1

39026

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR JOZEF BRODY

TITLE OF THESIS/TITRE DE LA THÈSE STOCHASTIC MACHINES AND
PROBABILISTIC AUTOMATA

UNIVERSITY/UNIVERSITÉ Carleton University

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE Ph.D.

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1978

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE Dr. Stanislas Kłasa

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE 30/6/78 SIGNED/SIGNÉ [Signature]

PERMANENT ADDRESS/RÉSIDENCE FIXE 5795 Sir Walter Scott, Apt. 804
Montreal, Quebec, H4W-2T7

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

The undersigned recommend to the Faculty
of Graduate Studies acceptance of the thesis

"STOCHASTIC MACHINES AND PROBABILISTIC AUTOMATA"

submitted by Josef Brody, M.Sc.
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Karel Culik
Dr. K. Culik
External Examiner

Stanislaw Klawns
Thesis Supervisor

L. Culik
Chairman, Department of Mathematics

Carleton University

1978

STOCHASTIC MACHINES
AND
PROBABILISTIC AUTOMATA

J. Brody

© 1978 Josef Brody

ABSTRACT

Let $R = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ be a set of actions. The set of all finite sequences, including the empty sequence ϵ , of elements of R will be denoted by R^* .

A finite deterministic automaton A over R is a system $A = (Q, q_1, F)$, where Q is a finite set of states. The state $q_1 \in Q$ is the initial state and $F \subset Q$ is the set of final states.

There is also a transition function $f : Q \times R \rightarrow Q$, which is naturally extended to a function from $Q \times R^*$ to Q . A sequence w is accepted by A if and only if $f(q_1, w)$ is in F .

Let X be a set and R be a finite non empty set of binary relations on X . An X -machine of type R is a finite automaton $M = (Q, q_1, F)$.

If a sequence $w = (\alpha_1 \alpha_2 \dots \alpha_n)$ is accepted by M as an automaton, then $r = \alpha_1 \circ \alpha_2 \circ \dots \circ \alpha_n$ is accepted by the machine M . The event r is again a binary relation on X .

In this thesis, we arrive at the notion of initiated stochastic machine, to generalize the above deterministic devices. Let P be a function from R to the monoid of the substochastic square matrices of order s , the cardinality of Q . A stochastic X -machine M of type R is a pair (Q, P) .

The substochastic matrix $P(\alpha) = [p_{ij}(\alpha)] : \alpha \in R$ has entries equal to the probabilities $\Pr(q_j, \alpha | q_i)$. The machine M can stop in some states of Q . This possibility is a special state t , not in Q , called the terminal state. The function P produces a complementary substochastic vector τ .

satisfying $\sum_{\alpha \in R} P(\alpha) \cdot \eta + \tau = \eta = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. The vector $\tau = [p]_1$ has entries equal to the probabilities $\Pr(t|q_1)$.

We center our discussion around the description of different types of these machines. Various reductions, which can be performed and the relations between the machines are the broad topics of the thesis. Although the whole work is new and original, let us point out some of the results of the main importance.

Every machine can be represented by an oriented graph. The state q is accessible if there is a path $q_1 \rightarrow q$. Similarly a state q is coaccessible if there exists a path $q \rightarrow t$. In this thesis we establish a way, how to locate and remove all non accessible and non coaccessible states from the machines.

Besides the "useless" states, the machine may perform some "useless" moves. This happens if the identity relation i on X gives non zero matrix $P(i)$. The removal of all such moves is done globally and is considered the most important result of this work.

Let X be $\Gamma^* \times \Sigma^*$, where Σ and Γ are finite input and output alphabets, R be a set of the following actions:

$$R_Y \times I : (g, s) R_Y \times I = (gY, s),$$

$$R_Y \times L_\sigma^{-1} : (g, \sigma s) R_Y \times L_\sigma^{-1} = (gY, s) \quad \text{and}$$

$$I \times L_\sigma^{-1} : (g, \sigma s) I \times L_\sigma^{-1} = (g, s),$$

where $\gamma \in \Gamma$, $\sigma \in \Sigma$ and $(g, s) \in X$. The stochastic machine then becomes a stochastic transducer. If every input (ϵ, s) produces output (g, ϵ) , with $|g| = |s|$, the transducer is length-preserving. We have produced a constructive proof of the following theorem: Every length-preserving transducer can be accelerated, i.e. made of type $\{R_Y \times L_\sigma^{-1} : \gamma \in \Gamma, \sigma \in \Sigma\}$.

CONTENTS

Introduction	1
0. Preliminaries	10
1. Notation and basic definitions	19
2. Equivalence relations and reduction of ISM	33
3. Accessibility	46
4. Stochastic transducers	66
5. Two-way stochastic machines	100
References	110
List of symbols	113
Index	117

INTRODUCTION :

The main task of this thesis is to present some new developments in automata theory concerning stochastic machines. These developments were motivated by stochastic learning models, but we do not try to deal with these models in this thesis. To motivate the reader who is interested in such applications, in the introduction we present some of the aspects related to automata theory. No attempt is made to establish the psychological utilities of these models, but some references as (4), (8), (15) and (26) are provided for the reader who wishes to pursue the questions.

Generally we may describe a learning process by the following structure. At the beginning of the k -th trial, the subject is characterized by his state of learning q_k , which takes on values in a state space Q . On this trial, an event α_k occurs, in accordance with a probability distribution

$$p(q_k, \alpha_k) = \Pr(\alpha_k \in S | q_k)$$

over a subset S of an event space R . This in turn, effects a transformation

$$q_{k+1} = f(q_k, \alpha_k)$$

of state.

The transformation function f is called the event operator.

It is possible to study different mathematical models related to the finiteness or not of the set Q and R .

The finite state models are those whose state spaces Q have only a finite number of elements. In distance diminishing models, Q is a metric space with some metric d . In this case all the event operators $f(.,\alpha)$ are non expansive

$$d(f(w,\alpha), f(u,\alpha)) \leq d(w,v) \quad , \text{ for all } w, v \in Q .$$

The sequence q_n of states is a Markov Process. Thus the study of learning models leads to Markov Processes.

In learning models by means of problems we take the event $r \in R$ as a relation between the questions and the answers. This was described by the author in (6) and (7). The state space Q and the event space R are finite which leads to the study of finite automata of finite type R .

Later we shall show a connection of this thesis to one specific learning model, but before that we introduce the notions of automata and machines.

Let $R = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ be a finite non empty set (the set of actions), called the alphabet. Greek letters $\alpha, \beta \dots$ (possibly with subscripts) will usually denote elements of R . The set of all finite sequences of elements of R will be denoted by R^* . The elements of R^* will be called words. The empty word (i.e. the sequence of length zero) will be denoted by

ϵ . Subsets of R^* (i.e. sets of words) will sometimes be called events. Note that R^* with associative multiplication $v \cdot w$ (the concatenation of words v and w) and identity ϵ is a free monoid.

A finite deterministic automaton A over R is a system $A = (Q, q_1, F)$, where Q is a finite set (the set of states) of n states, $q_1 \in Q$ (the initial state) and $F \subset Q$ (the set of final states).

There is also a transition function $f : Q \times R \rightarrow Q$, which can be extended to a function from $Q \times R^*$ to Q by

$$\begin{aligned} f(q, \epsilon) &= q \\ f(q, w\alpha) &= f(f(q, w), \alpha) \end{aligned}$$

where $q \in Q$, $w \in R^*$ and $\alpha \in R$.

A word w is accepted by A if and only if $f(q_1, w)$ is in F . The set of all words accepted by A is denoted by $|A|$. An event $E \subset R^*$ is called a regular event if for some finite automaton A is $E = |A|$.

Eilenberg in (10) described a generalization of automata to machines.

Let X be a set and R be a finite nonempty set of relations on X . An X -machine of type R is a finite automaton $M = (Q, q_1, F)$.

If $w \in R^*$ is accepted by M as an automaton, then

$$r = F(w) = F(\alpha_1 \alpha_2 \dots \alpha_s) = \alpha_1 \circ \alpha_2 \circ \dots \circ \alpha_s$$

is accepted by the machine M . Note that $F(w)$ is a relation on X obtained by composing the relations $\alpha_i \in R$. The set M_p of all accepted relations $r \in F(R^*)$ is called a production set. The union

$$|M| = \bigcup_{r \in M_p} r$$

is called the behavior of M .

Let us follow the concept of the probabilistic automata (PA) in Rabin (23).

A probabilistic automaton A over R is a system $A = (Q, q_1, F)$, where Q, q_1, F are as in the deterministic case.

The transition function $f : Q \rightarrow \text{Fun}(Q \times R, [0,1])$ is a matrix valued function

$$f(q) = \begin{bmatrix} p_1(q, \alpha_1) & \dots & p_1(q, \alpha_m) \\ \dots & \dots & \dots \\ p_n(q, \alpha_1) & \dots & p_n(q, \alpha_m) \end{bmatrix}, \text{ where}$$

$$\sum_{i: q_i \in Q} \sum_{j: \alpha_j \in R} p_i(q, \alpha_j) = 1, \text{ for all } q \in Q.$$

Probabilistic automata are models for systems capable of a finite number of states $q_1, q_2, \dots, q_n$. When in state q , then the system can go into any of the states $q_i \in Q$ and receive an action $\alpha_j \in R$. The probability of this transition is the j -th element in the i -th row: $p_i(q, \alpha_j)$ of $f(q)$. These transition probabilities are assumed to remain fixed and be independent of time and previous actions.

A careful look on the function f shows that we really deal with three variables, two states and one action. Reorganizing the notation we get also $P : R \rightarrow \text{Fun}(Q \times Q, [0,1])$. This enables us to do the following.

For $\alpha \in R$ and $w \in R^*$ we define $n \times n$ substochastic matrices $P(\alpha)$ and $P(w)$ by

$$P(\alpha) = [p_j(q_i, \alpha)]$$

$$P(w) = P(\alpha_1 \alpha_2 \dots \alpha_s) = P(\alpha_1) \dots P(\alpha_s) = [p_j(q_i, w)]$$

An easy calculation will show that $p_j(q_i, w)$ is the probability of A for moving from q_i to q_j with the word w . The probability

$$p(w) = \sum_{i: q_i \in F} p_i(q_i, w)$$

is the probability for A to accept the word $w \in R^*$. In this thesis we say that w is accepted by A if $p(w) > 0$. The set of all accepted words by A is denoted by $|A|$ and is again regular,

The pair $\langle |A|, p \rangle$ is called the probabilistic language of A .

In this thesis we are not considering the notion of cut points as in (22) and (23). This is the reason why the probabilistic language has the support $|A|$ always as a regular event.

A generalization of PA has been described in (21) by Paz who defined the stochastic sequential machines.

A stochastic sequential machine (SSM) is a system $M = (Q, X, Y, P)$, where Q is as in the case of deterministic automata, X and Y are finite sets (the sets of input and output words) and P is a function from $X \times Y$ to the monoid of $n \times n$ - substochastic matrices

$$P(x, y) = [p_{ij}(y|x)] \text{ and}$$

$$p_{ij}(y|x) \geq 0 \text{ with } \sum_{y \in Y} \sum_{j: q_j \in Q} p_{ij}(y|x) = 1$$

for all $x \in X$ and $q_i \in Q$.

We may imagine the machine M as being at some state $q_1 \in Q$. This is arranged by an initial distribution η , where η is an n -stochastic row vector. The machine is then fed sequentially with a word $x = x_1 x_2 \dots x_s$, it prints the word $y = y_1 y_2 \dots y_s$ and moves on the next state q_j .

This transition is controlled by the matrices $P(x, y)$. The entry $p_{ij}(y|x)$ is the conditional probability of going to the state q_j and printing the output word y , given it had been in state q_i and fed the input word x .

The probability $p_\eta(v|u)$ of printing the word v if word u is sequentially fed into M is given by

$$p_\eta(v|u) = \eta \cdot P(u, v) \cdot \eta \text{ , where } \eta = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

The substochastic matrix $P(u, v)$ is given as a product

$$P(u,v) = P(u_1 u_2 \dots u_s, v_1 v_2 \dots v_s) = P(u_1, v_1) \cdot \dots \cdot P(u_s, v_s) .$$

Note that the definition of SSM uses all states as final.

We shall now present a mathematical model of learning by means of problems. This type of learning models has attracted many mathematicians and probabilists, who studied them in relation to Markov Processes. The reader is here referred to (16) and (19) as well as to (7) .

The set of states Q represents a set of states of abilities to function. The states $q_i \in Q$ are related to some topic studied by the learner.

Let Σ be a set of selected problems and Γ be a set of all possible solutions to these problems. The set $X = \Gamma * \Sigma^*$ shall then be a set of pairs of all finite sequences from Γ and Σ .

The finite set $R = \{\alpha, \beta, \dots\}$ of relations on X contains the following actions

$$(\gamma_1 \dots \gamma_n, \sigma_{n+1} \sigma_{n+2} \dots \sigma_m) \alpha = (\gamma_1 \dots \gamma_n \gamma_{n+1}, \sigma_{n+2} \dots \sigma_m)$$

i.e. α takes the problem σ_{n+1} and produces some of its solution γ_{n+1} .

At each state q_i of ability to function a learner may perform different actions. According to this performance α the state of ability to function q_i shall change with some probability $p_{ij}(\alpha)$ to another state q_j . We have

$$p_{ij}(\alpha) = \Pr(q_j, \alpha | q_i) .$$

If a learner attains some state q_i , p_i gives the probability of mastering the learned topic. Let t be the final state, which represents that situation. This gives the following

$$p_i = \Pr(t|q_i) .$$

The reader can easily realize that neither of the previous machines nor automata is adequate for such model. In this thesis we arrive at the notion of initiated stochastic machines (ISM) to accomodate the required properties of the above mathematical model of learning.

After the preliminaries (Chapter 0), where we give all the necessary mathematical prerequisites, we introduce the basic notions and notations (Chapter 1). Probabilistic languages and their relation to stochastic languages are discussed in this part, however we have not intended to go into great detail. This can be done as a generalization of formal languages described in (14).

In Chapter 2 we define some relations among the ISMs and give rather short proofs for various well-known problems.

In the next chapter (Chapter 3), we discuss accessibility and coaccessibility of ISMs. Different possible reductions are constructed. One specific reduction is especially useful.

The stochastic X-machine of type R , with identity relation 1 on X ($1 \in R$) and such that $P(1) \neq 0$, is called singular. As we move from one state to another we encounter some moves with the label 1 . These moves of the ISM are shown to be redundant. Various constructions were shown in order to remove such moves. This reduction is crucial in the construction of compilers for digital computers (13). The matrix al-

gebra, and the association of Markov Processes to ISMs, enables us to produce a new and quite compact construction of a new, non singular, ISM of type $R - 1$. The new machine gives the same stochastic language as the singular one and in that sense they are equivalent.

In Chapter 4 we offer some examples of stochastic transducers. This is a natural generalization of the known transducers (10). A transducer can be imagined as a machine changing the symbols of the word fed in into the new output word. The input word is on a tape and is fed into the machine sequentially, with a possibility to stop, while the machine may produce some output symbols. It is possible that the machine does not produce any symbol, while the input tape is running. Both, the input and the output tapes, can move only in one direction. The transducers are therefore referred to as one-way machines. If the length of the input and output word are always equal, we call it a length-preserving transducer.

It is known that any length-preserving transducer can be accelerated (10). This means that it is always possible to replace it by an equivalent one which operates without stopping either of its tapes. The usual proof of the acceleration is existential.

The generalization in this thesis enables us to produce a constructive proof of the corresponding theorem for length-preserving stochastic transducers.

In Chapter 5 we conclude the thesis with a study of the two way machines (10) and (22).

A one-way machine can be imagined as having simply a key board representing the symbols of the alphabet and as having the sequence for the tape fed in by successively stroking the keys. Thus no permanent record of the tape is required for the operation of the machine. A two-way machine, on the other hand, does need a permanent, actual tape on which it can run back and forth in trying to complete the answer.

As we follow the generalization of machines and probabilistic automata, we arrive to the idea of stochastic two-way machine. We construct a method which enables us to compute quickly the probability of a word to be accepted by the machine.

Although the whole concept of this thesis is completely new and original, some of the results are of the main importance. To emphasize this fact, we refer to them in the thesis as theorems.

It is a pleasure to acknowledge the encouragement and assistance of my supervisor S. Kłasa, as well as, of his wife Jacqueline. I am also much obliged to K. Culik for suggesting a number of changes in one of my drafts. The manuscript was proof-read by G. Matus to whom I am very grateful for all her comments.

Finally, the thesis is dedicated to my wife, Soňa, in gratitude for her inexhaustible patience.

0. PRELIMINARIES .

1. Functions and relations.

We shall write as $|A|$ the cardinality of a set A . The cardinality $|A|$ of an empty set $A = \emptyset$ is zero. A set A such that $|A| = 1$ is called singleton. We shall use the usual operations on sets.

- a/ the union as $\cup$
- b/ the intersection as $\cap$
- c/ the difference as $-$
- d/ the cartesian product as $\times$

Given two sets X and Y we shall use the notation

$$(0.1) \quad f : X \rightarrow Y \quad \text{or} \quad X \xrightarrow{f} Y$$

to denote a function defined for all x in X with values in Y . If $g : Y \rightarrow Z$ is another function, we denote $f \circ g : X \rightarrow Z$ the composition of f and g

$$(0.2) \quad f \circ g : X \xrightarrow{f} Y \xrightarrow{g} Z$$

We may also write

$$(0.3) \quad (f \circ g)(x) = g(f(x))$$

We shall also use an alternate parenthesis free notation xfg . In this notation the element x in X may be interpreted as a function $x : I \rightarrow X$, where I is a singleton. Then

$g \circ f$ is the composition of functions

$$(0.4) \quad x \circ f \circ g : 1 \rightarrow X \rightarrow Y \rightarrow Z$$

Let $\hat{X}$ be the power set of X , i.e. the set of all subsets of X (including the empty subset $\emptyset$). A relation f from X to Y , written $f : X \rightarrow Y$, is a function

$$(0.5) \quad f : \hat{X} \rightarrow Y$$

which satisfies the additivity condition

$$(0.6) \quad f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i), \text{ for any index set } I.$$

If for each x in X the set $f(x)$ is a singleton then f is a function. Thus, relations are generalization of functions and we shall therefore use the same notation for both. Composition of relations is defined in the obvious way using the original definition (0.5).

The graph of a relation $f : X \rightarrow Y$ is defined as the subset of $X \times Y$

$$(0.7) \quad \{ (x, y) : y \in f(x) \}$$

In this work we shall not distinguish between the notion of a relation and its graph.

A relation $f : X \rightarrow Y$ is called a partial function if for each x in X the set $f(x)$ has

at most one

2. Free monoids.

Let Σ be a finite set of distinct elements $\{\sigma_1, \dots, \sigma_n\}$ $n > 0$. The elements σ_i are called letters and Σ is called an alphabet. A sequence w of k letters is called a word of length k , denoted $|w| = k$. If u is another word of length l , then the word wu of length $k + l$ is formed by concatenation. This operation is associative and every word w of length k is the product of k letters. The word of length 0 is denoted by ϵ and is a unit element for multiplication. The set Σ^* of all words on Σ with the above multiplication is a free monoid. In general a monoid, not necessarily free, may be defined as a set with an associative binary operation with an identity element.

An example of a monoid is the set of all $s \times s$ real matrices with multiplication as the operation.

Let $A = [a_{ij}]$ be a matrix. We shall call a_{ij} its entry and (i, j) the address. Let $B = [b_{ij}]$ be another matrix. We shall use the following relations

a/ $A \leq B$ if $a_{ij} \leq b_{ij}$ for all (i, j) .
b/ $A < B$ if $a_{ij} < b_{ij}$ for all (i, j) .
and obviously $A = B$ if $a_{ij} = b_{ij}$ for all (i, j) .

In the case of vectors we shall use similar relations as for matrices.

We shall denote by I_s the unit matrix of order s

and we shall write I if the order is obvious. The vector η_s of s ones shall also be written as η if the order is clear from its use.

Let A be a matrix or a vector. The transposition of A , $A^t = [a_{ij}]^t = [a_{ji}]$ changes the address (i,j) to (j,i) .

Certain functions will occur in this work. They are

a/ the left multiplication by σ , $L_\sigma : \omega L_\sigma = \sigma \omega$

b/ the right multiplication by σ , $R_\sigma : \omega R_\sigma = \omega \sigma$.

We shall also use the inverse partial functions

a/ $L_\sigma^{-1} : \rho \omega L_\sigma^{-1} = \omega$ if $\sigma = \rho$
 $= \phi$ otherwise

b/ $R_\sigma^{-1} : \omega \rho R_\sigma^{-1} = \omega$ if $\sigma = \rho$
 $= \phi$ otherwise.

Let Γ^* be another free monoid and let $G : \Gamma^* \rightarrow \Gamma^*$ and $F : \Sigma^* \rightarrow \Sigma^*$ be two functions. The function

$$(0.8) \quad G \times F : \Gamma^* \times \Sigma^* \rightarrow \Gamma^* \times \Sigma^*$$

is defined by

$$(0.9) \quad (g,s)G \times F = (gG,sF)$$

3. Probability.

In any probability analysis of an experiment, there are two logical steps. First one chooses a set of possibilities U , cal-

led the sample space, and then assigns a probability measure. In this work we shall only consider countable sample spaces. The probability measures shall be discrete.

We get a probability measure on U if we assign to each element a in U a non negative number $\Pr(a)$ satisfying

$$(0.10) \quad \sum_{a \in U} \Pr(a) = 1.$$

The measure $\Pr(A)$ of a subset A of U is

$$(0.11) \quad \Pr(A) = \sum_{a \in A} \Pr(a).$$

It often happens that a probability measure is assigned to a set U when new information comes to us. We shall now determine a new probability measure of this set, using the original measure $\Pr$.

Let U be a set for which a measure $\Pr$ has been assigned. The conditional probability measure on U with respect to its subset A is defined as

$$(0.12) \quad \Pr(a|A) = \frac{\Pr(a)}{\Pr(A)} \quad \text{if } a \in A \\ = 0 \quad \text{otherwise.}$$

In (0.12) we shall assume that the value $\Pr(A)$ given in (0.11), is positive. It follows from (0.11) that $\Pr(A) \leq 1$.

Let U be a sample space with a probability measure $\Pr$. This measure induces a subprobability measure,

re on every subset A of U .

Let Pr be a measure defined on U . The set

$$(0.13) \quad A = \{ a \in U : Pr(a) > 0 \}$$

is called the support of the probability measure Pr .

4. Random variables and Markov Processes.

Each subset A of the sample space U is called an event. The set of all events over a sample space is closed under the intersection, union and complementation. This structure is called σ -algebra.

A random variable X is a function from the sample space U into the set of real numbers R .

$$(0.14) \quad X : U \rightarrow R.$$

Given a σ -algebra, a probability measure, and a random variable over it, a related distribution function from the real numbers to the interval $[0,1]$ is defined as follows.

Let A_x be the event such that

$$(0.15) \quad A_x = \{ a \in U : X(a) \leq x \}.$$

Then the distribution function $F : R \rightarrow [0,1]$ is given by

$$(0.16) \quad F(x) = Pr(A_x)$$

Sometimes the notation $\Pr(X(a) \leq x)$ is used for $\Pr(A_x)$ and the notation $\Pr(X(a) = x)$ is used for $\Pr(B_x)$, where

$$(0.17) \quad B_x = \{ a \in U : X(a) = x \}$$

Let X and Y be two random variables over a σ -algebra and a probability measure over it. Then we define

$$(0.18) \quad \Pr(X = x | Y = y) = \frac{\Pr(X = x \wedge Y = y)}{\Pr(Y = y)}$$

where $\Pr(Y = y) \neq 0$ and

$$(0.19) \quad (X = x) \wedge (Y = y) = B_x \cap B_y$$

The random variables X and Y are said to be independent if

$$(0.20) \quad \Pr(X = x | Y = y) = \Pr(X = x)$$

From (0.18) and (0.20) it follows that the random variables are independent if

$$(0.21) \quad \Pr(X = x \wedge Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

If A and B are independent events then

$$(0.22) \quad \Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

Let $X_1, X_2, \dots$ be a sequence of random variables such

that for any n

$$(0.23) \quad \Pr(X_n = j | X_1 = n_1, \dots, X_{n-1} = i) = \Pr(X_n = j | X_{n-1} = i),$$

i.e. the random variable X_n depends only on the random variable X_{n-1} . Such a system is a discrete Markov process. Such stochastic processes can be represented in the following model.

The sample space U is represented by a finite, or countable number of vertices. The random variable X_n represents the position of a moving point at time $t = n$. $\Pr(X_n = j)$ is the probability that the point will be at the vertex v_j at time $t = n$.

The conditional probability

$$(0.24) \quad \Pr(X_n = j | X_{n-1} = i) = p_{ij}(n)$$

is the probability that the point will be at the vertex v_j at time $t = n$ provided that it has been at vertex v_i at time $t = n-1$.

If $p_{ij}(n) = p_{ij}$ for all natural numbers n , then the stochastic process is called homogeneous, and otherwise it is called non homogeneous.

It is tacitly assumed throughout this thesis that the stochastic process is homogeneous and discrete, i.e. the transition from state to state occurs at discrete time intervals.

The probabilities

$$(0.25) \quad p_{ij} = \Pr(j|i)$$

can be arranged in a matrix form and such matrix is called stochastic or Markov. Clearly any stochastic matrix $[p_{ij}]$ has the property that

$$(0.26) \quad p_{ij} \geq 0 \quad \text{and} \quad \sum_j p_{ij} = 1$$

The matrix $[p_{ij}]$ such that

$$(0.27) \quad p_{ij} \geq 0 \quad \text{and} \quad \sum_j p_{ij} \leq 1$$

is called substochastic.

Each row of a stochastic matrix is a stochastic row vector. Similarly each row vector of a substochastic matrix is called a substochastic row vector. A stochastic vector with only one nonzero coordinate is a degenerate stochastic vector.

Let $[p_{ij}]$ be a stochastic matrix associated with some stochastic process. The vertex v_i in U is called an absorbing barrier if $p_{ii} = 1$. This means that once a moving point enters the vertex v_i it will stay there for ever.

1. NOTATION AND BASIC DEFINITIONS .

Let X be the set called machine vocabulary and $R = \{\alpha, \beta, \dots\}$ a finite set of binary relations on X called the set of actions .

DEFINITION 1.1 . A stochastic X -machine M of type R is a pair $M = (Q, P)$, where $Q = \{q_1, \dots, q_s\}$ is a finite set called the set of states . P is a function from R to the monoid of the substochastic square matrices of order s .

The machine M . can stop in some states of Q . This possibility may be considered as a special state t of the machine , obviously not in Q , called the terminal state . The function P produces a complementary substochastic vector τ . We shall require that τ satisfies

$$(1.1) \quad \sum_{\alpha \in R} P(\alpha) \cdot \eta + \tau = \eta$$

If we use

$$(1.2) \quad SP = \sum_{\alpha \in R} R(\alpha) \cdot ,$$

then from (1.1) it follows that

$$(1.3) \quad \tau = [I - SP] \eta$$

The purpose of a stochastic machine (SM) is to define a function $p : R^* \rightarrow [0,1]$. Later we shall show that p is

a subprobabilistic measure on R^* .

We may represent any SM : M by a valuated labeled directed graph $G_M = (V, E)$. The set of vertices $V = Q \cup \{t\}$ and the set of directed edges $E = E_1 \cup E_2$, where $E_1 \subset Q \times Q$ and $E_2 \subset Q \times \{t\}$. The subset of edges E_1 is labeled by the actions $\alpha_i \in R$ and carries values given by conditional probabilities

$$(1.4) \quad p_{ij}(\alpha) = \Pr(q_j, \alpha | q_i)$$

The above probabilities can be found as the entries of the substochastic matrices $P(\alpha)$, i.e.

$$(1.5) \quad P(\alpha) = [p_{ij}(\alpha)]$$

The directed edges from E_1 take the form of

$$q_i \xrightarrow{\alpha/p_{ij}(\alpha)} q_j$$

The subset of edges E_2 is not labeled but does have values given by the conditional probabilities

$$(1.6) \quad p_i = \Pr(t | q_i)$$

These probabilities are the entries of the substochastic column vector τ , i.e.

$$(1.7) \quad \tau = [p_i]$$

The requirement (1.1) is to assure that at every state q_i of Q one of the edges has to be followed. The terminal state t acts as an absorbing barrier in the Markov Process associated with the stochastic square matrix of order $s+1$

$$(1.8) \quad \tilde{P} = \begin{bmatrix} SP & \tau \\ 0 \dots 0 & 1 \end{bmatrix}, \text{ where } SP \text{ is given in (1.2) .}$$

EXAMPLE 1.2 .

Let $X = [0, \infty[$ be the set of all non negative real numbers. Let the set of actions be given as $R = \{ \alpha, \beta, \gamma \}$, where $\alpha = (x, x)$; $\beta = (x, 0)$; $\gamma = (x, x+1)$, where $x \in X$. Let $Q = \{ q_1 \dots q_5 \}$ and the function P be given as

$$P(\alpha) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}; \quad P(\beta) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix};$$

$$P(\gamma) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}; \text{ giving } \tau = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

We may now see that the pictorial representation of this stochastic machine $M = (Q, P)$ is

Figure 1.9

The matrices SP and H are

$$SP = P(\alpha) + P(\beta) + P(\gamma) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{6} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{bmatrix};$$

$$H = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{6} & 0 & \frac{1}{3} & \frac{1}{6} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

H is a stochastic matrix, i.e. $H \cdot \eta = \eta$.

Let R^* be the free monoid of all finite sequences of elements of R , called the set of generalized actions. The empty sequence ϵ is the unit element in R^* and the concatenation of sequences is the binary operation. The generalized actions $w \in R^*$ will be sequences of n actions from R , where $n = 0, 1, \dots$.

Each path $q_i \rightarrow t$, as a juxtaposition of directed edges from E , will carry a label $w = \alpha_1 \alpha_2 \dots \alpha_n$, where α_i are labels of the directed edges of $q_i \rightarrow t$ in E_1 . The value of $q_i \rightarrow t$ will be the product of the values of all the directed edges involved in the juxtaposition.

The directed graph G_M is not necessarily simple and therefore one path may have more than one value and more than one label.

In our example 1.2 the path $q_1 \rightarrow t = q_1 \rightarrow q_2 \rightarrow q_5 \rightarrow q_2 \rightarrow t$ has label $w = \gamma \beta \gamma$ and value $\frac{1}{2} \times \frac{1}{3} \times \frac{1}{4} \times \frac{1}{6} = \frac{1}{144}$. There are also other paths in G_M which do not terminate at t but we shall not consider them now.

It is not yet clear which state should initiate the move along which path. This choice is made by a stochastic row vector $\mathbb{P}$ with coordinates

$$(1.10) \quad \mathbb{P}_i = \text{Pr}(q_i)$$

$\mathbb{P}_i$ are the probabilities of starting a path at the state q_i .

DEFINITION 1.3. An initiated stochastic X -machine M of type R is a triple $M = (Q, \mathbb{P}, P)$.

We shall say that a generalized action w is produced by the initiated stochastic machine (ISM) M if there exists a path $q_i \rightarrow t$ with the label w and also $\eta_i > 0$.

In order to calculate the probability $\tilde{p}_M(w)$, of w being produced by M , we shall encounter all the paths with label w and the probability of their realization. If we denote by $c_i(w)$ the set of all paths initiated at q_i and labeled by w , then

$$\begin{aligned}\tilde{p}_M(w) &= \sum_{i=1}^S \Pr(q_i \xrightarrow{w} t \cap q_i) = \sum_{i=1}^S \Pr(q_i) \sum_{c \in c_i(w)} \Pr(c|q_i) \\ &= \sum_{i=1}^S \eta_i \sum_{c \in c_i(w)} \Pr(c|q_i) = \eta \cdot P(w) \cdot \tau.\end{aligned}$$

In the above calculation we used the matrix $P(w)$ given by

$$(1.11) \quad P(w) = P(\alpha_1 \alpha_2 \dots \alpha_n) = P(\alpha_1) \cdot P(\alpha_2) \cdot \dots \cdot P(\alpha_n)$$

This gives the following

PROPOSITION 1.4. The probability $\tilde{p}_M(w)$ of w being produced by M is given by

$$(1.12) \quad \tilde{p}_M(w) = \eta \cdot P(w) \cdot \tau,$$

where $P(w)$ is given by (1.11).

In this work a special emphasis is given to ISMs with degenerate probability distribution of initial states

$$(1.13) \quad \mathbb{I}^i = (0, \dots, 1, \dots, 0)$$

with 1 in i -th position.

We may always rename the states in Q in such way that the initial state will be the state q_1 . We shall use the triple $M = (Q, q_1, P)$, rather than $M = (Q, \mathbb{I}^1, P)$.

The reason for the emphasis on ISMs with degenerate initial distribution is mainly to follow Eilenberg (10) but is also motivated by the following optimization problem.

If w is the generalized action to be produced we shall look for such ISM M giving the maximal probability $\tilde{p}_M(w)$. From (1.12) follows that the best result is guaranteed by such ISM: $M = (Q, \mathbb{I}^i, P)$ for which the probability

$$\tilde{p}_M(w) = \max_i \mathbb{I}^i \cdot P(w) \cdot \tau$$

Until now we have been distinguishing different generalized actions. We have not yet used the fact that the actions $\alpha \in R$ are binary relations on the set X . To utilize this property we shall introduce the following homomorphism F as

$$(1.14) \quad F(w) = F(\alpha_1 \alpha_2 \dots \alpha_n) = \alpha_1 \circ \alpha_2 \circ \dots \circ \alpha_n,$$

where $\circ$ is the composition of relations. If $w = \epsilon$, i.e. the empty sequence, and $x \in X$, we shall define

$$(1.15) \quad F(\epsilon) = 1 = \{(x, x)\} \text{ the identity relation on } X$$

To clarify the way the composition in (1.14) is working

we say that $(x, y) \in r = F(\alpha_1 \alpha_2 \dots \alpha_n)$ if and only if there exists a sequence $x = x_0, x_1, \dots, x_n = y$ such that $(x_{i-1}, x_i) \in \alpha_i$ for $i = 1, \dots, n$.

In general F is not injective. F maps R^* to the power set of $X \times X$.

The set $F(R^*)$ is therefore a monoid of binary relations on X with composition $\circ$ as its binary operation and 1 as its identity.

The measure $\tilde{p}_M$ is defined on R^* . We shall now introduce a new measure p_M on $F(R^*)$

$$(1.16) \quad p_M(r) = \sum_{w \in F^{-1}(r)} \tilde{p}_M(w) = \sum_{w \in F^{-1}(r)} P(w). \tau$$

The inverse image $F^{-1}(r)$ of r is defined as follows -

$$(1.17) \quad F^{-1}(r) = \{ w \in R^* : F(w) = r \}$$

for all r in $F(R^*)$.

The value $p_M(r)$ is the probability of the event r from $F(R^*)$ to be produced by M . We shall later prove that $\tilde{p}_M$ and p_M are subprobability measures on R^* and $F(R^*)$ respectively.

We have seen so far that each $ISM : M$ produces two types of events :

a/ generalized actions from R^* , with probability given by the measure $\tilde{p}_M$, and

b/ binary relations on X from $F(R^*)$, with subpro-

bability measure p_M .

We can easily see that unless F is an isomorphism these two measures are very different.

COMMENT 1.5 . If the monoid $F(R^*)$ is free, then we shall obtain a classical notion of a probabilistic automaton (PA). The homomorphism F is then injective.

It is now possible to introduce the following two new sets.

DEFINITION 1.6 . The set M_p , given as

$$(1.18) \quad M_p = \{ r \in F(R^*) : p_M(r) > 0 \},$$

is called the production set of M . The set $|M|$, defined by

$$(1.19) \quad |M| = \bigcup_{r \in M_p} r,$$

is called the behavior of M .

There is a considerable difference between the sets M_p and $|M|$. The elements of $M_p \subset F(R^*)$ are relations. M_p is the support for the subprobability measure p_M .

On the other hand $|M|$ is a relation on X .

If M is an initiated probabilistic automaton (IPA), the elements of M_p can be considered as generalized actions. The subprobability measures $\tilde{p}_M$ and p_M may be identified as follows from (1.16). We shall call M_p , in this case, the language of M and the pair $\langle M_p, p_M \rangle$

the stochastic language of M .

EXAMPLE 1.2 (continuation).

The relations α, β, γ satisfy the following properties

- a/ $\alpha = 1$ is the identity relation on X .
- b/ $\gamma^n = (x, x+n)$, where $n = 1, 2, \dots, x \in X$
- c/ $\beta \circ \gamma^n = (x, n)$, for $n = 1, 2, \dots, x \in X$
- d/ $r \circ \beta = \beta$, for any $r \in F(R^*)$.

This gives $F(R^*) = \{(x, x+n), (x, n) : n = 0, 1, \dots, x \in X\}$.

Although we shall not yet provide the subprobability measure p_M , we can already come to conclusion that

$$a/ \quad p_M(1) = \sum_{n=0}^{\infty} P^n(\alpha) \cdot \tau = 0.$$

b/ To find the probability $p_M(x, x+n)$, we are going to evaluate all paths with label giving this relation. The relation $(x, x+n)$ is obtained from the path

$$q_1 \rightarrow q_2 \rightarrow q_1 \rightarrow q_2 \rightarrow \dots \rightarrow q_1 \rightarrow q_2 \rightarrow t,$$

where the edge $q_1 \rightarrow q_2$ is involved n -times.

This gives

$$p_M(x, x+n) = \frac{1}{2} \times 6^{-n}.$$

c/ To produce the relation (x, n) , we have to take any path $q_1 \rightarrow \dots \rightarrow q_2 \rightarrow t$ with some restrictions. At least one edge with label β must be involved. After taking this edge for the last time we have to take an edge with label γ in

order to get to the state q_2 . Let us denote by a the sum of the values of all such paths. From this position we shall take the circuit $q_2 \rightarrow q_1 \rightarrow q_2$ with label $\alpha \gamma$ $(n-1)$ -times before finally terminating at t . This shall provide the probability

$$p_M(x, n) = a/6^n$$

d/ The relation $(x, 0)$ is produced by any path.

$$q_1 \rightarrow \dots \rightarrow q_2 \rightarrow q_5 \rightarrow q_5 \rightarrow \dots \rightarrow q_5 \rightarrow t$$

To find the value of such paths we shall use the results from b/ and c/

$$p_M(x, 0) = \left(\frac{1}{2} \sum_{n=0}^{\infty} 6^{-n} + a \cdot \sum_{n=0}^{\infty} 6^{-n} \right) \cdot \frac{1}{3} \cdot \sum_{n=0}^{\infty} 4^{-n} \cdot \frac{1}{2} = \frac{2}{15} + \frac{4}{15} a$$

Unfortunately, we can not yet calculate the value a but we shall be able to do it later. It is clear though that $a \neq 0$ and therefore $M_p = F(R^*) - 1$. The behavior $|M|$ is the relation with the graph

EXAMPLE 1.7 .

In this example the ISM : M is given as a triple $M = (Q, q_1, P)$, where $Q = \{q_1, q_2, q_3\}$ i.e. $s = 3$. Let us take the set of actions $R = \{\alpha\}$, i.e. there is only one action in R . Let

$$P(\alpha) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}, \text{ giving } \tau = \begin{bmatrix} 0 \\ \frac{1}{4} \\ 1 \end{bmatrix}$$

We may illustrate our ISM : M by the following graph

The set of actions R has only one element and therefore we have $\mathcal{SP} = P(\alpha)$

Let us use the following abbreviation

$$w = (\alpha\alpha\dots\alpha) = \alpha^n \quad \text{i.e.} \quad |w| = n,$$

for $n = 1, 2, \dots$, with $\epsilon = \alpha^0$. This gives

$$R^* = \{ \alpha^n : n = 0, 1, \dots \}$$

The stochastic matrix H is given as

$$H = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let us now assume that the action α satisfies the following conditions

$$\alpha = \alpha_1, \alpha \circ \alpha = \alpha_2 \neq \alpha_1 \quad \text{and} \quad \alpha \circ \alpha \circ \alpha = 1$$

Then the behavior $|M|$ of M is given as

$$|M| = 1 \cup \alpha_1 \cup \alpha_2$$

To find $p_M(r) = \Pr(r \text{ being produced by } M)$, we shall use the formula (1.16), i.e.

$$p_M(r) = \frac{1}{\sum_{w \in F^{-1}(r)} 1} \sum_{w \in F^{-1}(r)} P(w) \cdot \tau = \frac{1}{\sum_{w \in F^{-1}(r)} 1} \sum_{w \in F^{-1}(r)} \tilde{p}_M(w)$$

If we realize that $M_p = \{1, \alpha_1, \alpha_2\}$ and

$$F^{-1}(1) = \{\alpha^{3i} : i = 0, 1, \dots\},$$

$$F^{-1}(\alpha_1) = \{\alpha^{3i+1} : i = 0, 1, \dots\},$$

$$F^{-1}(\alpha_2) = \{\alpha^{3i+2} : i = 0, 1, \dots\},$$

we get the following results

$$p_M(1) = \sum_{i=0}^{\infty} \tilde{p}_M(\alpha^{3i}) = \sum_{i=0}^{\infty} 7 \cdot 3^{3i-2} / 2 \cdot 4^{3i} = \frac{7}{4},$$

$$p_M(\alpha_1) = \sum_{i=0}^{\infty} \tilde{p}_M(\alpha^{3i+1}) = \frac{1}{8} + \sum_{i=1}^{\infty} 7 \cdot 3^{3i-1} / 2 \cdot 4^{3i+1} = \frac{25}{4},$$

$$p_M(\alpha_2) = \sum_{i=0}^{\infty} \tilde{p}_M(\alpha^{3i+2}) = \sum_{i=0}^{\infty} 7 \cdot 3^{3i} / 2 \cdot 4^{3i+2} = \frac{1}{37}.$$

2. EQUIVALENCE RELATIONS AND REDUCTION OF ISM .

The probability measures $\tilde{p}_M$ and p_M define two types of languages.

DEFINITION 2.1 . Let M be an ISM. The couple $\langle R^*, \tilde{p}_M \rangle$ is called the probabilistic language of M . The pair $\langle F(R^*), p_M \rangle$ is called the stochastic language of M . We shall also say that $\langle R^*, \tilde{p}_M \rangle$ and $\langle F(R^*), p_M \rangle$ are recognized by M .

As we mentioned before, the probabilistic and the stochastic languages are essentially equal if M is an IPA.

In order to understand the above notions, we shall now introduce the following equivalence relations.

DEFINITION 2.2 . Let M and M' be two ISMs.

a/ M and M' are c-equivalent (computationally equivalent) if and only if they have the same probabilistic languages. We shall denote this by $M \stackrel{c}{\sim} M'$.

b/ M and M' are e-equivalent if and only if they have the same stochastic languages. We shall denote this by $M \sim M'$.

c/ M and M' are b-equivalent (behaviorally equivalent) if and only if they have the same behavior. We shall denote this by $M \stackrel{b}{\sim} M'$.

If M and M' are b-equivalent, we do not compare their subprobability measures. This enables us to compare any two machines, not necessarily stochastic ones.

We shall now introduce the notion of initiated machines as done by Eilenberg in (10).

DEFINITION 2.3. An initiated X -machine of type R is a triple (Q, q_1, T) , where Q and q_1 are as in the case of ISMs and $T \subset Q$ is the set of terminal states.

Besides the above triple there is a set of labeled edges E with elements $(q_i, \alpha, q_j) : q_i \xrightarrow{\alpha} q_j$, where q_i and q_j are from Q and α belongs to R .

A path $c : q_1 \rightarrow q_j$ is a sequence of edges with identical connecting states. Each path initiates at q_1 and terminates at some q_j from T . All such paths are labeled by $w = \alpha_1 \alpha_2 \dots \alpha_n$, where α_i are labels of the edges of the path c . Each label $w \in R^*$ of such a path shall be called a successful generalized action. The associated relation $r = F(w)$ shall be called a successful event. We shall use M_p for all successful events from M . The behavior $|M|$ is again given by (1.19).

PROPOSITION 2.4. For every ISM there exists a b-equivalent initiated machine (IM) .

PROOF.

Let $M = (Q, q_1, P)$ be an ISM. Let us define IM $\bar{M}$ by $\bar{M} = (Q, q_1, T)$ as : q_1 is a terminal state of $\bar{M}$ if and only if $p_1 > 0$. The set of edges E shall then con-

tain the edge $q_i \xrightarrow{\alpha} q_j$ if and only if $p_{ij}(\alpha) > 0$. The rest is obvious.

The IM $\bar{M}$ and the ISM M , are not only b-equivalent, they have also the same production sets, i.e. $M_p = \bar{M}_p$. This enables us to conclude the following.

PROPOSITION 2.5. Let A be a subset of $F(R^*)$. There exists an ISM M such that $M_p = A$ if and only if A is recognizable by some IM.

It is now obvious that if $\langle F(R^*), p_M \rangle$ is a recognizable stochastic language then the support of p_M must be recognized also by some IM.

The question remains whether the stochastic language reveals more than a machine language. In other words, does every subprobability measure p on $F(R^*)$ with the support recognizable by an IM give a recognizable stochastic language $\langle F(R^*), p \rangle$?

Although the above question is beyond the scope of this work, we shall present the following conjecture.

CONJECTURE 2.6. Let $A \subset F(R^*)$ be a set recognized by some IM. Let $B \subset A$ be a set not recognized by any IM. Let p and p' be two subprobability measures on $F(R^*)$ with the common support A . Let $p(r) \neq p'(r)$ if and only if $r \in B$. Then at least one of the two stochastic languages $\langle F(R^*), p \rangle$ or $\langle F(R^*), p' \rangle$ is not recognized by any ISM.

As the above conjecture compares stochastic languages with recognizable subsets of the monoid $F(R^*)$, we shall now do the

same comparison for probabilistic languages. Let us first introduce the notion of automata.

DEFINITION 2.7 . Let $M = (Q, q_1, T)$ be an initiated X -machine of type R . If $F(R^*)$ is a free monoid then we shall call M an initiated R -automaton (IA).

Similarly as we proved the proposition 2.4, we may present a proof of the following proposition.

PROPOSITION 2.8 . For every IPA A there exists an IA $\bar{A}$ such that $A_p = \bar{A}_p$.

We should realize that the production sets A_p and $\bar{A}_p$ are in this case essentially subsets of R^* , i.e. sets of sequences from R . The fact that R is a set of relations on X is not important.

We can present the following conjecture involving the probabilistic languages now.

CONJECTURE 2.9 . Let $A \subset R^*$ be a recognizable set by some IA, i.e. A is a regular event. Let $B \subset A$ be a non regular event. Let p and p' be two subprobability measures on R^* with the same support A . Let $p(w) \neq p'(w)$ if and only if $w \in B$, then at least one of the two probabilistic languages $\langle R^*, p \rangle$ or $\langle R^*, p' \rangle$ is not recognized by any IPA.

EXAMPLE 2.10 .

Let $R = \{ \alpha, \beta \}$. R^* is a regular event (10). It is also known that the event $B = \{ \alpha^n \beta^n : n \geq 0 \}$ is not regu-

lar. (10). Let p and p' be the following subprobability measures on R^*

$$p(w) = \frac{1}{2} \cdot 4^{-n}, \text{ where } n = |w|$$

$$p(w) = 0 \text{ if } w \neq \alpha^n \beta^n$$

$$p(w) = p(w) = \frac{1}{4} \cdot 8^{-2n} \text{ if } w = \alpha^n \beta^n \text{ and } n \text{ is even,}$$

$$p(w) = \frac{1}{4} \cdot 8^{-2n} \text{ if } w = \alpha^n \beta^n \text{ and } n \text{ is odd.}$$

From the conjecture 2.9 it follows that the probability language $\langle R^*, p' \rangle$ is not recognizable.

This is true, because $\langle R^*, p \rangle$ is recognized by the IPA $A = (\{q_1\}, q_1, P)$, where

$$P(\alpha) = P(\beta) = \left[\frac{1}{2} \right] \text{ giving } \tau = \left[\frac{1}{2} \right].$$

Let us now employ the other two equivalences described in the definition 2.2.

PROPOSITION 2.11. Let M and M' be two ISMs of the same type R .

$$a/ \text{ If } M \sim M' \text{ then } M \stackrel{b}{\sim} M'.$$

$$b/ \text{ If } M \stackrel{c}{\sim} M' \text{ then } M \sim M'.$$

PROOF.

a/ If $M \sim M'$ then $\langle F(R^*), p_M \rangle = \langle F(R^*), p_{M'} \rangle$, i.e. the support of p_M is the same as of $p_{M'}$. This gives $M_p = M'_p$ and from (1.19) also $|M| = |M'|$ giving $M \stackrel{b}{\sim} M'$.

b/ If $M \stackrel{c}{\sim} M'$ then $\langle R^*, \tilde{p}_M \rangle = \langle R^*, \tilde{p}_{M'} \rangle$. From (1.16) it follows that $p_M = p_{M'}$, giving $\langle F(R^*), p_M \rangle = \langle F(R^*), p_{M'} \rangle$, i.e. $M \sim M'$.

We shall now use an example which shows that the converse to the above proposition 2.11 is false.

EXAMPLE 2.12.

Let $X =]0,1[$ and R be the set of the following relations

$$R = \{ \gamma = (x, x) ; \alpha = (x, \sqrt{x}) ; \beta = (x, x^2) , x \in X \}$$

Let M^1 , M^2 and M^3 be initiated stochastic X -machines of type R such that :

$$Q_1 = Q_2 = \{ q_1, q_2, q_3 \} ; Q_3 = \{ q_1, q_2 \}$$

$$P_1(\gamma) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} ; P_1(\alpha) = P_2(\alpha) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} ;$$

$$P_1(\beta) = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} ; P_2(\beta) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \quad P_2(\gamma) = 0.$$

$$\text{giving } \tau_1 = \tau_2 = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

$$P_3(\alpha) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} ; \quad P_3(\beta) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} ; \quad \text{and} \quad P_3(1) = 0 .$$

This gives $\tau_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

The monoid $F(R^*) = \{ \alpha_i = (x, x^{2^i}) , i \in \mathbb{Z} \}$, with $\alpha = \alpha_{-1}$
and $\beta = \alpha_1$.

Figure 2.1

From the graphs in the figure 2.1 it follows that

$$\tilde{p}_{M^1}(1\beta) = 1/6 \quad \text{but} \quad \tilde{p}_{M^2}(1\beta) = \tilde{p}_{M^3}(1\beta) = 0 .$$

this gives M^1 not c-equivalent to M^2 nor to M^3 .

It is also clear that the only paths with positive values in M^1 have labels $w = (1*\beta\alpha)1*\beta$. The relations α and β are mutually inverse and therefore $F(w) = \beta$. The direct computation results in

$$p_{M^1}(\beta) = \sum_{i=0}^{\infty} \left(\sum_{j=0}^{\infty} 2^{-j} \cdot \frac{1}{4} \right)^i \cdot \sum_{k=0}^{\infty} 2^{-k} \cdot \frac{1}{8} = \frac{1}{2}$$

Similarly the only paths with positive values in M^2 have labels $v = (\beta\alpha)*\beta$ with $F(v) = \beta$. This gives

$$p_{M^2}(\beta) = \sum_{i=0}^{\infty} 2^{-i} \cdot \frac{1}{4} = \frac{1}{2}$$

From the above, it follows that the subprobability measures p_{M^1} and p_{M^2} are equal. We have therefore $M^1 \sim M^2$. It is easy to realize that $p_{M^3}(\beta) = .1$ i.e. M^1 is not equivalent to M^3 as well as M^2 is not equivalent to M^3 . The three subprobability measures p_{M^1} , p_{M^2} and p_{M^3} have common support $\{\beta\}$ and therefore give the same behavior for all the three ISMs. This gives $M^1 \stackrel{b}{\sim} M^2 \stackrel{b}{\sim} M^3$.

We may now conclude the example with the realization of the relations

M^1 is not c-equivalent to M^2 but $M^1 \sim M^2$;
 M^2 is not equivalent to M^3 but $M^2 \stackrel{b}{\sim} M^3$.

In the above example 2.12 we saw that $M^1 \sim M^2$. Yet M^1 and M^2 may be considered of different type. M^1 is of type R while M^2 may be studied as an ISM of type R - 1. We

can now see the following.

Let ι be the identity relation on X . If $\iota \in R$, then any ISM of type R might encounter some edges with the label ι . Moving along these edges does never change the involved relations and therefore are, in this sense, useless. These so called singular moves lead us to the following definition.

DEFINITION 2.13. Let ι be the identity relation on X and let $\iota \in R$. Let $M = (Q, q_1, P)$ be an initiated stochastic X -machine of type R such that $P(\iota) \neq 0$. We shall call such ISM an initiated singular stochastic machine (ISSM).

The question arises whether it is possible, and under what conditions this is so, to find for every ISSM M an equivalent ISM $\bar{M}$. The answer is given in the following theorem.

THEOREM 2.14. Let M be an ISSM of type R . Let $P(\iota)$ be a matrix such that $I - P(\iota)$ is non singular. Then there exists an ISM $\bar{M}$ of type $R - \iota$ and $M \sim \bar{M}$.

PROOF.

Let us take any event $r \in F(R^*)$. To find the probability $p_M(r)$ requires a consideration of all the probabilities $\tilde{p}_M(w)$ where $w \in F^{-1}(r)$. The set $F^{-1}(r)$ has the following property.

$$(2.2) \quad \tilde{w} = \alpha_1 \dots \alpha_n \in F^{-1}(r)$$

if and only if all the sequences $\iota(w) \in F^{-1}(r)$, where

$$l(w) = l * \alpha_1 l * \dots * l * \alpha_n l *$$

Let us denote by $R(r)$ the set

$$(2.3) \quad R(r) = \{ w \in (R - 1)^* : F(w) = r \}$$

Using the notation of (2.2) we see that with each sequence of type $w \in R(r)$ all the sequences $l(w)$ are also in $F^{-1}(r)$. From (2.2) follows

$$(2.4) \quad F^{-1}(r) = R(r) \cup \{ l(w) : w \in R(r) \} = \{ l(w) : w \in R(r) \}.$$

According to (1.16) we have

$$(2.5) \quad p_M(r) = \sum_{w \in R(r)} \tilde{p}_M(l(w))$$

The probability $\tilde{p}_M(l(w))$ in (2.5), using (2.2), equals

$$(2.6) \quad \begin{aligned} \tilde{p}_M(l(w)) &= \prod_{i=0}^1 P^i(l) P(\alpha_1) \prod_{i=0}^{\infty} P^i(l) \dots \prod_{i=0}^{\infty} P^i(l) \tau \\ &= \prod_{i=0}^1 (I - P(l))^{-1} P(\alpha_1) \dots (I - P(l))^{-1} \tau \end{aligned}$$

The formulae (2.5) and (2.6) suggest we define a new function $\bar{P} : R - 1 \rightarrow \bar{P}(R)$ given by

$$(2.7) \quad \bar{P}(\alpha) = (I - P(l))^{-1} P(\alpha) \quad \text{for all } \alpha \in R - 1$$

The new ISM : $\bar{M} = (Q, q_1, \bar{P})$, with

$$(2.8) \quad \bar{\tau} = (I - P(l))^{-1} \tau$$

will give us the required result.

To prove this we still have to prove that the new $\bar{M}$ satisfies the equation (1.1), i.e.

$$(2.9) \quad \sum_{\alpha \in R-1} \bar{P}(\alpha) \cdot \eta + \bar{\tau} = \eta$$

Let us use the fact that M is an ISM of type R i.e. satisfies the equation (1.1). Using (2.7) and (2.8) we have

$$\begin{aligned} \sum_{\alpha \in R-1} \bar{P}(\alpha) \eta + \bar{\tau} &= (I - P(1))^{-1} \left(\sum_{\alpha \in R-1} P(\alpha) \eta + \tau \right) \\ &= (I - P(1))^{-1} \left(\sum_{\alpha \in R} P(\alpha) \eta + \tau - P(1) \eta \right) \\ &= (I - P(1))^{-1} (\eta - P(1) \eta) \\ &= (I - P(1))^{-1} (I - P(1)) \eta = \eta \end{aligned}$$

The above proof of the theorem 2.14 is constructional and gives us through formulae (2.7) and (2.8), the description of the non singular ISM $\bar{M}$. This construction is of global character, i.e. it treats all the edges with the label 1 at the same time. The corresponding theorem for non stochastic machines is known, and one of its forms is presented in Gries (13). The construction of the equivalent non singular machine is of local character, i.e. they deal with each edge and state separately. Those constructions are therefore lengthy and tedious.

REMARK . The construction of a non singular ISM $\bar{M}$ for a singular one M always produces another example of a pair of ISMs which are equivalent but never c-equivalent.

EXAMPLE 2.15

In the example 1.2 the ISM M is an ISSM, for

$$P(1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} \neq 0, \text{ giving}$$

$$(I - P(1))^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{4}{3} \end{bmatrix}$$

The construction of the non singular ISM $\bar{M}$, as indicated by (2.7) and (2.8), is possible and results in

$$\bar{P}(B) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \bar{P}(Y) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \end{bmatrix}$$

giving $\bar{\tau} = \begin{bmatrix} 0 \\ \frac{1}{6} \\ 0 \\ \frac{1}{6} \\ \frac{2}{3} \end{bmatrix}$

EXAMPLE 2.16 .

In the example 2.12 we have a pair of ISMs : M^1 and M^2 .
 The ISM M^1 is ISSM and the ISM M^2 is non singular and
 equivalent to M^1 . Let us use the above construction on M^1
 to produce $\bar{M}^1$ and compare it with the ISM M^2 .

$$P_1(1) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{giving} \quad (I - P_1(1))^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This gives

$$\bar{P}_1(\alpha) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} ; \quad \bar{P}_1(\beta) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \bar{T} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

and we see that $\bar{M}^1 = M^2$.

3. ACCESSIBILITY.

The question, we shall pursue now, is whether, for a state q in Q , there is some path c of any length initiating at q and terminating at t . In this case, if we say path, we always mean a path with positive value. The vector τ indicates with its non zero coordinates all the states in Q directly connected with the terminal state t .

DEFINITION 3.1 . Let M be an ISM. If there is a path in M initiating at $q \in Q$ and terminating at t , then q is called a coaccessible state.

If all states in Q are coaccessible, we shall call M a coaccessible ISM.

To be able to trace all coaccessible states we shall introduce the following proposition.

PROPOSITION 3.2 . Let $M = (Q, q_1, P)$ be an ISM, then the following conditions are equivalent.

a/ M is coaccessible.

b/ For every $q_1 \in Q$ there exists a non negative integer $n(i)$ such that

$$(3.1) \quad q^i \cdot SP^{n(i)} \cdot \tau > 0$$

c/ For every $i = 1, \dots, s$, the behavior $|M_i|$ of $M_i = (Q, q_i, P)$ is non empty.

PROOF .

a/ $\rightarrow$ b/

If M is coaccessible, then for every $q_i \in Q$ there exists a path c of a length $n(i) + 1$ connecting q_i and t . Let $w = \beta_1 \beta_2 \dots \beta_{n(i)}$ be the label of this path. Then the vector $\eta(w) = [p_i(w)]$ given by

$$(3.2) \quad \eta(w) = P(\beta_1) \cdot P(\beta_2) \cdot \dots \cdot P(\beta_{n(i)}) \cdot \tau$$

will have a positive i -th coordinate. This coordinate, $p_i(w)$, gives the value of the path c .

The $n(i)$ -th power of SP , given by the formula (1.2), encounters all the labels of paths of the length $n(i)$, i.e. also the path c . This gives

$$(3.3) \quad SP^{n(i)} \cdot \tau \geq \eta(w)$$

From (3.3) and the fact that $p_i(w) > 0$ it follows (3.1).

b/ $\rightarrow$ c/

If $q_i \in Q$, then there exists $n(i)$ such that (3.1) is true. From (1.2) we may deduce

$$(3.4) \quad SP^{n(i)} = \sum_{w \in R(n(i))} P(w)$$

where $R(n) = \{ w \in R^* : |w| = n \}$.

From (3.3) we can see that there exists a generalized action $w \in R(n(i))$ such that

$$(3.5) \quad \mathbb{P}^i \cdot P(w) \cdot \tau > 0$$

The event $r = \bar{r}(w)$ gives positive probability $p_{M_i}(r)$. This shows that the behavior $|M_i| \neq \emptyset$.

c/ $\rightarrow$ a/

If $|M_i| \neq \emptyset$, then there exists a path $c_i : q_i \rightarrow t$ i.e. q_i is coaccessible. This is true for all $q_i \in Q$.

The ISM in the example 1.2, as well as the non singular one in the example 2.15, is non coaccessible. This follows from the proposition 3.2, by realizing that $\mathbb{P}^3 \cdot SP^n \cdot \tau = 0$ for every positive integer n .

If the ISM M is coaccessible then the terminal state t is the only absorbing barrier in the Markov Process described by the stochastic matrix H , given by (1.8). Using this we will show the following.

PROPOSITION 3.3. Let M be an ISM, then the following properties are equivalent.

a/ M is coaccessible.

b/ $\lim_{n \rightarrow \infty} SP^n = 0$.

c/ The matrix $I - SP$ is non singular.

PROOF.

a/ $\rightarrow$ b/

From the proposition 3.2 we have (3.1) and therefore we can state that for some n

$$(3.6) \quad \eta_n = \sum_{i=0}^n SP^i \cdot \tau > 0$$

Repeating (1.1) and multiplying it by SP we reach the result that

$$(3.7) \quad SP^{n+1} \cdot \eta = \eta - \eta_n < \eta$$

Let us put

$$\lambda = \max_i \eta^i \cdot SP^{n+1} \cdot \eta$$

then $0 \leq \lambda < 1$. If we now multiply SP^{n+1} and SP we get a matrix with all entries less than or equal to λ^2 .

Repeating this argument we get b/

$$b/ \rightarrow c/$$

$$\lim_{i \rightarrow 0} \left(\sum_{i=0}^n SP^i \right) (I - SP) = I - \lim_{i \rightarrow 0} SP^{n+1} = I$$

This gives the inverse of $I - SP$

$$(3.8) \quad (I - SP)^{-1} = \sum_{n=0}^{\infty} SP^n$$

and proves c/

$$c/ \rightarrow a/$$

Let us suppose that M is not coaccessible, i.e. there exists a subset $\bar{Q}$ of Q containing all non coaccessible states. In the matrix SP we may assume that the states from $\bar{Q}$ are the last k states. This enables us to write SP as

$$(3.9) \quad SP = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

where C is a stochastic square matrix of order k . Therefore

$$(3.10) \quad I - SP = \begin{bmatrix} I_{s-k} - A & 0 \\ 0 & I_k - C \end{bmatrix},$$

with $I_k - C$ singular, i.e. $I - SP$ also singular.

The proposition 3.3 finally enables us to prove some important properties of the measures $\tilde{p}_M$ and p_M for an ISM M .

PROPOSITION 3.4 . Let M be a coaccessible ISM, then $\tilde{p}_M$ and p_M are probability measures on R^* and $F(R^*)$ respectively. This means that

$$(3.11) \quad \sum_{w \in R^*} \tilde{p}_M(w) = \sum_{r \in F(R^*)} p_M(r) = 1.$$

PROOF .

From (1.3) and the proposition 3.3, it follows that

$$(3.12) \quad (I - SP)^{-1} \cdot \tau = \eta$$

From (1.12) we have

$$\begin{aligned} \sum_{w \in R^*} \tilde{p}_M(w) &= \sum_{r \in F(R^*)} p_M(r) = \mathbb{1}^1 \sum_{w \in R^*} P(w) \tau = \mathbb{1}^1 \sum_{n=0}^{\infty} SP^n \tau \\ &= \mathbb{1}^1 (I - SP)^{-1} \tau = \mathbb{1}^1 \eta = 1 \end{aligned}$$

We can now show why $\tilde{p}_M$ and p_M are always subprobability measures, i.e.

$$(3.13) \quad \sum_{w \in R^*} \tilde{p}_M(w) = \sum_{r \in F(R^*)} p_M(r) \leq 1.$$

PROPOSITION 3.5 : Let M be an ISM, then (3.13) holds.

PROOF :

If M is coaccessible the proposition 3.5 obviously holds.

If M is not coaccessible then for every non coaccessible state q we may choose one edge : $q \rightarrow q_1$ and decrease its value by some $\varepsilon > 0$. This gives us a possibility to introduce an edge $q \rightarrow t$ and obtain a new, coaccessible, machine $\bar{M}$. It is obvious that

$$\tilde{p}_{\bar{M}}(w) \geq \tilde{p}_M(w) \quad \text{for every } w \in R^*.$$

The property (3.13) then follows from the proposition 3.4.

We should realize that the measure p_M is always probabilistic if M is coaccessible. The converse is not true though, as we can see in the following simple example.

EXAMPLE 3.6 :

Let $R = \{ \alpha \}$ and $M = (Q, q_1, P)$. Let Q be equal to $\{ q_1, q_2 \}$ and P be given as

$$P(\alpha) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}, \text{ giving } \tau = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}.$$

The matrix

$$I - SP = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}$$

is singular and from the proposition 3.3, it follows that M is not coaccessible.

On the other hand it is true that

$$\sum_{w \in R^*} P(w) = \sum_{n=0}^{\infty} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}^n \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} = \sum_{n=1}^{\infty} 2^{-n} = 1$$

We see that p_M is a probability measure on $F(R^*)$ although M is not coaccessible.

From the definition of the stochastic language it is clear that the non coaccessible states do not contribute anything. These states are therefore "useless" and we shall try to remove them.

Before we tackle this problem, we shall realize that it is impossible to construct an ISM M' , such that it is coaccessible and $M' \sim M$, where M is a non coaccessible ISM. This follows from the fact that p_M is not necessarily a probability measure but $p_{M'}$ is always a one.

We shall therefore try at least to find $M' \sim^b M$ with $p_{M'} \geq p_M$ and common support.

We shall therefore offer the following definition.

DEFINITION 3.7 . Let M and M' be two initiated stochastic X-machine of type R . We shall call M' better than M , and denote it by

$$M' > M, \text{ if } M'_p = M_p \text{ and } p_{M'} \geq p_M.$$

We shall now prove the following proposition.

PROPOSITION 3.8. Let M be an ISM; then there exists a coaccessible $\bar{M}$ such that $\bar{M} > M$.

PROOF.

If we keep all the edges $q_i \rightarrow q_j$, where q_i and q_j are coaccessible states and remove all the other ones, we shall obtain a coaccessible ISM. On the other hand let us reevaluate all these, remaining, edges in such a way that the values shall not decrease, then we get the required ISM $\bar{M}$. The change in the values shall be done to satisfy the property (1.1) of any ISM.

Let $\bar{Q} \subset Q$ be the set of all coaccessible states in Q . Let us delete all the rows and columns from all the $P(\alpha)$ and also from the vector τ corresponding to the non coaccessible states in Q . The new $P'(\alpha)$ and τ' give

$$(3.14) \quad 0 < \sum_{\alpha \in R} P'(\alpha) \cdot \eta + \tau' \leq \eta.$$

Let us introduce the following probabilities λ_i

$$(3.15) \quad \lambda_i = \frac{1}{\sum_{\alpha \in R} P'(\alpha) \cdot \eta + \tau'} \left(\sum_{\alpha \in R} P'(\alpha) \cdot \eta + \tau' \right) = \sum_{q \in \bar{Q}} \Pr(q|q_i).$$

The state q_i is also coaccessible.

Let us define $\bar{P}$ as

$$(3.16) \quad \bar{P}(\alpha) = [\bar{p}_{ij}(\alpha)], \text{ where } \bar{p}_{ij}(\alpha) \cdot \lambda_i = p_{ij}(\alpha)$$

giving $\bar{\tau} = [\bar{p}_i]$, where $\bar{p}_i \cdot \lambda_i = p_i$, where q_i and q_j

are in $\bar{Q}$.

Using (3.14) and (3.15) we see that $\bar{p}_{ij}(\alpha) \geq p_{ij}(\alpha)$ for all $\alpha \in R$ and $\bar{p}_i \geq p_i$ for all q_i and q_j in $\bar{Q}$.

The proof will be completed by showing that $\bar{M}$ given as $(Q, q_1, \bar{P})$ satisfies (1.1).

$$\begin{aligned}
 (3.17) \quad SP\eta + \bar{\tau} &= \sum_{\alpha \in R} \bar{P}(\alpha)\eta + \bar{\tau} = \sum_{\alpha \in R} \left(\sum_{q_j \in \bar{Q}} \bar{p}_{ij}(\alpha) + \bar{p}_i \right) \\
 &= \sum_{\alpha \in R} \left(\left(\sum_{q_j \in \bar{Q}} p_{ij}(\alpha) + p_i \right) / \lambda_i \right) \\
 &= \left(\sum_{\alpha \in R} P'(\alpha)\eta + \tau' \right) / \lambda_i = 1
 \end{aligned}$$

EXAMPLE 3.9.

Let us continue with the example 1.2 as we left it in the example 2.15. It is easy to see that, q_3 is a non coaccessible state in the ISM $\bar{M}$. From the proposition 3.8 follows that we can remove q_3 and "improve" on the ISM $\bar{M}$. Let us in this example use M instead of $\bar{M}$ and reserve this for the new ISM we shall now construct.

The matrices $P'(\beta)$ and $P'(\gamma)$ and the vector τ are

$$P'(\beta) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & \frac{1}{3} \\ \frac{1}{6} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix} ; \quad P'(\gamma) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \end{bmatrix} ;$$

and giving

$$\tau' = \begin{bmatrix} 0 \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{2}{3} \end{bmatrix}$$

This gives the vector $\lambda = [\lambda_i]$ as

$$\lambda = (P'(\beta) + P'(\gamma)) \cdot \eta + \tau' = \begin{bmatrix} 1 \\ \frac{5}{6} \\ \frac{5}{6} \\ 1 \end{bmatrix}$$

From (3.16) we get $\bar{P}$ and $\bar{\tau}$ as

$$\bar{P}(\beta) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad \bar{P}(\gamma) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}; \quad \bar{\tau} = \begin{bmatrix} 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

The above construction results in the coaccessible non singular

ISM : $\bar{M} = (\bar{Q}, q_1, \bar{P})$, where $\bar{Q} = \{q_1, q_2, q_3, q_4\}$.

In the example 1.2 we arrive at the probabilities $p_M(r)$ on $F(R^*)$. These probabilities still depend on a parameter a . The ISM : M in the example was non coaccessible and therefore (3.13) holds. The equality in (3.13) is true only if all λ_i , given by (3.15), are equal to 1. In the above example 3.9 we see that (3.15) will be a strict inequality. Let us replace, in the

example 1.2, the edge : $q_3 \xrightarrow{8/1} q_3$ by $q_3 \xrightarrow{1} t$, which shall make M coaccessible. From the proposition 3.4, it follows that (3.15) becomes an equality.

Let us calculate the probabilities $p'_M(r)$ for all paths through q_3 .

a/ $r_{1,n} = (x, x+n)$, for $n = 2, 3, \dots$ gives

$$p'_M(r_{1,n}) = 2^{-1} 6^{1-n}, \text{ using } q_1 \rightarrow q_2 \rightarrow q_1 \rightarrow \dots \rightarrow q_2 \rightarrow q_3 \rightarrow t.$$

b/ $r_{2,n} = (x, n)$, for $n = 2, 3, \dots$ gives

$$p'_M(r_{2,n}) = a \cdot 6^{-n}, \text{ using}$$

$$q_1 \rightarrow \dots \xrightarrow{8} q \xrightarrow{1} q_2 \rightarrow q_1 \rightarrow q_2 \rightarrow \dots \rightarrow q_2 \rightarrow q_3 \rightarrow t$$

c/ $r_3 = (x, 0)$ can never be produced by a path through q_3 , i.e. $p'_M(r_3) = 0$.

This gives us

$$\begin{aligned} \sum_{n=1}^{\infty} p_M(x, x+n) + \sum_{n=1}^{\infty} p_M(x, n) + p_M(x, 0) + \sum_{n=2}^{\infty} p'_M(x, x+n) + \\ + \sum_{n=2}^{\infty} p'_M(x, n) = 1, \end{aligned}$$

where p_M are calculated in the example 1.2. After a short calculation we get $a = 4/3$ and

$$p_M(x, x+n) = 2^{-1} \cdot 6^{-n}, \quad p_M(x, n) = 4/3 \cdot 6^{-n} \quad \text{and} \quad p_M(x, 0) = 2/3.$$

giving

$$\sum_{r \in F(R^*)} p_M(r) = \frac{27}{96}$$

Using a similar method as the example 1.2 we shall find the probabilities $p_{\bar{M}}(r)$ for $\bar{M}$ in our example 3.9.

$$p_{\bar{M}}(x, x+n) = \frac{1}{2} \cdot 5^{-n} > p_M(x, x+n) \quad ; \quad n = 1, 2, \dots$$

$$p_{\bar{M}}(x, n) = \frac{1}{4} \cdot 5^{-n} > p_M(x, n) \quad ; \quad n = 1, 2, \dots$$

$$\text{and} \quad p_{\bar{M}}(x, 0) = \frac{1}{4} > p_M(x, 0)$$

It is easy to verify that $p_{\bar{M}}$ is a probability measure on $F(R^*)$.

As we construct a coaccessible ISM from a non coaccessible one, as shown in the proposition 3.8, we need to locate all the non coaccessible states. To do this, we shall take the following approach.

Let us define the vector

$$(3.18) \quad \tau_j = SP^j \cdot \tau \quad \text{giving} \quad \tau_{k+1} = SP \cdot \tau_k$$

This way we get the following sets and vector spaces

$$(3.19) \quad \mathcal{K}^M = \{ \tau_0, \tau_1, \dots \}$$

$$(3.20) \quad \mathcal{K}(n) = \text{sp}\{ \tau_0, \tau_1, \dots, \tau_n \} \quad \text{and} \quad \mathcal{K} = \text{sp} \mathcal{K}^M$$

The rank of $\mathcal{K}(i)$ is less or equal to the rank of $\mathcal{K}(j)$ and this less or equal than the rank of $\mathcal{K}$ which is equal to

$|Q| = s$, for all $i \leq j$. We are going to prove the following.

PROPOSITION 3.9 . If $\mathcal{H}(m-1) = \mathcal{H}(m)$ for some m , then

$$\mathcal{H}(m-1) = \mathcal{H}.$$

PROOF .

We shall proof by induction that $\tau_i \in \mathcal{H}(m-1)$ for $i = 0, 1, \dots, j$ and for all j .

a/ If $j = m-1$, then the assertion follows from the definition of $\mathcal{H}(m-1)$.

b/ Let $\tau_i \in \mathcal{H}(m-1)$ for $i = 0, 1, 2, \dots, k$.

Then

$$\tau_{k+1} = SP.\tau_k = SP \sum_{i=0}^{m-1} \lambda_i \tau_i = \sum_{i=0}^{m-1} \lambda_i SP.\tau_i = \sum_{i=0}^m \lambda_{i-1} \tau_i,$$

giving $\tau_{k+1} \in \mathcal{H}(m) = \mathcal{H}(m-1)$. This shows that $\mathcal{H}^M \subset \mathcal{H}(m-1)$ and therefore $\mathcal{H} = \mathcal{H}(m-1)$.

The vectors in (3.18) will help us to locate all non coaccessible states in M .

PROPOSITION 3.10 . The state $q_i \in Q$ is non coaccessible in M if and only if the i -th coordinate of the vector θ , given by

$$(3.21) \quad \theta = \sum_{i=0}^{m-1} \tau_i = \sum_{i=0}^{m-1} SP^i.\tau$$

is equal to zero.

PROOF .

Let θ_i be the i -th coordinate of the vector θ then

$$(3.22) \quad \theta_i = \eta^i \sum_{j=0}^{m-1} SP^j \tau = \eta^i \sum_{|w| \leq m-1} P(w) \cdot \tau = \sum_{|w| \leq m-1} \tilde{p}_{M(i)}(w)$$

If q_i is non coaccessible then all probabilities $p_{M(i)}$, where $M(i) = (Q, q_i, P)$, are equal to zero, i.e. $\theta_i = 0$.

If $\theta_i = 0$, then from the proposition 3.9 follows that all τ_j have the i -th coordinate equal to zero. We then conclude that

$$(3.23) \quad \sum_{w \in R^*} \tilde{p}_{M(i)}(w) = \eta^i \sum_{w \in R^*} P(w) \cdot \tau = \eta^i \sum_{i=0}^{\infty} SP^i \tau = 0$$

giving that q_i is non coaccessible.

The following is then readily apparent.

COROLLARY 3.11 . Let M be an ISM, then M is coaccessible if and only if $\text{rank } K = s = |Q|$.

In the theorem 2.14 we assumed that the matrix $I - P(1)$ was non singular. In the light of coaccessibility, especially as a result of the proposition 3.3, we can see that :

COROLLARY 3.11 . Let M be a coaccessible ISM of type R . Then we have

a/ $I - P(\alpha)$ is non singular for all $\alpha \in R$.

b/ If M is a SISM there exists a non singular ISM : $\bar{M}$ and $\bar{M} \sim M$.

PROOF :

a/ From the proposition 3.3 follows that $\lim SP^n = 0$. For any $\alpha \in R$ is $P(\alpha) SP$, i.e. $\lim P^n(\alpha) = 0$ as well. This is equivalent with the assertion that $I - P(\alpha)$ is non singular.

b/. From a/ follows that $I - P(1)$ is non singular and the rest is the result of the theorem 2.14.

COROLLARY 3.12. Let M be a SISM. Then there exists a non singular ISM: $\bar{M}$, and $\bar{M} > M$.

• PROOF.

From the proposition 3.8 follows that there exists M' , which is coaccessible and $M' > M$. From the corollary 3.11 then follows that there exists a non singular $\bar{M}$ such that $\bar{M} \sim M'$, giving the required result.

A very similar situation, as coaccessibility, is with "accessible" states. But first the definition.

DEFINITION 3.13: Let M be an ISM. If there is a path in M , starting at the initial state q_1 and reaching q , the state q is called an accessible state.

If all q in Q are accessible, we shall call M an accessible ISM.

We can easily see that all non accessible states are "useless" as far as $\tilde{p}_M$, p_M , M_p and $|M|$ are concerned.

To remove all non accessible states presents no difficulty, if we know about them and can locate them. If we cross-out all the corresponding rows and columns in every $P(\alpha)$, where $\alpha \in R$ as well as in τ , the new machine becomes accessible. This operation will not change any of the values of the remaining edges. To compare this process with the removal of the non coaccessible states, it does not require any additional changes. This simplification is assured by the fact that there is no edge from an accessible

to a non accessible state. We have just shown that :

PROPOSITION 3.13 . Let M be a non accessible ISM, then there exists an accessible one $\bar{M}$ and $\bar{M} \stackrel{L}{\sim} M$.

In order to trace all non accessible states, we shall approach the problem similarly as in the case of the search for the non coaccessible states.

PROPOSITION 3.14 . Let $\xi_j = \mathbb{I}^1 SP^j$ and $L^M = \{ \xi_0, \xi_1, \dots \}$.
Let $L(n) = \text{sp}\{ \xi_0, \xi_1, \dots, \xi_n \}$ and
 $L = \text{sp } L^M$. If $L(n-1) = L(n)$ for some n ,
then $L(n-1) = L$.

PROOF .

We may follow the proof of the proposition 3.9 almost step by step. It is perhaps slightly different in showing that

$\xi_i \in L(n-1)$ for $i = 0, 1, \dots, k$ then also $\xi_{k+1} \in L(n-1)$.

Let us therefore show this derivation.

$$\xi_{k+1} = \xi_k SP = \sum_{j=0}^{n-1} c_j \cdot \xi_j \cdot SP = \sum_{j=0}^n c_{j-1} \cdot \xi_j \in L(n) = L(n-1).$$

We shall show that again it is possible to locate all non accessible states.

PROPOSITION 3.15 . The state q_1 in Q is non accessible in M if and only if the i -th coordinate of the vector $\psi = [\psi_i]$ given as

$$(3.24) \quad \psi = \sum_{j=0}^{n-1} \xi_j = \mathbb{I}^1 \sum_{j=0}^{n-1} SP^j \quad \text{is equal to zero.}$$

PROOF .

Let ψ_i be the i -th coordinate of ψ , then

$$(3.25) \quad \begin{aligned} \psi_i &= \mathbb{1}^1 \sum_{j=0}^{n-1} SP^j \cdot (\mathbb{1}^1)^t = \mathbb{1}^1 \sum_{|w| \leq n-1} P(w) \cdot (\mathbb{1}^1)^t \\ &= \sum_{|w| \leq n-1} \Pr(q_i, w | q_1) \end{aligned}$$

If q_i is non accessible, then all the the probabilities $\Pr(q_i, w | q_1)$ are equal to zero, giving $\psi_i = 0$.

Consequently, if $\psi_i = 0$, from the proposition 3.14 follows that all ξ_j have i -th coordinate equal to zero, i.e.

$$(3.26) \quad \sum_{w \in R^*} \Pr(q_i, w | q_1) = \mathbb{1}^1 \sum_{w \in R^*} P(w) \cdot (\mathbb{1}^1)^t = \sum_{j=0}^{\infty} \xi_j \cdot (\mathbb{1}^1)^t = 0.$$

This proves that q_i is non accessible.

The following is also true.

COROLLARY 3.16 : Let M be an ISM and let $|Q| = s$.

M is accessible if and only if $\text{rank } L = s$.

EXAMPLE 3.17 (continuation of the example 3.9).

We shall use the vector ψ , given in the proposition 3.14, to locate all non accessible states in $\bar{M}$. The ISM $\bar{M}$ is the one from the example 3.9, to which we shall refer as M in this example. We get

$$\xi_0 = [1, 1, 0, 0, 0] , \quad \xi_1 = \mathbb{1}^1 SP = [1/2, 1/2, 0, 0]$$

$$\xi_2 = [1/2, 1/2, 0, 0, 1] \quad \text{and} \quad \xi_3 = [1/4, 1/4, 0, 1/2]$$

This shows that $\xi_3 \in L(2)$, i.e. $L(2) = L$. This gives that $\text{rank } L = 3$. Using the formula (3.24) we get

$$\psi = \begin{bmatrix} \frac{57}{20} & \frac{17}{20} & 0 & \frac{1}{5} \end{bmatrix} \text{ and we can see that } \psi_3 = 0.$$

ψ_3 is the only zero coordinate of ψ , i.e. q_3 is the only non accessible state in Q .

To construct an accessible ISM: $\bar{M} \sim M$, where M , as mentioned above, is the ISM from the example 3.9, requires the deletion of the 3rd row and column in all $P(\alpha)$, where $\alpha \in R$ and in τ . This gives

$$\bar{P}(\beta) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{5} & 0 & \frac{2}{5} \\ 0 & 0 & 0 \end{bmatrix}; \quad \bar{P}(\gamma) = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & \frac{1}{5} & 0 \end{bmatrix} \quad \text{and} \quad \bar{\tau} = \begin{bmatrix} 0 \\ \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}.$$

We can now see the pictorial representation of $\bar{M}$ as

Figure 3.26.

As a matter of convenience we shall add the following proposition.

PROPOSITION 3.18 . Let M be a coaccessible ISM. The state q_1 in Q is non accessible if and only if the i -th coordinate of the vector $\zeta = [\zeta_i]$, given as

$$(3.27) \quad \zeta = \mathbb{I}^1 (I - SP)^{-1}$$

equals to zero.

PROOF .

From the proposition 3.14, it follows that instead of the vector space $L(n-1)$ we may use the vector space L , for they are equal. This allows us to replace the vector ψ , given by (3.24), by the vector

$$\zeta = \sum_{n=0}^{\infty} \xi_n = \mathbb{I}^1 \sum_{i=0}^{\infty} SP^i = \mathbb{I}^1 (I - SP)^{-1}$$

using the proposition 3.3.

Let us introduce the following notion.

DEFINITION 3.19 . Let M be an ISM. If M is both, accessible and coaccessible, we shall call it an initiated stochastic trim.

Let us conclude this chapter by the following trimming algorithm.

Step 1 . Let $M = (Q, q_1, P)$. If $\det(I - SP) \neq 0$, put $\bar{M} = M$ and proceed to the step 4. Otherwise continue.

Step 2 . Find the vector $\psi = \sum_{j=0}^{s-2} SP^j \cdot \tau = [\psi_i]$ and continue.

Step 3 . Let $\bar{Q} = \{q_i \in Q : \psi_i > 0\}$. If $q_1 \notin \bar{Q}$, put $M_t = (\{q_1\}, q_1, \phi.)$ and stop.

Otherwise put

$$\lambda_i = \sum_{q_j \in \bar{Q}} p_{ij} + p_i$$

Change P to $\bar{P}$ by

$$\bar{p}_{ij}(\alpha) = p_{ij}(\alpha)/\lambda_i, \text{ for all } \alpha \in R \text{ and}$$

q_i, q_j in $\bar{Q}$. Delete all $p_{ij}(\alpha)$'s if q_i or q_j are not in $\bar{Q}$. Change τ to $\bar{\tau}$ by putting $\bar{p}_i = p_i/\lambda_i$, for all $q_i \in \bar{Q}$ and delete all other p_i 's .

Step 4 . Find the vector $\zeta = \tau^1 (I - SP)^{-1} = [\zeta_i]$. If $\zeta > 0$, put $M_t = \bar{M} = (\bar{Q}, q_1, \bar{P})$ and stop. Otherwise continue.

Step 5 . Let $Q_t = \{q_i \in \bar{Q} : \zeta_i > 0\}$. Delete all rows and columns in $\bar{P}(\alpha)$, where $\alpha \in R$ with addresses (i,j) , where q_i or q_j are not in Q_t . Also delete all entries $\bar{p}_i$ in $\bar{\tau}$ with addresses corresponding to q_i in Q_t . This way we get P_t . Put M_t equal to (Q_t, q_1, P_t) and stop.

4. STOCHASTIC TRANSDUCERS .

In this and the following chapters we shall study some types of ISMs. These special types will be the stochastic transducers and the two-way machines.

Let us describe the machine vocabulary X , the type R and then the set of generalized actions for the stochastic transducers.

Let $\Sigma = \{ \sigma_1, \sigma_2, \dots, \sigma_n \}$ be the set of input letters, called the input alphabet. The set Σ^* is called the input vocabulary.

In a similar way we shall introduce the output alphabet $\Gamma = \{ \gamma_1, \gamma_2, \dots, \gamma_m \}$ and the output vocabulary Γ^* .

The machine vocabulary X is given as

$$(4.1) \quad X = \Gamma^* \times \Sigma^* .$$

The machine vocabulary X is a monoid with the identity $1 = (\epsilon, \epsilon)$. The product concatenation $(g_1, s_1)(g_2, s_2)$ equals to $(g_1 g_2, s_1 s_2)$, where g_1, g_2 are output words and s_1, s_2 are input ones, is the binary operation on X .

The set R contains four types of relations on X .

$$(4.2) \quad R = \{ R_{\gamma} x L_{\sigma}^{-1}, R_{\gamma} x I, I x L_{\sigma}^{-1}, I x I : (\gamma, \sigma) \in \Gamma \times \Sigma \} .$$

These relations are defined as follows

$$(4.3) \quad \begin{aligned} (g, \omega s) R_{\gamma} x L_{\sigma}^{-1} &= (g\gamma, s) & \text{if } \sigma = \omega \\ &= \phi & \text{otherwise} \end{aligned}$$

$$(4.4) \quad \begin{aligned} (g, \omega s) Ix L_{\sigma}^{-1} &= (g, s) & \text{if } \sigma = \omega \\ &= \phi & \text{otherwise} \end{aligned}$$

$$(4.5) \quad (g, s) R_{\gamma} x I = (g\gamma, s) \quad , \quad \text{and}$$

$$(4.6) \quad (g, s) Ix I = (g, s)$$

The symbols g , s , γ , σ , ω are

$$g \in \Gamma^* \quad , \quad s \in \Sigma^* \quad , \quad \gamma \in \Gamma \quad , \quad \sigma , \omega \in \Sigma$$

The above notation is adopted from (10). It is to be noted that the actions from R are used here as right operators.

We can realize that $R_{\gamma} x L_{\sigma}^{-1}$, and $Ix L_{\sigma}^{-1}$, as described in (4.3) and (4.4), are partial functions. The $Ix I$, as given in (4.6), is the identity relation on X .

DEFINITION 4.1 . Let X be given by (4.1) and R by (4.2) using the formulae (4.3) to (4.6). The initiated stochastic X - machine of type R , given as $T = (Q , q_1 , P)$, is called an initiated stochastic transducer (IST) . .

If $P(Ix I) \neq 0$ then the IST T is singular. We can use the theorem 2.14 to realize the following.

PROPOSITION 4.2 . Let $T = (Q, q_1, P)$ be a singular IST and let $\lim P^n(IxI) = 0$, then there exists a non singular IST $\bar{T} = (Q, q_1, \bar{P})$ such that $\bar{T}$ is equivalent to T .

To appreciate the importance of the above proposition 4.2 we shall first describe the process of computation in an IST. Let us now define the generalized actions from R^* inductively as follows.

Let $g = \bar{g}\gamma$ be an output word and $s = \bar{s}\sigma$ and input one. The γ and σ are output and input letters. We have

$$(4.7) \quad R_g x L_s^{-1} = (R_{\bar{g}} x L_{\bar{s}}^{-1}) \circ (R_{\gamma} x L_{\sigma}^{-1}) = (R_{\bar{g}} x L_{\bar{s}}^{-1}) \circ (I x L_{\sigma}^{-1}) \\ = (R_{\bar{g}} x I) \circ (R_{\gamma} x I)$$

$$(4.8) \quad R_g x I = (R_{\bar{g}} x I) \circ (R_{\gamma} x I), \text{ and}$$

$$(4.9) \quad I x L_s^{-1} = (I x L_{\bar{s}}^{-1}) \circ (I x L_{\sigma}^{-1})$$

Let us from now on use the notation $P(\gamma, \sigma)$ instead of $P(R_{\gamma} x L_{\sigma}^{-1})$ and $P(g, s)$ instead of $P(R_g x L_s^{-1})$.

The complexity of the formula (4.7) introduces the basic difficulty in the computation of the substochastic matrix $P(g, s)$, necessary for the calculation of the probability

$$(4.10) \quad p_T(R_g x L_s^{-1}) = \mathbb{1}^1 \cdot P(g, s) \cdot \tau$$

We have

$$(4.11) \quad P(g,s) = P(\bar{g},\bar{s}) \cdot P(\gamma,\sigma) + P(g,\bar{s}) \cdot P(\epsilon,\sigma) + P(\bar{g},s) \cdot P(\gamma,\epsilon) \quad .$$

On the other hand the simplicity of (4.5) and (4.6) offer a straight-forward computation of $P(g,\epsilon)$ and $P(\epsilon,s)$ as

$$(4.12) \quad P(g,\epsilon) = P(\gamma_1 \dots \gamma_k, \epsilon) = P(\gamma_1, \epsilon) \cdot \dots \cdot P(\gamma_k, \epsilon) \quad \text{and}$$

$$(4.13) \quad P(\epsilon,s) = P(\epsilon, \sigma_1 \dots \sigma_l) = P(\epsilon, \sigma_1) \cdot \dots \cdot P(\epsilon, \sigma_l) \quad .$$

The computation described by (4.10), although long, is always finite. This is due to the proposition 4.2, without which we might have encountered some infinite paths.

In the view of the trimming algorithm, described in the chapter 3, and the proposition 4.2, we shall assume that the IST T is a nonsingular stochastic trim. The measure p_T on $F(R^*)$ is a probabilistic measure.

We shall also, as a matter of convenience, use the following abbreviated notation.

$$(4.14) \quad p_T(R_g \cdot L_s^{-1}) = p_T(g,s) \quad .$$

This introduces a measure on the machine vocabulary X and due to the above remarks gives

$$(4.15) \quad \sum_{(g,s) \in X} p_T(g,s) = 1 \quad .$$

The machine vocabulary is different from the input one Σ^* as well as from the output vocabulary Γ^* . We shall not accept that the IST requires an input as a machine word and gives a simi-

lar output. The following translation shall remove this problem.

DEFINITION 4.3 . The mapping $\Phi : \Sigma^* \rightarrow X$ defined as

$s\Phi = (\epsilon, s)$ is called an input code .

The partial function $\Omega : X \rightarrow \Gamma^*$, given as

$(g, s)\Omega = g$ if $s = \epsilon$ and

$= \phi$ otherwise

is called an output code .

Let us introduce a relation $\Lambda_T : \Sigma^* \rightarrow \Gamma^*$ as follows.

$$(4.16) \quad g \in \Lambda_T(s) \text{ if and only if } p_T(g, s) > 0, \text{ i.e.}$$

$$(4.17) \quad \Lambda_T(s) = \{ g \in \Gamma^* : ((s\Phi)R_X L_s^{-1})\Omega = g \text{ and } p_T(g, s) > 0 \} .$$

We may give a pictorial representation of the above as

Figure 4.18

It is clear from the definition of Λ_T , as well as, from the figure 4.18, that Λ_T is a "one-to-many mapping" . This is caused by the fact that $p_T(g, s) \cdot p_T(\cdot, s)$ may be positive, although the

output words g and h are different. It is also possible that for some $s \in \Sigma^*$ the set $\Lambda_T(s) = \emptyset$. In this case we shall say that the input word s is not accepted by T . We shall clarify this, and a similar situation for $g \in \Gamma^*$ not to be in any $\Lambda_T(s)$, in the following definition.

DEFINITION 4.4 . Let $s \in \Sigma^*$ and $g \in \Gamma^*$ be input and output words respectively. Let T be an IST. We shall say that s is accepted by T if $\Lambda_T(s) \neq \emptyset$.

In a similar way we shall say that g is realized by T if there exists an accepted input word s such that $g \in \Lambda_T(s)$.

The subset of Σ^* : $D_T = \{ s \in \Sigma^* : \Lambda_T(s) \neq \emptyset \}$ is called the domain of T and the subset of Γ^* : $R_T = \{ g \in \Gamma^* : \Lambda_T^{-1}(g) \neq \emptyset \}$ is called its range.

The inverse Λ_T image of g , as used above, is given as

$$(4.19) \quad \Lambda_T^{-1}(g) = \{ s \in \Sigma^* : g \in \Lambda_T(s) \} \quad \text{and} \quad g \in \Gamma^*$$

We shall construct two ISMs, I_T and O_T such that their production sets will be D_T and R_T respectively. On top of this, these ISMs will also give the probabilities $p_D(s)$ of $s \in \Sigma^*$ being accepted by T and $p_R(g)$ of $g \in \Gamma^*$ being realized by T respectively.

PROPOSITION 4.5 . Let T be a coaccessible IST. Let $s \in \Sigma^*$ be in the domain and $g \in \Gamma^*$ be in the range of T . The probabilities $p_D(s)$ of s

being accepted and $p_R(g)$ of g being realized by T , satisfy the following relations

$$(4.20) \quad p_D(s) = \sum_{g \in \Lambda_T(s)} p_T(g, s)$$

$$(4.21) \quad p_R(g) = \sum_{s \in \Lambda_T^{-1}(g)} p_T(g, s)$$

The functions p_D and p_R are probability measures on Σ^* and Γ^* with the supports D_T and R_T respectively.

PROOF .

The formulae (4.20) and (4.21) follow directly from the definitions of $s \in \Sigma^*$ to be accepted and $g \in \Gamma^*$ to be realized by T . It is perhaps worth mentioning that $s \in D_T$ and $g \in R_T$ if and only if $p_T(g, s) \neq 0$. It is also clear that the two events (g, s) and (h, s) , where $g \neq h$, are exclusive. To prove that p_D and p_R are probability measures on Σ^* and Γ^* respectively is to show that

$$(4.22) \quad \sum_{s \in \Sigma^*} p_D(s) = 1 \quad \text{and}$$

$$(4.23) \quad \sum_{g \in \Gamma^*} p_R(g) = 1$$

If we use (4.20) and (4.21) we get

$$\sum_{s \in \Sigma^*} p_D(s) = \sum_{s \in \Sigma^*} \sum_{g \in \Lambda_T(s)} p_T(g, s) = \sum_{(g, s) \in X} p_T(g, s) \quad \text{and}$$

$$\sum_{g \in \Gamma^*} p_R(g) = \sum_{g \in \Gamma^*} \sum_{s \in \Lambda_T^{-1}(g)} p_T(g, s) = \sum_{(g, s) \in X} p_T(g, s)$$

The above equalities follow from the fact that $s \in \Sigma^* - D_T$ give $p_T(g, s) = 0$ and the same is true for $g \in \Gamma_1^* - R_T$.

From the assumed coaccessibility of T we get (4.22) and (4.23).

Let us use the formulae (4.20) and (4.21) to construct the above mentioned ISMs I_T and O_T . Let $s \in \Sigma^*$ be an output word in the form of

$$s = \sigma_1 \sigma_2 \dots \sigma_{n-1} \sigma_n = \bar{s} \sigma_n, \text{ where } \bar{s} = \sigma_1 \sigma_2 \dots \sigma_{n-1}$$

We have then the following

$$(4.24) \quad p_D(s) = \mathbb{1} \sum_{g \in \Gamma^*} P(g, s) \cdot \tau$$

From (4.11), it follows that

$$\sum_{g \in \Gamma^*} P(g, s) = \sum_{g \in \Gamma^*} P(g, \bar{s}) \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \sigma_n) + \sum_{g \in \Gamma^*} P(g, s) \sum_{\gamma \in \Gamma} P(\gamma, \epsilon)$$

It is clear that the above can be rewritten as

$$\sum_{g \in \Gamma^*} P(g, s) = I - \sum_{\gamma \in \Gamma} P(\gamma, \epsilon) = \sum_{g \in \Gamma^*} P(g, \bar{s}) \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \sigma_n)$$

If we use the assumption of coaccessibility of T , we get

$$(4.25) \quad p_D(s) = \mathbb{1} \sum_{g \in \Gamma^*} P(g, \bar{s}) \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \sigma_n) \cdot (I - \sum_{\gamma \in \Gamma} P(\gamma, \epsilon))^{-1} \cdot \tau$$

Let us define matrix T by

$$(4.26) \quad T = (I - \sum_{\gamma \in \Gamma} P(\gamma, \epsilon))^{-1} = \sum_{n=0}^{\infty} (\sum_{\gamma \in \Gamma} P(\gamma, \epsilon))^n.$$

Then

$$(4.27) \quad p_D(s) = \mathbb{1}^1 \cdot T \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \sigma_1) \cdot \dots \cdot T \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \sigma_n) \cdot T \cdot \mathbb{1}.$$

This suggests the following definition of P_I .

$$(4.28) \quad P_I(\sigma) = T \sum_{\gamma \in \Gamma \cup \{\epsilon\}} P(\gamma, \epsilon) \quad \text{for all } \sigma \in \Sigma \text{ and}$$

$$(4.29) \quad \tau_I = T \cdot \tau.$$

This proves the following.

PROPOSITION 4.6 . Let $T = (Q, q_1, P)$ be a coaccessible IST. Let $I_T = (Q, q_1, P_I)$, where P_I is described by (4.28). Then the production set $(I_T)_P = D_T$, where D_T is the domain of T , and the probability measure p_D is defined as

$$(4.30) \quad p_D(\sigma_1 \dots \sigma_n) = \mathbb{1}^1 \cdot P_I(\sigma_1) \cdot \dots \cdot P_I(\sigma_n) \cdot \tau_I.$$

It is clear that the type R_I of the initiated stochastic Σ^* -machine is

$$(4.31) \quad R_I = \{ R_\sigma : \sigma \in \Sigma \}, \text{ where } R_\sigma : sR_\sigma = s\sigma$$

and the free monoid of generalized actions R_I^* is

$$(4.32) \quad R_I^* = \{ R_s : s \in \Sigma^* \}, \text{ with } R_s : \bar{s} R_s = \bar{s} s.$$

The actions R_0 , as well as, the generalized actions R_s are relations on Σ^* . The homomorphism F , as defined by (1.14) is therefore a bijection, i.e. $F(R^*)$ is also a free monoid. From the comment 1.5, it follows that I_T is an IPA. This is also clear from the formula (4.30) giving the unique way of calculating the probabilities $p_D(s)$.

In a similar way we may prove the following.

PROPOSITION 4.7. Let $T = (Q, q_1, P)$ be a coaccessible IST. Let $O_T = (Q, q_1, P_0)$, where P_0 is

$$(4.33) \quad P_0(\gamma) = \psi \sum_{\sigma \in \Sigma \cup \{\epsilon\}} P(\gamma, \sigma), \text{ for all } \gamma \in \Gamma, \text{ where}$$

$$(4.34) \quad \psi = (I - \sum_{\sigma \in \Sigma} P(\epsilon, \sigma))^{-1} = \sum_{n=0}^{\infty} (\sum_{\sigma \in \Sigma} P(\epsilon, \sigma))^n.$$

The production set $\langle O_T \rangle_P = R_T$, the range of T . The probability measure p_R is given by

$$(4.35) \quad p_R(\gamma_1 \dots \gamma_n) = \tau_0^1 \cdot p_0(\gamma_1) \cdot \dots \cdot p(\gamma_n) \cdot \tau_0.$$

From (4.33) follows that

$$(4.36) \quad \tau_0 = \psi \cdot \tau.$$

The type R_0 of the initiated stochastic Γ^* -machine O_T is

$$(4.37) \quad R_0 = \{ R_\gamma : \gamma \in \Gamma \}$$

and the free monoid of generalized actions, R_0^* is

$$(4.38) \quad R_0^* = \{ R_g : g \in \Gamma^* \}$$

From (4.35), it follows that O_T is an IPA.

EXAMPLE 4.7 .

Let $T = (Q, q_1, P)$ be the IST, given by

$$Q = \{q_1, q_2, q_3\}, \quad \Sigma = \{\rho, \sigma\} \quad \text{and} \quad \Gamma = \{\gamma, \delta\}$$

This gives the type R as

$$R = \{ \gamma ; \delta ; \rho ; \sigma ; (\gamma, \rho) ; (\gamma, \sigma) ; (\delta, \rho) ; (\delta, \sigma) \}$$

where we use the following abbreviation

$$\gamma \text{ for } R_{\gamma} x I, \quad \sigma \text{ for } I x L_0^{-1} \text{ and } (\gamma, \sigma) \text{ for } R_{\gamma} x L_0^{-1}$$

The mapping P is given by

$$P(\gamma, \sigma) = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad P(\delta) = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P(\delta, \rho) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}; \quad P(\gamma, \rho) = \begin{bmatrix} 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$P(\gamma) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix} ; P(\sigma) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{4}{5} & 0 \end{bmatrix} ;$$

$$P(\rho) = P(\delta, \sigma) = 0$$

This gives the vector T as

$$T = \begin{bmatrix} 0 \\ \frac{3}{4} \\ 0 \end{bmatrix}, \text{ and the following pictorial representation of } T$$

Figure 4.39

In order to find the domain D_T and the range R_T of the IST T we shall construct the input and the output automata I_T and O_T . The matrices T and Ψ , given by (4.26) and (4.34) are

$$T = [I - (P(\gamma) + P(\delta))]^{-1} = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and}$$

$$\Psi = [I - (P(\rho) + P(\sigma))]^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{4}{5} & 1 \end{bmatrix}$$

To get the mappings P_I and P_O , as well as, the vectors τ_I and τ_O , we shall use the construction given by the formulae (4.28), (4.29), (4.33) and (4.36).

$$P_I(\rho) = T \cdot (P(\rho) + P(\gamma, \rho) + P(\delta, \rho)) = \begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{10} \\ 0 & 0 & \frac{1}{20} \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

$$P_I(\sigma) = \begin{bmatrix} \frac{1}{3} & \frac{2}{5} & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & \frac{4}{5} & 0 \end{bmatrix}, \quad \text{giving} \quad \tau_I = \begin{bmatrix} 0 \\ \frac{3}{4} \\ 0 \end{bmatrix}$$

$$P_O(\gamma) = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{5} \end{bmatrix} = \Psi \cdot (P(\gamma) + P(\gamma, \rho) + P(\gamma, \sigma)),$$

$$P_O(\delta) = \begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}, \quad \text{giving} \quad \tau_O = \begin{bmatrix} 0 \\ \frac{3}{4} \\ \frac{3}{5} \end{bmatrix}$$

The pictorial representation of the IPA I_T is

Similarly we get the pictorial representation of the O_T as

To find the probability $p_T(\gamma\gamma\delta, \sigma\sigma\sigma)$ of accepting the word $\sigma\sigma\sigma$ and realizing the word $\gamma\gamma\delta$ by our T , requires, in general, a long tedious calculation. This is indicated

by the formula (4.11) and appears in this case as follows

$$p_T(\gamma\gamma\delta, \sigma\sigma\sigma) = 1^1 (P^2(\gamma)P(\delta)P^3(\sigma) + P^2(\gamma)P(\delta, \sigma)P^2(\sigma) + \dots + P^2(\gamma, \sigma)P(\delta, \sigma)) \cdot \tau$$

Fortunately in this situation, a consideration of the figure 4.39 indicates, that the above complication can be avoided. It is easy to see that the only path with the required label is

$$q_1 \xrightarrow{(\gamma, \sigma)/\frac{1}{3}} q_1 \xrightarrow{(\gamma, \sigma)/\frac{1}{3}} q_1 \xrightarrow{\delta/\frac{1}{2}} q_3 \xrightarrow{\sigma/\frac{1}{3}} q_2 \xrightarrow{\frac{3}{4}} t$$

This gives $p_T(\gamma\gamma\delta, \sigma\sigma\sigma) = \frac{1}{30}$

It is possible, using the input IPA I_T and the output one O_T , to find the following conditional probabilities :

$$p_T(\gamma\gamma\delta | \sigma\sigma\sigma) = p_T(\gamma\gamma\delta, \sigma\sigma\sigma) / p_D(\sigma\sigma\sigma)$$

which is the conditional probability of $\gamma\gamma\delta$ being realized if $\sigma\sigma\sigma$ has been already accepted, and

$$p_T(\sigma\sigma\sigma | \gamma\gamma\delta) = p_T(\gamma\gamma\delta, \sigma\sigma\sigma) / p_R(\gamma\gamma\delta)$$

the conditional probability of $\sigma\sigma\sigma$ being accepted if we know that $\gamma\gamma\delta$ is realized by our IST T .

We still need to calculate the probabilities $p_D(\sigma\sigma\sigma)$ and $p_R(\gamma\gamma\delta)$. These probabilities are easy to calculate through the input IPA I_T and the output one O_T as

$$p_D(\sigma\sigma\sigma) = \mathbb{I}^1 \cdot P_I^3(\sigma) \cdot \tau_I = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} & \frac{2}{5} & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & \frac{4}{5} & 0 \end{bmatrix}^3 \cdot \begin{bmatrix} 0 \\ \frac{3}{4} \\ 0 \end{bmatrix} = \frac{49}{750}$$

and

$$p_R(\gamma\gamma\delta) = \mathbb{I}^1 \cdot P_O^2(\gamma) \cdot P_O(\delta) \cdot \tau_O = \frac{23}{600}$$

The required conditional probabilities are equal to

$$p_T(\gamma\gamma\delta|\sigma\sigma\sigma) = \frac{25}{49} \quad \text{and} \quad p_T(\sigma\sigma\sigma|\gamma\gamma\delta) = \frac{20}{23}$$

The formula (4.11) generally leads to long and tedious calculations, although in our example 4.7 we were quite lucky. Let us now view this under scrutiny.

Let (g, s) be an event such that

$$(4.42) \quad g = \gamma_1 \gamma_2 \dots \gamma_m \quad \text{and} \quad s = \sigma_1 \sigma_2 \dots \sigma_n$$

Let $\mathcal{E}(g, s)$ be the set of all successful paths with the label (g, s) . The length ℓ of any path from $\mathcal{E}(g, s)$ satisfies

$$(4.43) \quad \max(m, n) \leq \ell \leq m + n$$

In order to fully appreciate the complexity of the formula (4.11) we shall now calculate the cardinality of the set

$$F^{-1}(g, s) = \{ w \in R^* : F(w) = (g, s) \}$$

Let us follow the process of computation of the event (g, s)

and introduce the following states.

a/ We shall start with the input word s of length n . The IST then starts its computation with the machine word (ϵ, s) . We shall associate the state $(0, n)$ with this initial machine word to reflect the length of ϵ and s .

b/ In the process of computation we shall get some machine words $(\bar{g}, \bar{s})$, where the lengths i of $\bar{g}$ and j of $\bar{s}$ satisfy the inequalities

$$(4.44) \quad 0 \leq i \leq m \quad \text{and} \quad 0 \leq j \leq n$$

We shall associate the state (i, j) with such a machine word.

c/ We want the computation to end with the output word g of length m , i.e. the IST shall end its computation with the machine word (g, ϵ) . This machine word is associated with the state $(m, 0)$.

We have now an initial state $(0, n)$, transient states (i, j) and a terminal state $(m, 0)$. From the inequalities (4.44) it follows that we have all together $(m+1)(n+1)$ different states. We shall order these states lexicographically increasing with the first coordinate i and decreasing with the second coordinate j as follows

$$(4.45) \quad (0, n) < (0, n-1) < \dots < (0, 0) < (1, n) < \dots < (m, 0)$$

If we index all the states according to the order described by (4.45) our task becomes one in which we must find out how many different ways lead from the first to the last state. To find this

we shall use the following one step transition matrix S .

$$(4.46) \quad S = [s_{uv}] ; \quad u, v = 1, \dots, (m+1)(n+1),$$

where s_{uv} is the number of ways from state u to state v . From the state (i, j) we can go to $(i, j-1)$ if $j > 0$ or to $(i+1, j)$ if $i < m$ and also to the state $(i+1, j-1)$ if $i < m$ and $j > 0$. This shows that the matrix S has the following structure.

$$(4.47) \quad S = \begin{bmatrix} A & B & 0 & \dots & 0 & 0 \\ 0 & A & B & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & A & B \\ 0 & 0 & 0 & \dots & 0 & A \end{bmatrix},$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

and $B = I + A$.

Each row of S in (4.47) consists of $m+1$ submatrices, all of the order $n+1$. From (4.47), it follows that A and B commute and $A^i = 0$ for all $i > n$. This shows that

$$(4.48) \quad S^i = \begin{bmatrix} A^i & iA^{i-1}B & \dots & \binom{i}{j}A^{i-j}B^j & \dots \\ 0 & A^i & \dots & \binom{i}{j-1}A^{i-j+1}B^{j-1} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & A^i \end{bmatrix}$$

where $\binom{i}{j}$ are the binomial coefficients, i.e. $S^i = 0$ for all $i > m+n$.

If we denote the last submatrix in the first row of S^i by S_i , then the first time $S_i \neq 0$ occurs when $i = m$. From (4.48), it follows that

$$(4.49) \quad S_m + S_{m+1} + \dots + S_{m+n} = B^m \sum_{i=0}^n \binom{m+i}{m} A^i$$

$$= \sum_{j=0}^m \binom{m}{j} \sum_{i=0}^n \binom{m+i}{m} A^{i+j}$$

To follow our original task, we shall only count the entries with addresses $(1, n+1)$ of the summands in (4.49). The only matrix giving this entry different from zero is coming from A^n . In this case the entry is equal to 1. We shall now examine two different situations.

a/ $m \geq n$ and b/ $n > m$.

If we put $i+j = n$ in the formula (4.49) we shall get the following two results.

$$a/ \quad |F^{-1}(g, s)| = \sum_{i=0}^n \binom{m}{n-i} \binom{m+i}{m} \quad \text{if } m \geq n, \text{ and}$$

$$b/ \quad |F^{-1}(g,s)| = \sum_{i=0}^m \binom{m}{i} \binom{m+n-1}{m-i} \quad \text{if } n > m,$$

where $|F^{-1}(g,s)|$ denotes the cardinality of the set of all generalized actions producing the event (g,s) satisfying (4.42).

As we have already seen in the example 4.7, it is possible that many of the paths encountered in the formulae for the cardinality of $F^{-1}(g,s)$ shall not be successful, making the actual computation of the $p_T(g,s)$ shorter. This shows therefore that the right sides in these formulae are only the upper bounds for the cardinality of the set of all "successful" generalized actions producing the event (g,s) .

Looking at the formula (4.43) we see that there are two "extremal" kinds of paths in $\mathcal{E}(g,s)$:

a/ of length $\ell = \max(m,n)$

b/ of length $\ell = m + n$

In the first case we arrived at such paths $c \in \mathcal{E}(g,s)$ which take input word s and change it as fast as possible into the output word g . This leads to the following special type of ISTs.

DEFINITION 4.8 . Let $T = (Q, q_1, \cdot P)$ be an IST of type R , where the set of actions R is given as

$$(4.50) \quad R = \{ R_{\gamma} x L_{\sigma}^{-1} : (\gamma, \sigma) \in \Gamma \times \Sigma \}$$

Then T is an accelerated initiated stochastic transducer (AIST).

The name AIST comes from the fact that all paths satisfy the equality $a/$. Unfortunately every AIST has limited power in its computation, stated in the following proposition.

PROPOSITION 4.8 . Let T be an AIST. Then for every event $(g,s) \in T_p$, it is true that $|g| = |s|$.

The mapping $\Lambda_T : \Sigma \rightarrow \Gamma$ in this case is therefore length-preserving.

Let $g = \gamma_1 \gamma_2 \dots \gamma_n$ and $s = \sigma_1 \sigma_2 \dots \sigma_n$. Let T_p be the production set of the AIST T . Let $(g,s) \in T_p$. Then

$$(4.51) \quad p_T(g,s) = \mathbb{1} \cdot P(\gamma_1, \sigma_1) \cdot \dots \cdot P(\gamma_n, \sigma_n).$$

From the formula (4.51), it follows that all the paths in $\mathcal{E}(g,s)$ are of length $\ell = n$. It is also clear that the mapping F is an isomorphism and $F(R^*)$ is a free monoid. We conclude with the following.

COROLLARY 4.9 . Every AIST is an IPA.

As stated in the proposition 4.8, every AIST computes only words with input and output words of the same length. We shall therefore call such machines length-preserving ISTs. It is obvious that if an IST is not length-preserving, then it is not possible to find an AIST which is even b-equivalent to the first one. The question is now as follows.

Is there always an AIST which is equivalent to a given IST? The answer is affirmative and is stated in the following theorem.

THEOREM 4.10 . Let $T = (Q, q_1, P)$ be a length-preserving IST. It is possible to construct an equivalent $\bar{T}$ which is already an AIST.

PROOF

We shall produce a construction of $\bar{T}$ equivalent to T . The construction will start with the following tagging process.

a/ Tag the initial state q_1 by 0.

b/ If $q_i \in Q$ has already a tag n and q_j is directly connected with q_i by an edge e_{ij}^α , then

(i) tag q_j by n , if $\alpha = R_Y \times L_O^{-1}$,

(ii) tag q_j by $n+1$, if $\alpha = R_Y \times I$, and

(iii) tag q_j by $n-1$, if $\alpha = I \times L_O^{-1}$.

We shall now prove that such tagging is always possible. In other words we shall show that there is no chance to follow a path c_1 from q_1 to q_1 and get a tag n_1 , and by following another path c_2 end up with another tag n_2 . This, due to the assumption that T is length-preserving, is not possible. To show this, we shall assume that T is a trim. The accessibility gives that every state shall be tagged and the coaccessibility guaranties a path c_3 from q_1 to t . This gives us two successful paths

$c_1 c_3$ with label (g_1, s_1) and

$c_2 c_3$ with label (g_2, s_2) .

Realizing that T is length-preserving we have

$$(4.52) \quad |g_1| - |s_1| = |g_2| - |s_2| = 0$$

If c_1 has label $(\bar{g}_1, \bar{s}_1)$, c_2 has label $(\bar{g}_2, \bar{s}_2)$ and c_3 has the one equal to $(\bar{g}_3, \bar{s}_3)$ then from the tagging procedure, it follows that

$$(4.53) \quad |\bar{g}_1| - |\bar{s}_1| = n_1 \quad \text{and} \quad |\bar{g}_2| - |\bar{s}_2| = n_2$$

We also have $g_1 = \bar{g}_1 \bar{g}_3$, $s_1 = \bar{s}_1 \bar{s}_3$ and $g_2 = \bar{g}_2 \bar{g}_3$, $s_2 = \bar{s}_2 \bar{s}_3$ giving

$$\begin{aligned} (4.54) \quad n_1 &= |\bar{g}_1| - |\bar{s}_1| = |g_1| - |\bar{g}_3| - |s_1| + |\bar{s}_3| \\ &= |g_2| - |\bar{g}_3| - |s_2| + |\bar{s}_3| \\ &= |\bar{g}_2| - |\bar{s}_2| = n_2 \end{aligned}$$

This justifies our tagging procedure.

The tagging defines on Q equivalent classes Q_n given as $\{q \in Q : \text{tag of } q \text{ is } n\}$. Let us now define the integer N as follows.

- a/ If there are some positive tags, then $N = \max n$,
- b/ otherwise put $N = \min n$.

If at any time we get $N = 0$, we shall terminate the process, because our new machine is already an AIST. This implicates that our effort shall be aimed towards the reduction of Q_N , when $N \neq 0$, until $Q_N = \phi$ and N has to be redefined. This process shall then eventually end with desired $N = 0$.

Let us therefore take any $q \in Q_N$. If $N > 0$ then we may picture the situation as shown in the figure 4.55.

22

OF/DE

Figure 4.55

From the tagging procedure and from the definition of N , it follows that there are no edges terminating at q with labels of type IxL_O^{-1} and no edges of type $R_Y xI$ initiating at q .

We shall now replace the state $q \in Q_N$ by two states q' and q'' as indicated in the figure 4.56. This process shall produce the iterate, T_i with q' and q'' in Q_{N-1} and reduce the cardinality of Q_N by one. The change of edges, as well as the change of states, as seen from the figures 4.55 and 4.56, preserves all labels of the paths with the required probabilities.

If we follow the paths from q_a to q_b in both figures 4.55 and 4.56, we get the same label (γ_1, σ_2) with the probability $p_1 p_2$. The same is true for $q_a \rightarrow q_c$ giving the label $(\gamma_1 \gamma_3, \sigma_3)/p_1 p_3$, for $q_d \rightarrow q_b$ with the label $(\gamma_4, \sigma_4 \sigma_2)/p_2 p_4$ and for $q_d \rightarrow q_c$ with the label $(\gamma_4 \gamma_3, \sigma_4 \sigma_3)$.

Figure 4.56

and the probability $p_3 p_4$

It is now time to make the following comments.

In the figure 4.55 we put only one edge of each type. It is though generally possible that there are more than one of the same type. The above process has then to be applied for each such group separately.

If, on the other hand, some types of edges are missing, we may imagine them with zero probabilities giving such edges in the figure 4.56. Such edges are then also missing from the new iterate

T_i

If some state q ; discussed above, has a loop then it can be simulated by a situation where $q = q_c = q_d$ in the figure 4.15.

If $T = (Q, q_1, P)$ is not a length preserving IST

then it can not be accelerated by an AIST. The question however remains.

Is there always an equivalent IST such that all its successful paths producing (g,s) are of length $\max(|g|, |s|)$?

We know already the answer for the length-preserving case. The question in general is quite complex and has not been solved in this thesis.

We shall now tackle this problem from exactly the opposite point of view. We are going to study whether it is worth while and possible to decelerate an IST, in other words, is it always possible to find an equivalent IST such that all its successful paths producing the event (g,s) shall be of length $|g| + |s|$. Such IST offers some advantages.

Let us first realize the advantages of such an IST. The IST, we are discussing now, produces events using the slowest ways, i.e. the longest paths. This IST is therefore of type R , where

$$(4.57) \quad R = \{ R_Y x I \quad , \quad I x L_{\sigma}^{-1} \quad : \quad (Y, \sigma) \in \Gamma \times \Sigma \}$$

These transducers do not use any action of type $R_Y x L_{\sigma}^{-1}$, because these actions operate too fast. The above IST then offers the following advantage.

The formula (4.11) is simplified to

$$(4.58) \quad P(g,s) = P(\bar{g},s) \cdot P(Y,\epsilon) + P(g,\bar{s}) \cdot P(\epsilon,\sigma)$$

where $g = \bar{g}Y$ and $s = \bar{s}\sigma$. The limitation of R by (4.57) shall decrease the number of paths in the process of computation

of the event (g, s)

In the computation of the event (g, s) we must use exactly m actions of the type R_{x1} and n actions of the type $I_{xL_0}^{-1}$. Each type by itself has to be used in a fixed order. The only choice we are left with is how to choose and where to place the m actions of type R_{x1} among all $m+n$ actions. This can be done in exactly $\binom{m+n}{n}$ different ways. We then get the following result.

PROPOSITION 4.11. Let $T = (Q, q_1, P)$ be an IST of type R , where R is given by the formula (4.57). Let $|g| = m$ and $|s| = n$. Then

$$(4.59) \quad |F^{-1}(g, s)| = \binom{m+n}{n}$$

If we compare (4.59) with the formulae giving the limitations for the cardinality of $F^{-1}(g, s)$ we can see that (4.59) reduces the sum in these estimates to one single term. We get only the last term ($i = n$) in the case of $m \geq n$ and only the first one ($i = 0$) in the case of $n > m$. This answers one part of our question and makes it worth while to study.

Before we tackle the existence part of our problem, let us introduce the following definition

DEFINITION 4.12. Let $T = (Q, q_1, P)$ be an IST of type R , where R is given by (4.57). We shall call T a decelerated initiated stochastic transducer (DIST).

In this terminology we can now state the following.

PROPOSITION 4.13. Let $T = (Q, q_1, P)$ be an IST. It is always possible to construct an equivalent DIST.

PROOF.

From the definition of DIST, it follows that we have to replace all the edges with labels of the type $R_Y X L^{-1}$ in T . All the other edges can remain untouched. This may be done as follows.

Leave all the states which do not initiate any edge with the label of type $R_Y X L^{-1}$.

Replace all the edges e_{ij}^α , with labels of type $R_Y X L^{-1}$, by a pair of edges $e_{ii(k)}^{\alpha_1}$ and $e_{i(k)j}^{\alpha_2}$, where $\alpha_1 = R_Y X I$ and $\alpha_2 = I X L^{-1}$. The probabilities are given by

$$(4.60) \quad p_{ii(k)}(\alpha_1) = p_{ij}(\alpha) \quad \text{and} \quad p_{i(k)j}(\alpha_2) = 1.$$

This requires the addition of some states $q_{i(k)}$. Each edge with the label $\alpha = R_Y X L^{-1}$ and initiating at q_i shall require one additional state. We get $k = 1, 2, \dots, \ell(i)$, where $\ell(i)$ is the number of edges with label $\alpha = R_Y X L^{-1}$ initiating at q_i .

All other edges shall then remain the same as in T .

From the construction, described in the proof of the proposition 4.13, it follows that the simplification of the computation of the probabilities $p_T(g, s)$ is more than compensated by an enlargement of the set of states Q . To illustrate this, let us continue with the example 4.7.

EXAMPLE 4.14

Let us decelerate our T , following the construction outlined in the proof of the proposition 4.13. We shall follow all the states in Q one by one.

q_1 . There are two edges with labels $R_Y x L_\sigma^{-1}$ and $R_Y x L_\rho^{-1}$. We have to add therefore two states to Q , $q_{1(1)}$ and $q_{1(2)}$ and make the following changes. We shall replace the edge $e_{11}^{(Y,\sigma)}$ by $e_{11(1)}^{(Y,\epsilon)}$ and $e_{1(1)1}^{(\epsilon,\sigma)}$ with probabilities $p_{11(1)}(Y,\epsilon) = p_{11}(Y,\sigma) = \frac{1}{3}$ and $p_{1(1)1}(\epsilon,\sigma) = 1$. Similarly we shall replace the edge $e_{12}^{(Y,\rho)}$ by $e_{11(2)}^{(Y,\epsilon)}$ and $e_{1(2)2}^{(\epsilon,\rho)}$ with probabilities $p_{11(2)}(Y,\epsilon) = p_{12}(Y,\rho) = \frac{1}{3}$ and $p_{1(2)2}(\epsilon,\rho) = 1$.

q_2 . No change.

q_3 . There is one edge with label $R_\delta x L_\rho^{-1}$. Therefore we have to add another state $q_{3(1)}$ and replace the edge $e_{33}^{(\delta,\rho)}$ by $e_{33(1)}^{(\delta,\epsilon)}$ and $e_{3(1)3}^{(\epsilon,\rho)}$ with the probabilities $p_{33(1)}(\delta,\epsilon) = p_{33}(\delta,\rho) = \frac{1}{3}$ and $p_{3(1)3}(\epsilon,\rho) = 1$.

This gives the required DIST $\bar{T} = (\bar{Q}, q_1, \bar{P})$, where $\bar{Q} = \{q_1, q_{1(1)}, q_{1(2)}, q_2, q_3, q_{3(1)}\}$ has cardinality six.

In $\bar{P}(R)$ we have "only" 4 substochastic matrices, but all of them are of order 6, given as follows.

$$\bar{P}(\gamma, \epsilon) = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad \bar{P}(\delta, \epsilon) = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\bar{P}(\epsilon, \delta) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \quad \bar{P}(\epsilon, \sigma) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

giving $\bar{T} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{4} \\ 0 \\ 0 \end{bmatrix}$

We may also easily realize that T is a length-preserving IST and therefore, as guaranteed by the theorem 4.10 we can accelerate it as follows.

a/ The tagging process gives us the equivalence classes $Q_0 = \{q_1, q_2\}$ and $Q_1 = \{q_3\}$.

b/ We shall start the acceleration process with $N = 1$ and remove q_3 from Q_1 . This gives the next N equal to zero and the end of the acceleration. If we use the notation introduced in the figure 4.55, we get $q_1 = q_a$, $q_2 = q_a = q_b$ and $q_3 = q = q_c = q_d$.

Figure 4.62

We shall realize that the state q_2 is acting as q_a if the action $R_{\gamma} x I$ is taken and as q_b if we take the action $I x L_0^{-1}$ into consideration. This comparison with the figure 4.56 gives the following.

$$q_1 = q_a, \quad q_2 = q_a = q_b, \quad \text{and} \quad q_3 = q' = q'' = q_c = q_d.$$

This shall therefore conclude in the following reduction as given in the figure 4.63.

Figure 4.63

We are now ready to construct the AIST : $T' = (Q, q_1, P')$.
The mapping P' is given as follows.

$$P'(\gamma, \rho) = \begin{bmatrix} 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{20} \\ 0 & 0 & 0 \end{bmatrix} ; \quad P'(\gamma, \sigma) = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & 0 \end{bmatrix} ;$$

$$P'(\delta, \rho) = \begin{bmatrix} 0 & 0 & \frac{1}{10} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} ; \quad P'(\delta, \sigma) = \begin{bmatrix} 0 & \frac{2}{5} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \end{bmatrix}$$

giving $T' = \begin{bmatrix} 0 \\ \frac{3}{4} \\ 0 \end{bmatrix}$

This enables us to present the pictorial representation of the resulting $T' \sim T$ as follows.

Figure 4.64

To conclude this example let us take an event $w = (\gamma\delta\delta, \sigma\rho\sigma)$ and calculate the probability $p_T(w)$.

a/ If we calculated the probability $p_T(w)$ directly from the original IST T , we would have to take into consideration all computations and work with the estimate

$$|F^{-1}(\gamma\delta\delta, \sigma\rho\sigma)| = \sum_{i=0}^3 \binom{3}{3-i} \binom{3+i}{3} = 63$$

If we take advantage of the fact that the substochastic matrices $P(\delta, \sigma)$ and $P(\epsilon, \rho)$ are equal to zero, we shall be still left with 10 different ways.

b/ Similarly the effort to calculate the probability $p_T(w) = p_{\bar{T}}(w)$ using the DIST : $\bar{T}$, would lead to

$$|F^{-1}(\gamma\delta\delta, \sigma\rho\sigma)| = \binom{6}{3} = 20$$

Unfortunately, in this case, all involved substochastic matrices are different from zero, i.e. the above estimate is accurate. On top of this, all involved matrices are of order 6.

c/ The computation becomes the simplest if we use the AIST T' . This way we get

$$p_T(w) = 1^1 \cdot P'(\gamma, \sigma) \cdot P'(\delta, \rho) \cdot P'(\delta, \sigma) \cdot \tau' = \frac{1}{50}$$

5. TWO-WAY STOCHASTIC MACHINES

In this chapter we shall introduce another type of ISMs called two-way stochastic machines.

Let us first introduce the basic notation. The input and the output alphabets are both given by the same finite set of letters Σ . The machine vocabulary X is therefore defined in the following way

$$(5.1) \quad X = \Sigma^* \times \Sigma^*$$

The set of actions R is given as

$$(5.2) \quad R = \{ R_{\sigma} x L_{\sigma}^{-1}, R_{\sigma}^{-1} x L_{\sigma}, IxI : \sigma \in \Sigma \}$$

The relations on X in R are defined as

$$(5.3) \quad \begin{aligned} (g, \rho s) R_{\sigma} x L_{\sigma}^{-1} &= (g\sigma, s) & \text{if } \rho = \sigma \\ &= \phi & \text{otherwise.} \end{aligned}$$

$$(5.4) \quad \begin{aligned} (g\rho, s) R_{\sigma}^{-1} x L_{\sigma} &= (g, \sigma s) & \text{if } \rho = \sigma \\ &= \phi & \text{otherwise.} \end{aligned}$$

The action IxI is the identity relation on X . The other relations are partial functions in general.

As a matter of convenience we shall now simplify the notation and write

$$(5.5) \quad \sigma \quad \text{instead of} \quad R_{\sigma} x L_{\sigma}^{-1}$$

(5.6) $\bar{\sigma}$ instead of $R_{\sigma}^{-1} \times L_{\sigma}$ and

(5.7) ϵ instead of $I \times I$. . .

DEFINITION 5.1 . Let X be given by (5.1) and R by (5.2).

The stochastic X -machine $W = (Q, q_1, P)$ of type R is called a singular initiated two-way stochastic machine (SITSM) .

If R does not contain the action ϵ , we shall call W an initiated two-way stochastic machine (ITSM) .

Before we describe the ITSM W in more details, we shall use the theorem 2.14 in our situation.

PROPOSITION 5.2 . Let $W = (Q, q_1, P)$ be a SITSM and let $\lim P^n(\epsilon) = 0$. Then there exists a non singular ITSM $\bar{W}$ and $\bar{W} \sim W$.

Consequently we shall consider only non singular ITSMs. To see what such ITSM does, we shall study the structure of the monoid $F(R^*)$.

$$(5.8) \quad F(R^*) = \{ v, \bar{v} : v \in \Sigma^* \}$$

The relations v and $\bar{v}$ are the extensions of σ and $\bar{\sigma}$

$$(5.9) \quad \begin{aligned} (g, us)v &= (gv, s) & \text{if } u = v \\ &= \phi & \text{otherwise.} \end{aligned}$$

$$(5.10) \quad (gu, s)\bar{v} = (g, vs) \quad \text{if } u = v \\ = \phi \quad \text{otherwise.}$$

The ITSM W , as defined above, operates as follows.

Figure 5.11

The tape, as illustrated in the figure 5.11, contains a machine word (g, s) and can move either way. If W performs the action σ , the tape stops or moves one step to the left. This is done with some probability which also depends on the change in states. Similarly the action $\bar{\sigma}$ stops or moves the tape one step to the right. The action ϵ does not move the tape at all. We may also see that

$$(5.12) \quad (g, s)v = \phi \quad \text{if the machine stops on the way, or}$$

$$(5.13) \quad (g, s)v = (g', s') \quad , \text{ where } gs = g's'$$

We shall also assume that the tape can only enter W from the right side, i.e. in the position (ϵ, s) and leave on the left side in the position (s, ϵ) .

Formally this can be described by

DEFINITION 5.3 . The mapping $\phi : \Sigma^* \rightarrow X$, defined as

$$(5.14) \quad s \phi = (\epsilon, s)$$

is called the input code . Similarly the output code $\Omega : X \rightarrow \Sigma^*$ is defined as

$$(5.15) \quad (g, s) \Omega = \begin{cases} \epsilon & \text{if } s = \epsilon \\ \phi & \text{otherwise.} \end{cases}$$

This introduces a mapping $\Delta : \Sigma^* \rightarrow \Sigma^*$ as a composition

$$(5.16) \quad \Delta = \phi \circ v \circ \Omega$$

This makes Δ operate as follows.

$$(5.17) \quad s \Delta = ((s \phi) v) \Omega = \begin{cases} s & \text{if } v = s \\ \phi & \text{otherwise.} \end{cases}$$

Each move of the tape is caused by a change of state and implementation of an action α . This is done with some probability found in a substochastic matrix $P(\alpha) \in P(R)$. This again introduces a probability $p_w(s)$ with which the tape s gets through the ITSM W . We may also write that

$$(5.18) \quad s \Delta \neq \phi \quad \text{with probability } p_w(s)$$

Using the figure 5.11 we see that ϕ puts the tape into the machine, while Ω takes it out.

We shall now adopt the following phraseology.

If (g, s) is a machine word, then the integer $|g|$ gives the position of the reading head. We can see that at the beginning, when the input machine word (ϵ, s) enters the initial state q_1 , the reading head is in the initial position, i.e. $|\epsilon|=0$. At the end, after it scans the whole input word s for the last time and ends at the terminal state t , the reading head is in its rightmost position $|s|$. The generalized actions move the reading head as follows,

$$s = R_s x L_s^{-1} : |s| \text{ positions to the right or stops,}$$

$$\bar{s} = R_s^{-1} x L_s : |s| \text{ positions to the left or stops,}$$

while ϵ does not move the reading head, and does not stop.

Let us now follow all the possible ways of computation which lead to the same event $s \in W_p$. We can see that in this case the set $F^{-1}(s)$ contains an infinite number of generalized actions. If we use the formula (1.16) in the computation of $p_W(s)$, we have to encounter all infinite number of ways involved in it. The mentioned sum in the formula (1.6) shall therefore include infinite number of terms and we have to find an effective way for its calculation. The problem of convergency is easy to resolve, because all involved matrices are substochastic. This ensures the upper bound equal to one for all entries.

Let us now approach the problem of finding the probability $p_W(s)$ in the formula (5.18). To do what we want to do, requires to watch the movement of the reading head all its way. The input word s is changed into the input machine word (ϵ, s) by the

input code Φ and enters the initial state q_1 . At this moment our reading head is at the initial position.

Let $S = \sigma_1 \sigma_2 \dots \sigma_n$. This gives no choice and we have to move the reading head one place to the right into the new position $|\sigma_1| = 1$. In order to move the reading head for the first time to the position $|\sigma_1 \sigma_2| = 2$, we have many choices. We can move it back and forth between 0 and 1 as many times as we want. This movement gives the identity relation ϵ , associated with the matrix

$$(5.19) \quad \sum_{i=0}^{\infty} (P(\bar{\sigma}_1) \cdot P(\sigma_1))^i = (I - P(\bar{\sigma}_1) \cdot P(\sigma_1))^{-1}$$

The inversion on the right side of the formula (5.19) is possible if we assume coaccessibility. Let us now introduce the following notation.

$$(5.20) \quad P^{\sigma} = \sum_{i=0}^{\infty} (P(\bar{\sigma}) \cdot P(\sigma))^i = (I - P(\bar{\sigma}) \cdot P(\sigma))^{-1},$$

for every $\sigma \in \Sigma$. Using (5.20) we can easily see that the relation σ_1 is associated with the the substochastic matrix $P(\sigma_1) \cdot P^{\sigma_1}$. After this, the reading head moves one place to the right for the first time, into the position $|\sigma_1 \sigma_2| = 2$. This single movement to the right is caused by the action σ_2 and is associated with the substochastic matrix $P(\sigma_2)$. The reading head, before moving another place to the right, can move back and forth, one or two places now, again as many times as it wants.

This is associated with the substochastic matrix

$$(5.21) \quad \sum_{i=0}^{\infty} (P(\bar{\sigma}_2) \cdot P^{\sigma_1} \cdot P(\sigma_2))^i = (I - P(\bar{\sigma}_2) \cdot P^{\sigma_1} \cdot P(\sigma_2))^{-1}$$

Let us now give the recursive description of the following matrices.

$$(5.22) \quad P^{g^0} = \sum_{i=0}^{\infty} (P(\bar{\sigma}) \cdot P^g \cdot P(\sigma))^i = (1 - P(\bar{\sigma}) \cdot P^g \cdot P(\sigma))^{-1},$$

with P^0 given by (5.20). The relation $\sigma_1 \sigma_2$, which brings the reading head into the position $|\sigma_1 \sigma_2| = 2$, is associated with the substochastic matrix $P(\sigma_1) \cdot P^0 \cdot P(\sigma_2) P^{\sigma_1 \sigma_2}$. At this point, the reading head is moving another step to the right for the first time. The ITSM now takes the action σ_3 associated with $P(\sigma_3)$ and the whole process shall repeat again, until the reading head reaches its terminal position $|s| = n$ for the last time. This is arranged by the relation s and is associated with the substochastic matrix $P(s)$ given by

$$(5.23) \quad P(s) = \prod_{i=1}^n P(\sigma_i) \cdot P^{\sigma_1 \sigma_2 \dots \sigma_i}$$

Before we conclude our derivation, it is necessary to mention the difference between the action σ , associated with the substochastic matrix $P(\sigma)$, and the relation σ , associated with the substochastic matrix $P(\sigma) \cdot P^0$.

THEOREM 5.4. Let $p_W(s)$ be the probability of $s \Delta \neq 0$.
Then

$$(5.24) \quad p_W(s) = \tau^1 \cdot P(s) \cdot \tau,$$

where $P(s)$ is given by (5.23) and uses the matrices P^g defined recursively by (5.22) and (5.20).

Let us now illustrate the actual calculation in the following

example.

EXAMPLE 5.5

Figure 5.25

As seen from the figure 5.25, the input and output alphabets consist of two letters, i.e. $\Sigma = \{ \rho, \sigma \}$. The set of actions R is of cardinality four.

$$R = \{ \rho, \sigma, \bar{\rho}, \bar{\sigma} \}$$

The ITSM $W = (Q, q_1, P)$, where P is giving the following substochastic matrices.

$$P(\rho) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 \end{bmatrix} ; \quad P(\sigma) = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} ;$$

$$P(\bar{\rho}) = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} ; \quad P(\bar{\sigma}) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This leaves us with $\tau = \begin{bmatrix} 0 \\ \frac{1}{3} \\ 0 \end{bmatrix}$

Let s be $\rho\sigma\sigma$. To find the probability $p_W(s)$ of $s \in \phi$ requires to find the matrices P^ρ , $P^{\rho\sigma}$, $P^{\rho\sigma\sigma}$. We shall use the formulae (5.20) and (5.22) and get

$$P^\rho = (I - P(\bar{\rho}) \cdot P(\rho))^{-1} = \begin{bmatrix} 1 & \frac{1}{22} & 0 \\ 0 & \frac{11}{11} & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$$P^{\rho\sigma} = (I - P(\bar{\sigma}) \cdot P^\rho \cdot P(\sigma))^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{11}{11} & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$$P^{\rho\sigma\sigma} = (I - P(\bar{\sigma}) \cdot P^{\rho\sigma} \cdot P(\sigma))^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{11}{11} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We have now all the required matrices for the computation of $p_W(s)$. Using the formula (5.24) we get

$$p_W(s) = 1 \cdot P(\rho) \cdot P^\rho \cdot P(\sigma) \cdot P^{\rho\sigma} \cdot P(\sigma) \cdot P^{\rho\sigma\sigma} \cdot \tau = \frac{1}{110}$$

We shall now conclude this chapter and realize that the theorem 5.4 solves the following decision problem.

PROPOSITION 5.6 . Let $W_p = \{ s \in \Sigma^* : p_W(s) > 0 \}$ be the production set of the ITSM W . There exists an

effective procedure which for each s in Σ^*
decides whether s belongs to W_p or not.
This procedure also gives the probability of
 $s \in \Delta \neq \phi$.

REFERENCES .

- (1) M.A.ARBIB and L.S.BOBROV : Discrete Mathematics for Computer and Information Science, Sanders, Philadelphia, 1974.
- (2) M.A.ARBIB : Algebraic Theory of Machines, Languages and Semigroups, Academic Press, 1968.
- (3) M.A.ARBIB and G.E.MANES : Fuzzy Morphisms in Automata Theory, Proceedings of the First International Symposium on Category Theory applied to Computation and Control, Univ. of Mass., 1974.
- (4) R.C.ATKINSON, G.H.BOWER and E.J.CROTHERS : An Introduction to Mathematical Learning Theory, Wiley, New York, 1965.
- (5) J.BRODY and S.KLAŠA : Reduction of Stochastic Machines, article in progress.
- (6) J.BRODY : Error Frequency a Regulator of the Teaching Algorithm, Kybernetika, No 6, Prague, 548-569, 1968.
- (7) J.BRODY : A Probabilistic Model of Learning by Means of Problems, Networks, No 3, 75-80, 1973.
- (8) R.R.BUSH and F.MOSTELER : Studies in Mathematical Learning Theory, Stanford Univ. Press, Stanford, California, 293-307, 1959.
- (9) K.L.CHUNG : Markov Chains with Stationary Transition Probabilities, Springer-Verlag, 1960.
- (10) S.EILENBERG : Automata, Languages and Machines, Vol. A, Academic Press, 1974.
- (11) W.K.ESTES : Learning Theory and Mental Development, Academic Press, New York, 1970.

- (12) G.GORDON : System Simulation, Prentice Hall series in automatic computation, 1969.
- (13) D.GRIES : Compiler Construction for Digital Computers, John Wiley and Sons Inc., New York, 1971.
- (14) J.HOPCROFT and D.ULLMAN : Formal Languages and their Relations to Automata, Addison-Wesley, 1969.
- (15) M.IOSIFESCU and R.THEODORESCU : Random Processes and Learning, Springer-Verlag, 1969.
- (16) J.KLASA : Estimation Problem in Probabilistic Automata, Symposium on Statistics and Related Topics, 1974.
- (17) S.KLASA : On Involutions and Periodicity in Algebra, Communications in Algebra, 4(1), 1-9, 1976.
- (18) H.MINE and S.OSAKI : Markovian Decision Processes, Elsevier, 1970.
- (19) M.F.NORMAN : Markov Processes and Learning Models, Academic Press, 1972.
- (20) A.PAZ : Introduction to Probabilistic Automata, Academic Press, New York, 1971.
- (21) A.PAZ : Some aspects of Probabilistic Automata, Information and Control 9, 26-60, 1966.
- (22) M.O.RABIN and D.SCOTT : Finite Automata and their Decision Problems, I.B.M. Journal for Research Development 3, No 2, 114-125, 1963.
- (23) M.O.RABIN : Probabilistic Automata, Information and Control 6, No 3, 230-245, 1963.

- (24) A.SALOMAA : On Events represented by Probabilistic Automata of Different Types, Canad. Journal of Mathematics, 1966.
- (25) A.SALOMAA : Theory of Automata, Pergamon Press, 1969.
- (26) P.SUPPES, H.ROUANET, M.LEVINE and R.W.FRANKMAN : Studies in Mathematical Psychology, Stanford Univ. Press, Stanford, California, 358-379, 1964.

LIST OF SYMBOLS .

These symbols are used in the same or similar ways in this thesis. The definitions given on the pages indicated do not cover all local variations in meaning.

A , 3	η , 6
$ A $, 1	η_n , 49
$\langle A , p \rangle$, 5	$\eta(w)$, 47
a_{ij} , 12	f , 2
AIST , 85	F , 3, 25
A^t , 13	$f \circ g$, 10
A_x , 15	$f(q)$, 4
α_1 , 1	$F(R^*)$, 26
B_x , 6	$F^{-1}(r)$, 26
$c_i(w)$, 24	$\langle F(R^*), p_M \rangle$, 33
DIST , 92	Φ , 70, 103
D_T , 71	$G \times F$, 13
Δ , 103	G_M , 20
E , 20	γ , 76
E_1 , 20	(Y, σ) , 76
E_2 , 20	Γ , 66
ϵ , 23	Γ^* , 66
$\mathcal{E}(g, s)$, 81	H , 21

IA , 36
 IM , 34
 IPA , 27
 ISM , 24
 $ISSM$, 41
 IST , 67
 I_T , 71
 $ITSM$, 101
 i , 7, 25
 $i(w)$, 41
 $\mathcal{K}$, 57
 $\mathcal{K}^M$, 57
 $\mathcal{K}(n)$, 57
 $\ell(i)$, 93
 L , 61
 L^M , 61
 $L(n)$, 61
 L_σ° , 13
 L_σ^{-1} , 13
 λ_i , 53
 Λ_T , 70
 M , 3

$M \sim \bar{M}$, 33
 $M \stackrel{b}{\sim} \bar{M}$, 33
 $M \stackrel{c}{\sim} \bar{M}$, 33
 $|M|$, 4, 27
 $M > \bar{M}$, 53
 $M(i)$, 51
 M_p , 4, 27
 $\langle M_p, p_M \rangle$, 27
 M_t , 65
 O_T , 71
 Ω , 70, 103
 p_D , 71
 $p(q, \alpha)$, 1, 4
 p_i , 20
 $p_i(y|x)$, 6
 $p_i(w)$, 47
 p_{ij} , 18
 $p_{ij}(n)$, 17
 $p_{ij}(\alpha)$, 20
 $p_M(r)$, 26
 $\tilde{p}_M(w)$, 24
 p_R , 71

PA , 4

$P(g,s)$, 68

$P(x,y)$, 6

$P(w)$, 5, 24

$P(\alpha)$; 5, 20

$P(Y,\sigma)$; 68

p^g , 106

p^σ , 105

$Pr(a)$, 14

$Pr(A)$, 14

$Pr(a|A)$, 14

$Pr(\alpha \in S|q)$, 1

$\mathbb{I}$, 6

$\mathbb{I}^i$, 25

ψ , 61

ψ_i , 61

Ψ , 75

q' , 90

q'' , 90

q_a , 89

q_b , 89

q_c , 89

q_d , 89

q_1 , 1

Q , 1

(Q, q_1, P) , 25

(Q, q_1, T) , 34

$(Q, \mathbb{I}, P)$, 23

Q_n , 88

R , 1, 19

R^* , 2

$\langle R^*, \tilde{p}_M \rangle$, 33

$R(r)$, 42

$R(n)$, 47

R_I , 74

R_O , 75

R_T , 71, 75

R_O , 13

R_O^{-1} , 13

$R_O^{-1} \times L_O$, 100

S , 83

SITSM , 101

SM , 19

SSM , 6

SP , 19

$\hat{S}_i$, 84

σ , 76, 100

$\bar{\sigma}$, 101

Σ , 66

Σ^* , 66

T , 67

T_i , 89

T , 74

τ , 19

τ_n , 57

θ , 58

θ_i , 58

v , 101

$\bar{v}$, 101

w , 101

$|w|$, -12

X , 19

X , 11

ξ_j , 61

ζ , 64

ζ_i , 64

INDEX

- Absorbing barrier , 18
- Accelerated IST , 85
- Accepted word , 71
- Accessible ISM , 60
- Accessible state , 60
- Address of an entry , 12
- Alphabet , 12
- Automaton , 2
- B - equivalent , 33
- Behavior , 4, 27
- Better ISM , 52
- C - equivalent , 33
- Coaccessible ISM , 46
- Coaccessible state , 46
- Composition , 10
- Conditional probability , 14
- Cut - points , 5
- Decelerated IST , 92
- Discrete measure , 14
- Discrete stoch. process , 17
- Dist. diminishing mod. , 2
- Distribution function , 15
- Domain of T , 71
- Equivalent , 33
- Entry of a matrix , 12
- Event , 1, 3, 15
- Event operator , 2
- Finite det. automaton , 3
- Free monoid , 12
- Function , 10
- Generalized action , 23
- Graph of a relation , 11
- Homogeneous stoch. process , 17
- Independent rand. variables , 16
- Initiated automaton , 36
- Initiated machine , 34
- Initiated PA , 27
- Initiated SM , 7, 23
- Initiated sing SM , 41
- Initiated stoch transducer , 67
- Initiated stoch. trim , 64
- Input alphabet , 66
- Input code , 70, 103
- Input vocabulary , 66
- Length-preserving mapping , 86
- Length-preserving transducer , 8, 86
- Language , 27

Letter , 12
 Machine , 2
 Machine vocabulary , 19
 Markov matrix , 18
 Non homog. stoch. process , 17
 One-way machine , 8
 Output alphabet , 66
 Output code , 70, 103
 Partial function , 11
 Power set , 11
 Probabilistic automaton , 4, 27
 Probabilistic language , 5, 33
 Probability measure , 14
 Production set , 4, 27
 Random variable , 15
 Range of T , 71
 Reading head , 104
 Realized word , 71
 Regular event , 3
 Relation , 11
 Sample space , 14
 Set of actions , 19
 Set of states , 19
 Singular ITSM , 101
 State , 1
 State space , 1
 Stochastic language , 7, 28
 Stochastic sequential mach. , 6
 Stochastic transducer , 8, 85
 Stochastic two-way mach. , 9
 Stochastic vector , 18
 Stochastic X-machine , 19
 Subprobability measure , 14
 Substochastic matrix , 18
 Substochastic vector , 18
 Successful event , 34
 Successful generalized act. , 34
 Support , 15
 σ - algebra , 15
 Tagging process , 87
 Terminal state , 19
 Two-way stoch. machine , 101
 X - machine , 3
 Word , 12

END

1 6 0 8 7 9

FIN