prior knowledge of such a bound); the proof is straightforward and will be omitted here.

Theorem B1. *Let f be any external function, and let r_t be the maximal rank of any f-matrix formed from the t-segment of f. If for some integers t and u, we have $r_t = r_{t+1} = \cdots = r_{t+u} = t$, then either $r(f) = t$ or $r(f) \geqslant t + 2u$.*

As another instance of use of the factorization $F = GH$ and the relationship of G and H to the linear spaces Π and N, we can obtain the following formulation of sufficient and essentially necessary conditions for ss equivalence, subsuming the sufficient conditions of Theorem 2. Here, we omit the constant K mentioned in Theorem 2 (or we take $K = 1$), but it is easy to see that the additional flexibility afforded by introduction of K can be incorporated into the present formulation in the same way.

THEOREM B2. *Let A and A' be* ss, *and let G' be a matrix whose rows form a basis for the linear space Π' spanned by the vectors $\pi'(x)$. If there exists a matrix B such that*

$$\begin{aligned} \pi' B &= \pi, & & \\ G'BA(\sigma) &= G'A'(\sigma)B & &\text{for all } \sigma \in \Sigma, \\ G'\eta_i' &= G'B\eta_i & &\text{for all } i \in Z, \end{aligned}$$

then $A \sim A'$. Conversely, if $A \sim A'$ and A is in minimal form, then there exists B such that the above conditions hold.

Proof. For the first assertion we need to show that $\pi A(x)\,\eta_i = \pi' A'(x)\,\eta_i'$ for all strings x. We omit details; it suffices to note that the result follows by a straightforward induction on the length of x, first observing that the three conditions also hold for strings, i.e.,

$$\pi'(x)B = \pi(x), \qquad G'BA(x) = G'A'(x)B, \qquad G'\eta'(x) = G'B\eta(x),$$

which likewise follows by induction on string length, with the aid of the observation that according to the definitions of $\pi'(x)$ and G', for each x and each $i \in Z$ there is a row vector $v_{x,i}$ such that $G'\eta_i'(x) = v_{x,i}G'$, and for each x there is a matrix T_x such that $G'A'(x) = T_x G'$. Conversely, if $A \sim A'$ and the common external function is f of rank r, let F be an $r \times r$ nonsingular f-matrix; then $F = GH = G'H'$, where the matrices G and H for A, and G' and H' for A', are defined as in the discussion following (4). Then we also have

$$\pi' H' = \pi H, \qquad GA(\sigma)H = F(\sigma) = G'A(\sigma)\,H', \qquad G'\eta' = G\eta$$

into which the substitution $B = H'H^{-1}$ may be made to yield the desired result (B is well defined, since A is assumed minimal, so that H and G are $r \times r$ nonsingular matrices).

The above theorem is similar in spirit to its specialization, Theorem 2, but does not lead to useful computational criteria as in the case of the corollaries to Theorem 2, since in similar problem formulations the matrix G' is determined by the "unknown" system A' to be solved for; hence the number of "unknowns" is not manageable in practice when G' as well as B appears in the equations to be solved for A' based on A.

It should be noted that a companion to the above theorem can be obtained in which H is mentioned rather than G'; in this case, the conditions are

$$\pi' BH = \pi H, \qquad BA(\sigma)H = A'(\sigma)\, BH, \qquad \eta_i' = B\eta_i$$

and in the second half of the theorem, A' (rather than A) is required to be minimal and then B can be taken to be $(G')^{-1}G$.

As an easy corollary to either version of the theorem, we see that every minimal form of a given ss, or every minimal-state ss realization of a given finite-rank function, must be of the form (13′) for some choice of the matrix Q.

Finally, we see that if $A \sim A'$ and both A and A' are required to be minimal, B is square and nonsingular, so the relation $BA(\sigma) = A'(\sigma)B$ implies that $A(\sigma)$ and $A'(\sigma)$ have the same set of eigenvalues.

References

1. A. Paz, Some aspects of probabilistic automata, *Information and Control* **9** (1966), 26–60.
2. C. V. Page, Behavioral equivalences between probabilistic and deterministic sequential machines, *Information and Control* **9** (1966), 469–520.
3. J. W. Carlyle, On the external probability structure of finite-state channels, *Information and Control* **7** (1964), 385–397.
4. A. Paz, Minimization theorems and techniques for sequential stochastic machines, *Information and Control* **11** (1967), 155–166.
5. J. W. Carlyle, Stochastic finite-state system theory, *in* "System Theory" (L. A. Zadeh and E. Polak, Eds.), Chap. 10, McGraw-Hill, New York, 1969.
6. P. Turakainen, On probabilistic automata and their generalizations, *Ann. Acad. Sci. Fenn. Ser. A.I* **429** (1968), 1–52.
7. M. Nasu and N. Honda, Fuzzy events realized by finite probabilistic automata, *Information and Control* **12** (1968), 284–303.
8. J. W. Carlyle and A. Paz, On realization of probabilistic automata with prescribed external function, *Proc. 2nd Hawaii Conference on System Sciences* (1969), 77–80.