T H E   U N I V E R S I T Y   O F   M I C H I G A N
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Computer and Communication Sciences Department

HOMOMORPHIC SIMPLIFICATION OF SYSTEMS

Norman Y. Foo

July 1974

Technical Report No. 156

with assistance from

National Science Foundation
Grant No. GJ-29989
Washington, D.C.

ABSTRACT

HOMOMORPHIC SIMPLIFICATION OF SYSTEMS

by

Norman Yeow-khean Foo

Chairman: Bernard P. Zeigler

A simplification procedure for dynamical systems is one which reduces complexity and yet retains aspects of structure or behavior. An approach to establishing criteria for the goodness of simplification procedures is proposed using concepts from logic. Homomorphic simplification procedures are considered from this perspective and various instances of homomorphism schemes are tested with respect to system theoretic predicates including realizability, linearity, continuity, time-invariance and stability. In constructing the framework for this investigation a topological realization theory for systems and a measure for degrees of time-variance are developed.

ACKNOWLEDGMENTS

Much of what I know about systems and computer science was learned directly or indirectly from Bernard Zeigler, Arthur Burks, John Holland, John Meyer, Arch Naylor, William Riddle and William Porter. In particular my debt to Zeigler, who provided unfailing guidance and inspiration throughout my graduate study and subsequent research, is obvious in the following pages. I am grateful to the Logic of Computers Group for partial financial support. Its members have always been willing listeners; Andrew Barto and Sudhir Aggarwal were frequent and effective critics of my conjectures. Richard Laing was a source of extracurricular wisdom. Monna Whipp transcribed my handwritten notes into beautiful typescript. Finally I wish to express my gratitude to my wife Yoke-lin for her patience and encouragement.

TABLE OF CONTENTS

TABLE OF CONTENTS

(Cont'd)

Chapter 0

The art of modeling and simulation has been slow in acquiring scientific credentials. The reason for this is the inherent complexity of the subject and hence the difficulty in providing a mathematical description of the activity itself. In simulation the scientific aspects are currently embedded in predominantly statistical issues concerning experimental design, random number generation, monte-carlo techniques, and nondeductive inference. It can be argued that this is so because the sources of difficulties in simulation do not reside in the algorithmic specification of models but rather in the construction of models and the interpretation of simular experiments. An attempt should be made to provide a conceptual basis for the process of modeling and simulation in order to identify areas where effort can be directed to substitute more science for art, in particular to mathematise those areas which are amenable to such treatment.

One such attempt has been outlined by Zeigler in a series of papers ([Z1],[Z2],[Z3],[Z4]) and a forthcoming book [Z5]. We shall rely on this conceptual basis for our development without necessarily claiming that it is the best, although it is evidently both plausible and convenient. A brief summary of the aspects of Zeigler's ideas relevant to our investigations seems appropriate.

Our knowledge of the external world stems from experimentation and consists of input-output (I-O) observations. Because of both spatial and temporal limitations these observations cannot be infinite in extent. As a consequence we are unable to comprehend the entire external world but are restricted to some subset of it via a subset of all possible I-O observations. It is this restriction that compartmentalizes our hypotheses about how the external world functions and gives rise to the notion of a system. The process of rationalizing about I-0 observations consists in positing a structure for the system, a structure which purports to account for the I-0 observations which have been obtained and may be obtained in the future. Of necessity structure assignment and I-0 experimentation is interactive in practice, as is well argued by Mihram ([M1], [M2]). The fundamental postulate of this approach is that in fact the external world is structured, and that its structure can be at least approximately known. However, given a system and its posited structure, the complete verification of that structure cannot usually be accomplished because of the finiteness of I-0 experimentation. Hence we have the notion of an experimental frame which selects only certain aspects of the system to be verified. An experimental frame imposes a conceptual simplification of the system, giving rise to a simplified system. This pair is conveniently called the base system and the lumped system respectively. Verification then proceeds via the lumped system with respect to the experimental frame. Thus, the modeling situation involves, from this perspective, an I-0 system. Each of these systems may be given a formal description in some mathematical system theory. The choice of the theory is part of the translation process from the informal to the formal description, and hence cannot be algorithmically specified. The analogy here is with Church's thesis which asserts that intuitively computable functions are indeed computable in a formal sense. We do likewise, and assert that every informal system has some formal counterpart. Unlike the theory of computable functions however, it is not yet known if all of the significant mathematical system theories are equivalent. But this need not unduly concern us because the subsequent development will not postulate such equivalence; rather, we shall feel free to move from one theory to another f convenience.

The passage from an I-0 system to a base system is called realization or identification. In a formal sense the syntheses of acceptors for various classes of languages, the design of filters to process signals according to some prescription, the construction of a molecular biological theory to account for Mendelian genetics, are all properly called realization procedures. In this broad usage of the term realization the notion of a state space is not explicitly incorporated. Indeed it is possible, as in the case of perfect band-stop filters or nondeterministic automata, that classical notions of states are dispensable. However the subsequent development will be simplified to the extent that by realization we shall mean a state realization.

The relation between a base system and a lumped system will be called a simplification. Since base systems and lumped systems yield I-0 systems in a natural way, the term simplification will also be used to refer to the corresponding relation between the two I-0 systems. In fact this use of the term is the most primitive possible. More precisely however, a simplification relation between two mathematical systems presumes the existence of some complexity measure with respect to which the relation is monotone.

Our investigation will be concerned with both the realization and the simplification issues. In particular we shall address ourselves to topological issues in addition to those of an algebraic nature.

0.1 Realizations

The provision of a good state description for an I-0 system may be viewed as a measure of economy. For causal I-0 systems a valid state space always exists, namely the set of all past inputs to the system. This hardly qualifies as a good state description because it has maximal redundancy. On the other hand the Nerode right congruence relation on this set yields a minimal state space with no redundancy, and so provides an excellent state description. Intermediate between these two descriptions lie a host of compromises, each characterized by a right congruence relation refining Nerode congruence. We shall develop a theory of topological realization by introducing topologies into I-0 systems. Using this theory the properties of a class of simplification procedures will be examined.

There remain I-0 systems which may not be causal. For these a classical state description is not possible and the question of realization does not arise. However it is still possible to simplify such systems.

0.2 Simplifications

Because the most general systems are I-0 systems we will have occasion to look at these in some detail with regard to simplifications. To keep matters simple it will be expedient to investigate only one class of simplification procedures, namely, homomorphisms. Homomorphisms will be the main topic in the later chapters. In addition, the types of systems under consideration will be restricted to non-stochastic systems. These choices reflect our current state of confidence in simplification procedures because homomorphisms and non-stochastic systems are familiar concepts, whereas more exotic procedures and systems are less so.

0.3 Validity of Simplifications.

A theme which will recur throughout the investigation is that of the validity of simplification procedures. Unlike the validity of realizations, which is merely a requirement that an internal structure accounts precisely for an I-0 behavior, it is not immediately clear what is meant by the validity of a simplification procedure. Since verification of a lumped system is with respect to some experimental frame it is evident that validity must refer to a hypothetical set of experiments which we are prepared to conduct. In practice this is not by itself very useful because it is usually the case that complete verification necessitates an infinite number of such experiments. Hence we have recourse to statistics, to ergodicity assumptions, and the like. However, we feel that there is a more profound basis for the definition of validity which we shall now expound.

Given a particular system theory τ and the collection $\mathcal{S}$ of systems describable in that theory, we may draw up lists of n-ary <u>predicates</u> over $\mathcal{S}$ corresponding to familiar relations on and properties of systems. Each predicate selects a subset of some cartesian product of $\mathcal{S}$, namely, the extension of that predicate. Some typical predicates are discussed below. The predicates which are adduced by a particular experimemtal frame consists in the formalization of the kinds of system theoretic questions which may be asked relative to that frame. The validity of a simplification procedure is judged by the preservation of properties and relations (defined by some frame) under that procedure. More precisely, let $P:\mathcal{S} \to \mathcal{S}$ be some map on $\mathcal{S}$, ϕ_p its associated binary predicate, i.e., $\phi_p(S,S') \leftrightarrow P(S) = S'$, ϕ an n-ary predicate on $\mathcal{S}$; then P is <u>directly</u> <u>ϕ-valid</u> if

$$(S_1)\dots(S_n)((\phi_p(S_1,S_1')\wedge\dots\wedge\phi_p(S_n,S_n')) \supset$$

$$(\phi(S_1,\dots,S_n) \supset \phi(S_1',\dots,S_n'))$$

and P is <u>conversely</u> <u>ϕ-valid</u> if

$$(S_1)\dots(S_n)((\phi_p(S_1,S_1')\wedge\dots\wedge\phi_p(S_n,S_n')) \supset$$

$$(\phi(S_1',\dots,S_n') \supset \phi(S_1,\dots,S_n))$$

and P is strongly ϕ-valid if it is both directly and conversely ϕ-valid. For brevity we shall say ϕ-validity for direct ϕ-validity since this is the most common instance we shall consider.

The generalization of the above concepts is immediate. Suppose Ω is a collection of predicates which are meaningful in two system theories τ and τ', with associated system classes S and S'. Then a procedure $P:S\rightarrow S'$ is Ω-valid if it is valid with respect to every ϕ in Ω, and the unprimed and primed symbols are interpreted to be in S and S' respectively.

These definitions are designed to capture formally the essential meaning of valid simplification procedures. For instance, if we consider the predicates LINEAR, CONNECTED and CONTINUOUS, all of which are unary, a homomorphism $h:S\rightarrow S$ is LINEAR-, CONNECTED-, and CONTINUOUS- valid if

$$(S)((h(S) = S') \supset ((\mathrm{LINEAR}(S) \wedge \mathrm{CONNECTED}(S) \wedge \mathrm{CONTINUOUS}(S)) \supset (\mathrm{LINEAR}(S') \wedge \mathrm{CONNECTED}(S') \wedge \mathrm{CONTINUOUS}(S')))$$

with the corresponding modifications for converse and strong validity. It is presumed that these three predicates characterize some experimental frame, i.e., completely and exactly specify the questions of interest. In practice no such characterization is feasible because of the adaptive, interactive and iterative nature of the modeling and simulation process. Nevertheless what we have here is an idealization of an instant of this dynamic process, and it is hoped that an understanding of this will enable us to tackle the more ambitious enterprise of modeling dyanmics with improved confidence.

Several observations are pertinent. Many predicates of interest are not computable in the sense of Turing. A simple example of this is the predicate LINEAR on systems defined with vector spaces over an infinite field. This observation should give pause to those who are concerned with effectiveness but it turns out that most questions of any depth in systems with topology have this regretable character. (However, it may be possible to circumvent this by appealing to an abstract theory of analog computability [P4] or to some kind of a measure-theoretic approximation.) Next, if S is infinite there are more predicates than we can possibly name. This is because predicates are coextensive with subsets of cartesian products of S. Given S the identity map $I:S \to S$ is one which satisfies all conditions of predicate validity. Conversely, if a procedure $P:S \to S$ is required to be valid with respect to every predicate it has to be the identity map because it must for any $S \in S$ satisfy the assertion $\phi(S) \supset \phi(P(S))$ where ϕ is the predicate denoting membership in the singleton set $\{S\}$. Thus every meaningful simplification procedure must be required to be valid with respect to a proper subset of all possible predicates. More is required. If P is to be a true simplification it has to be monotonic in some complexity measure $\mu:S \to R$, i.e., $\phi_p(S_1,S_2) \supset (\mu(S_1) > \mu(S_2))$, the selection of μ being part of the modeling process. If P_1,P_2 are two simplification procedures, and P_1 is "stronger" than P_2, P_2 is Φ-valid implies P_1 is Φ-valid. For instance, if P_2 is a homomorphism and P_1 is a continuous homomorphism then P_1 is stronger than P_2.

It should be clear that in any useful simplification some system predicates will not be preserved. An essential feature of modeling is the deliberate exclusion of properties to which the modeler is indifferent. The choice of which properties to exclude is part of the modeling process.

An examination of some of the predicates quoted as examples will reveal that the language we are proposing to describe validity is very "high level". A predicate such as LINEAR presumes many concepts at a more primitive or lower level, including definitions of a vector space and a linear transformation. A CONTINUOUS predicate presumes the definitions of a continuous map, hence of a topological space. This means that we are using the predicates as abbreviations of sentences which are really much more complex than indicated. (The advantage gained in treating high level predicates as our primitives is analogous to the advantages afforded by high level programming languages or macros.) We need not apologize for this because although it is widely accepted that all mathematics can be reduced to axiomatic set theory, most mathematics is discussed in a language using concepts at much higher levels. Occasionally we may have to reach inside a system S to say something about individual points within the list that describes S. Suppose $S = (K,X)$ where $K:X \to X$ is the system operator, X is the input and output space; then x_0 is a fixed point of S if $Kx_0 = x_0$. To formalize assertions like this we consider formulas which are free in certain symbols. In this case let $\psi(S,x_0)$ be free in the symbols S and x_0, to be interpreted as "x_0 is a fixed point of S". Then if S' is a simplification of S via $P:\mathcal{S} \to \mathcal{S}$, P is <u>ψ-valid</u> if

$$(S)(\phi_p(S,S') \supset (\psi(S,x) \supset \psi(S',x')))$$

where x' is the point assigned by P to x. This suggests that the earlier definition of validity is a special case of a more complex definition involving formulas. Because we are usually concerned with global properties which are characterized by predicates, this more general definition will not be emphasized.

0.4 Higher order validity.

Thus far the definition of validity has been confined to first order properties of systems. In this case every simplification procedure is a homomorphism of an Ω-structure (of universal algebra). Let Ω be a collection of predicate letters and $\mathcal{S}$, $\mathcal{S}'$ be the carriers of the Ω-structure.

Then the condition for $P:S \to S'$ to be an Ω-structure homomorphism is precisely what we have defined to be direct Ω-validity.

In the study of systems, especially with regard to questions which prompt people to model and simulate, topological and quasitopological properties are as important as algebraic properties. Many of the former are expressible in a first order language*, but some are not. In the case that a second order language is required the definition of validity has to be generalized to allow quantification over predicates (and functions in particular). To make matters more precise, suppose $\alpha(f,\phi,S)$ is a formula free in function symbol f, unary predicate letter ϕ and variable S. The symbol f is to be interpreted into a map $f:Z \to S$, where Z is the integers. Then a simplification $P:S \to S$ is <u>α-valid</u> if

$$(\phi)(f)(S)(\alpha(f,\phi,S) \supset \alpha(P\cdot f,\phi\cdot P,P(S))).$$

This is a second order sentence in a 2-sorted language. More generally we would allow a number of function symbols, predicate letters (of arbitrary arity), and variables, and let $P:S \to S'$. But the generalization is obvious. The purpose of introducing a function on Z is to allow <u>sequences</u> to be treated. This is adequate as long as S has the minimal structure of a separable topological space, which will always be assumed. We consider the use of nets to be of marginal advantage.

There is another motivation for studying topological properties. A predicate is <u>fuzzy</u> if its extension has a characteristic function with values in an interval, according to Zadeh [Za1]. A fuzzy predicate such as ALMOST-LINEAR may be given a rigorous interpretation by identifying systems satisfying the LINEAR predicate, taking the quotient, and finding a topology for the quotient space which suitably "measures" nonlinearity. Such a procedure is carried out for a different fuzzy predicate in the closing chapter.

0.5 Simplification Procedures

We recall several procedures in common use for the simplification of systems. Given a coordinate system on the input-output or state set for a system, each coordinate defines a variable. Omission of some variables is one class of simplification procedures. Another class arises from the aggregation of variables. A technique for simplification which not only clarifies structure but is also a very powerful design tool is that of hierarchical decomposition. Other modes of decomposition are also used. In fact hierarchical and lateral decomposition are often employed together. Simon [S1] and Mesarovic, et al. [M3] contain good discussions of these techniques. Relatively unexplored are procedures which convert stochastic to deterministic systems and vice versa although the introduction of the ubiquitous (external or internal) random "noise" in modeling is one aspect of such procedures. It is expected that results in this area will be forthcoming. In our investigation we shall be dealing exclusively with a class of procedures collectively termed "homomorphisms". The notion of a system homomorphism arose from algebraic automata theory. Basically the idea is simple. For one automaton to be a homomorphic image of another certain commutative diagrams are presented. Depending on the "strength" of the homomorphism the objects of the commutative diagrams may consist of the input, output or state sets, which may in turn be coordinatized or not. Auxiliary conditions, such as the preservation of structural dependencies in state spaces, may be imposed.

The motivation for examining homomorphisms in detail stems from the conviction that they are simple yet realistic schemes for simplifi-tion. A complex biological model has been simplified by Zeigler and Weinberg using one instance of such a scheme [Z4] and successfully simulated on a digital computer. It is not known, however, what conditions on homomorphisms are really necessary and/or sufficient for different sets of system predicates to be preserved. It also turns out that many simplification procedures are instances of homomorphisms. For example, the aggregation procedures of economics, investigated in a control-theoretic framework by Aoki [A1], is a homomorphism of state automata.

From another perspective homomorphisms are idealized versions of simplifications. The nearly-decomposable systems of Ando, et al. [A2], and the method of singular perturbations [W1] in applied mathematics are approximate homomorphisms in the sense of diagrams that do not quite commute. A recent paper by Park [P1] in effect applies singular perturbation techniques to achieve an approximate homomorphic simplification of metabolic networks. To anticipate things a little, suppose $h:Q \to Q_1$ is a map from a state space to another, and $\delta:Q \to Q$ and $\delta_1:Q_1 \to Q_1$ are autonomous state transition functions. For a homomorphism we require that $h\delta = \delta_1 h$. Suppose Q_1 is a metric space with metric d. Then an approximate homomorphism will be one such that for all $q \varepsilon Q$ $d(h\delta(q),\delta_1 h(q)) < \varepsilon$ for some ε. The time interval over which δ and δ_1 are defined will be the domain of validity of the approximation. In any case strict homomorphisms arise from idealizing these approximations, with a view to establishing the scope and limits of such procedures.

Every set of commutative diagrams which define a class of homomorphisms in fact define a homomorphism scheme. Instances of the scheme are obtained by providing conditions on the maps and objects. Throughout the investigation we shall be considering such instances with regard to particular schemes. An instance i of some scheme

may contain all the conditions of another instance j of the same scheme. If we have a chain of such instances they may be ordered in a hierarchy. An example of such a hierarchy is the progression towards stronger conditions in the instances: input-output homomorphism, homomorphism, weak structure homomorphism, strong structure homomorphism [Z1]. Suppose ϕ_1 and ϕ_2 are the associated predicates of two instances of a homomorphism with ϕ_2 stronger than ϕ_1. A justifying condition (see Zeigler [Z1]) is a predicate on either the base or lumped system (or both) which will allow us to deduce ϕ_2 knowing ϕ_1, that is, climb up the hierarchy. More precisely, a justifying condition on the base system is a unary predicate J such that

$$J(S) \wedge \phi_1(S,S') \supset \phi_2(S,S')$$

0.6 Systems

The concept of a system was introduced earlier. Its formalization has been carried out in the literature, and we shall merely outline several possibilities here, leaving the details to later chapters. I-0 systems are considered in great detail in functional analytic approaches to systems theory; for examples see Porter [P2], Naylor [N1], Saeks [S2], Sandberg [S3], and Wilems [W2]. Systems with state arise from both control and automata theory; for examples see Arbib ([A3],[A4]), Kalman, Falb and Arbib [K1], and the enormous literature on either subject. Much of the theory has been developed for discrete time systems, and with purely algebraic structures. In the case of I-0 systems the setting is usually continuous and the structures are analytical. In our development we shall feel free to pass from one formulation to another depending on the convenience of the situation. In particular we shall sometimes be considering systems with state and sometimes I-0 systems.

Chapter 1

In this and the next chapter a topological realization theory for systems is developed. This will serve as a prerequisite for linking input-output descriptions to state descriptions of systems, and also as part of a continuing effort to reconcile algebraic automata theory with control theory. Some systems concepts relevant to the development will be presented in the next section but for more complete discussions the following references should be consulted: Kalman, Falb and Arbib [K1], Arbib ([A3], [A4]), Naylor and Sell [N1], and Porter ([P2], [P3]) In these two chapters we shall focus our attention on time-invariant, causal systems. These restrictions will be relaxed in later chapters.

1.1 Some systems concepts

In the later chapters we will encounter systems which are defined by an <u>input space</u> and an <u>output space</u> of time functions. The time functions are defined on $(-\infty,\infty)$ or on Z the set of integers, and have as range a <u>value space</u>. Time-invariance is easy to define using a time set $(-\infty,\infty)$ or Z if we first define a <u>shift operator</u> S_τ or $S(\tau)$, acting on time functions so that

$$(S_\tau x)(t) = x(t-\tau)$$

for any x a function of time. We will use the notation S_τ or $S(\tau)$ according to the desired emphasis, the latter being used whenever τ is to be regarded as an argument of interest. In the first two chapters the shift τ is only incidental, hence the operator will be denoted by S_τ, but in the closing chapter it becomes paramount, and there the operator is denoted by $S(\tau)$. Suppose $\hat{f}:\hat{\mathcal{U}} \to \hat{\mathcal{Y}}$ is a function from an input space to an output space of time functions, and represents an <u>input-output (IO) system</u> where the time set is $(-\infty,\infty)$. Then <u>time-in-</u>

variance of $\hat{f}$ is defined by

$$S_\tau \hat{f} = \hat{f} S_\tau \qquad \text{for all } \tau .$$

Suppose now the time set is to be altered to some semi-infinite interval, say $[0,\infty)$. A reasonable yet rigorous definition of time invariance demands that we take the following route. Let $\hat{U}$ be the space of partial functions on $(-\infty,\infty)$ such that each function $x \varepsilon \hat{U}$ is defined from some t onwards, i.e., on $[t,\infty)$. $\hat{Y}$ may be likewise defined. A subspace of $\hat{U}$ consisting of all functions defined on $[t,\infty)$ is denoted by $\hat{U}_t$. A shift operator S_τ will now be regarded as acting on $\hat{U}_t$ to produce another subpsace $\hat{U}_{t+\tau}$, i..e., it is presumed that $\hat{U}$ is closed under $\{S_\tau\}$, the group of shift operators. The shift operator is slightly modified in this instance, viz.,

$$(S_\tau x)(\alpha) = \begin{cases} x(\alpha-\tau) & \text{when } \alpha-\tau \geq t \\ \text{undefined otherwise} & \end{cases}$$

Using this modification the above definition for time invariance still holds.

$\hat{U}_t$ and $\hat{U}_{t+\tau}$ are topological subspaces of $\hat{U}$. S_τ is an isomorphism between the subspaces, and we shall assume that $\hat{U}_t$ and $\hat{U}_{t+\tau}$ are in fact homeomorphic. Thus S_τ is a homeomorphic isomorphism, i.e., an iseomorphism. Therefore we may record:

Definition 1.1.1

A system $\hat{f}$ is time-invariant if $\hat{f} S_\tau = S_\tau \hat{f}$ for all τ.

Evidently, when $\hat{f}$ is time-invariant it suffices to restrict consideration to $\hat{U}_o$. We shall do this and write $\hat{U}$ for $\hat{U}_o$ in this context.

A forward restriction function E^τ on $\hat{U}$ is defined by $(E^\tau \hat{u})(t) = \hat{u}(t)$ for $t \varepsilon [0,\tau)$, and undefined elsewhere. The family of such functions $\{E^\tau\}_{\tau \varepsilon R+}$ where R_+ is the non-negative reals, act on $\hat{U}$ to produce the set of initial segments denoted by $\mathcal{U}$. Thus each element u of $\mathcal{U}$ is a restriction of some $\hat{u}$ of $\hat{\mathcal{U}}$. $\mathcal{U}$ is called the segment space (of $\hat{\mathcal{U}}$). A backward restriction function E_τ on $\hat{\mathcal{U}}$ is defined likewise, restricting domains to $[\tau,\infty)$.

The binary operation of concatenation, denoted by $\circ$, is defined as follows. Suppose $u_1,u_2 \varepsilon \mathcal{U}$ and $\mathcal{D}(u_1) = [0,t_1)$, $\mathcal{D}(u_2) = [0,t_2)$. Then

$$(u_1 \circ u_2)(t) = \begin{cases} u_1(t) & t \in [0,t_1) \\ u_2(t-t_1) & t \in [t_1, t_1+t_2) \\ \text{undefined} & \text{otherwise} \end{cases}$$

We shall assume that under concatenation as algebraic composition $\mathcal{U}$ is closed. The empty segment Λ is an identity, and concatenation is associative. Hence $\mathcal{U}$ is a monoid.

Definition 1.1.2

A system f is causal if $\forall \hat{u},\hat{v} \varepsilon \hat{\mathcal{U}}\ \forall \tau \varepsilon R_+ \quad E^\tau \hat{u} = E^\tau \hat{v} \Rightarrow E^\tau \hat{f}\hat{u} = E^\tau \hat{f}\hat{v}$.

$\mathcal{U}$ corresponds to initial substrings in automata theory.

In many applications the system $\hat{f}$ exhibits inertia in the sense that it cannot distinguish between two input elements u_1 and u_2 which differ only on a set of (Lebesque) measure zero on R_+. When such is the case $\hat{\mathcal{U}}$ may be replaced by its quotient space where input functions differing on sets of measure zero are identified. In this instance it is clear that the construction of $\mathcal{U}$ may be accomplished by considering segments defined on closed (or open) intervals, since the function values at the end points of segments may be arbitrarily altered. This will be assumed whenever (Lebesque) measure is placed on R_+.

In other applications it may be desirable to place more structure on $\hat{\mathcal{U}}$, for instance $\hat{\mathcal{U}}$ may consist of piece-wise continuous functions. In this particular case it is clear that the monoid structure of $\mathcal{U}$ remains intact. Our concern later is to establish some cases when the algebraic monoid structure is embellished to a topological monoid structure. Particular instances of topologies on $\hat{U}$ which will be considered in the sequel will be defined as the situation arises.

The system $\hat{f}$ is <u>linear</u> when $\hat{U}$ and $\hat{Y}$ have the structure of linear spaces, and $\hat{f}$ is a linear transformation. More generally $\hat{f}$ is a <u>homomorphism</u> when $\hat{U}$ and $\hat{Y}$ have algebraic structures and $\hat{f}$ is an algebraic homomorphism (preserving the structure). The system $\hat{f}$ is <u>continuous</u> if $\hat{f}$ is continuous.

1.2 Topologies for U

The segment space U may be topologized in several ways giving rise to different modes of convergence. It is pertinent to ask what is meant by the assertion that a collection of segments converges to a given segment. To answer this a different way of looking at segments is explicated. (Dugundji [D4] is a good reference for topology.)

Consider the cartesian product $\hat{U} \times R_+$. The relation ρ on $\hat{U} \times R_+$ defined by

$$(\hat{u}_1,t_1)\rho\ (\hat{u}_2,t_2) \iff t_1 = t_2 \text{ and } \hat{u}_1|_{[0,t_1)} = \hat{u}_2|_{[0,t_2)}$$

is an <u>equivalence relation</u>. There is a bijection between the quotient space $\hat{U} \times R_+/\rho$ and the segment space U given by $[(\hat{u},t)]_\rho \to E^t\hat{u}$. That this is indeed a bijection is easily verified. In fact this bijection is a monoid isomorphism between $\hat{U} \times R_+/\rho$ and U, with the concatenation operation on the first space defined in the obvious way.

Now, $\hat{U}$ has a function space topology while R_+ has several possible topologies. (For discrete-time systems, R_+ may be replaced by Z, which is given the discrete topology). It is natural to equip

$\hat{U} \times R_+$ with the product topology and hence $\hat{U} \times R_+/\rho$ with the quotient topology from $\hat{U} \times R_+$. In this way the segment space is topologized in a natural manner. Note that the inverse image of a segment $[(\hat{u},t)]\rho$ in $\hat{U} \times R_+$ is a set $\{(\hat{v},t): \hat{v}|_{[0,t)} = \hat{u}|_{[0,t)}\}$

Let U and Y be the value spaces associated with the input and output spaces $\hat{U}$ and $\hat{Y}$ respectively (and hence of U and Y). We now specialize the development to the case where at least U is a metric space with a metric d. It may in fact be shown that the theory can be generalized to the case where U is a uniform space or simply a topological space, but this would complicate many proofs and obscure the essentially simple ideas on which the theory is based.

The metric d in U induces a metric d' in $\hat{U}$ as follows:

$$d'(\hat{u}_1,\hat{u}_2) = \sup_{t \varepsilon R_+} d(\hat{u}_1(t),\hat{u}_2(t))$$

In fact, d' is a <u>generalized</u> <u>metric</u> since it is a map into the extended non-negative reals. We shall write d for any metric when no confusion can arise.

It is noted that a metric d on $\hat{u}$ may indeed arise without reference to the metric (if any) on U. For our purposes either point of view is acceptable.

<u>Theorem 1.2.1</u>

The natural projection $p:\hat{U} \times R_+ \to \hat{U} \times R_+/\rho$ is both continuous and open.

<u>Proof</u>:

Continuity follows by definition of the quotient topology.

Let $\theta \times N$ be a basic open set in $\hat{U} \times R_+$, i.e. $\theta = \{\hat{u}|d(\hat{u},\hat{u}_o)<\varepsilon, \hat{u}_o \varepsilon \hat{U}\}$ and $N = (t_1,t_2)$. Then $p(\theta \times N) = \{[\hat{u},t)]: t \varepsilon (t_1,t_2)$ and restricted to $[0,t)$ $d(\hat{u},\hat{u}_o)<\varepsilon\}$, so that $p^{-1}(p(\theta \times N) = \{(\hat{u},t) | t>t_1$ and $d(\hat{u},\hat{u}_o)<\varepsilon\}$. Let $\hat{v}$ be an element in it. Then $d(\hat{v},u_o)<\varepsilon$, say $d(\hat{v},u_o) = \varepsilon-k$ where $k > 0$. The set $\theta' \times N'$ where $\theta' = \{\hat{u} | d(\hat{u},\hat{v}) < k/2\}$ and $N' = (t_1,\infty)$ contains $\hat{v}$, and is contained in $p^{-1}(p(\theta \times N))$, so that the latter is open. Thus $p(\theta \times N)$ is open, and p is open.

In the above proof it was tacitly assumed that the topology on R_+ is the usual one. Suppose instead the topology on R_+ was the discrete topology. Then every point is open. An examination of the proof shows that the theorem is also valid in this case.

Now we are in a position to answer the question posed at the beginnir of this section.

First, it is observed that both the usual and the discrete topologies on R_+ are metric topologies, given respectively by $d(t_1,t_2) = |t_1-t_2|$ and $d(t_1,t_2) = 1$ for all $t_1 \neq t_2$. Since $\hat{U}$ is a metric space it follows that $\hat{U} \times R_+$ is a metric space. However, in general $\hat{U} \times R_+/\rho$ is not metrizable, although it may be made so in a somewhat contrived fashion to be described later. Nevertheless it is not necessary to use the language of filters or nets when discussing convergence because it turns out that $\hat{U} \times R_+/\rho = U$ is first-countable, hence sequences are sufficient. To prove first-countability, note that p is open and continuous implies that open sets in U are precisely the images of open sets in $\hat{U} \times R_+$. $\hat{U} \times R_+$ is first-countable because it is metric space. Hence U is first-countable, and sequences are adequate.

Let $\{u_n\}$ be a sequence converging to u in U, and suppose R_+ has the usual topology. Let $(\hat{u},t) \in p^{-1}(u)$ and $\theta \times N$ be a basic neighborhood of $(\hat{u},t)$, i.e., θ is an ε-ball about u and $t \in N$, an open interval. Then $p(\theta \times N)$ is a neighborhood of u by the theorem just proved, so that $\{u_n\}$ must eventually lie in $p(\theta \times N)$. The preimage of u_n under

p is a set $\{(\hat{u}_\alpha, t_n) : \hat{u}_\alpha|_{[0,t_n)} = \hat{u}|_{[0,t_n)}\}$ so each element of the preimage is associated with the same t_n. Thus, to say that $\{u_n\}$ is eventually in $p(\theta \times N)$ is to say that (i) $t_n \to t$ and (ii) $E^t\hat{u}_n \to E^t\hat{u}$ where $\hat{u}_n$ and $\hat{u}$ are in the preimages of u_n and u respectively.

If instead the discrete topology for R_+ was used, then condition (i) is modified to $t_n = t$ eventually. In this case $\mathcal{U}$ is metrizable in a trivial way, namely, a generalized metric d is definable by

$$d(u,v) = \begin{cases} \infty & \text{if } \mathcal{D}(u) \neq \mathcal{D}(v) \\ d(\hat{u},\hat{v}) & \text{otherwise,} \end{cases}$$

In this case an alternative approach to topologizing $\mathcal{U}$ is possible, corresponding to a constructive specification of $\hat{\mathcal{U}}$ from $\mathcal{U}$: let $\mathcal{U}_t = U^{[0,t)}$, and then $\mathcal{U}$ may be regarded as the <u>free union</u> of $\{\mathcal{U}_t\}_{t \in R_+}$. This free union is analogous to input space I* of automata theory, except of course that it may not be a free semigroup under concatenation. To show that this topology for $\mathcal{U}$ coincides with that of $\hat{\mathcal{U}} \times R_+/p$ when R_+ has the discrete topology it suffices to observe that given $u \in \mathcal{U}$ we know that $u \in \mathcal{U}_t$ for some t: an open set in $\mathcal{U}$ containing u is, in particular, one which is open in $\mathcal{U}_t$ and empty in all other components.

We examine the consequence of the two topologies imposed on R_+, beginning with the discrete topology. Let $\{u_n\}$ be a sequence in $\hat{\mathcal{U}} \times R_+/\rho$ such that $u_n \to u$. Let $(\hat{u},t) \in p^{-1}(u)$. Observe that if $(\hat{v},t_v) \in p^{-1}(u)$ then $t = t_v$ and $E^t\hat{u} = E^t\hat{v}$. An open set about $(\hat{u},t)$ is $\theta_u \times \{t\}$ where θ_u is some ball of radius δ about $\hat{u}$. It is evident that

$$\{p^{-1}p(\theta_u \times \{t\})\} = \{(\hat{v},t) \mid E^t\hat{v} = E^t\hat{w} \text{ for some } \hat{w} \text{ where } d(\hat{w},\hat{u})<\delta\}$$
$$= \{(\hat{v},t) \mid d(E^t\hat{v}, E^t\hat{u})<\delta\} .$$

p is open implies $p(\theta_u \times \{t\})$ is open, so u_n is eventually in this set. Let $(\hat{u}_n,t_n) \in p^{-1}(u_n)$. Then $t_n = t$ eventually, for otherwise $u_n = p(\hat{u}_n,t_n)$ cannot be in $p(\theta_u \times \{t\})$ eventually. In addition, $E^t\hat{u}_n \to E^t\hat{u}$

uniformly. For, suppose otherwise. Then there exists $\varepsilon > 0$ and a subsequence $\{E^t\hat{u}_m\}$ such that $d(E^t\hat{u}_m, E^t\hat{u}) > \varepsilon$. Choosing $\delta=\varepsilon$ it is clear that $(\hat{u}_m, t)$ is not in $p^{-1}p(\theta_u x\{t\})$, hence $u_m = p(u_m, t)$ cannot be in $p(\theta_u x\{t\})$ eventually, which is a contradiction.

Conversely, suppose $\{(\hat{u}_n, t_n)\}$ is a sequence such that $t_n = t$ eventually and $E^t\hat{u}_n \to E^t\hat{u}$ uniformly. Let v_n be the extension of $E^t\hat{u}_n$ which agrees with $\hat{u}$ beyond t. Then $p(\hat{u}_n, t) = p(\hat{v}_n, t)$. Also $\hat{v}_n \to \hat{u}$ uniformly, and certainly $t_n \to t$. So $p(\hat{u}_n, t_n) \to p(\hat{u}, t)$.

Summarizing the above discussion we state the following principle which will be used in the sequel.

<u>Theorem 1.2.2</u>

Let $\hat{U}$ have the topology of uniform convergence and R_+ the discrete topology.

Suppose $u_n \to u$. Then for any sequence $\{(\hat{u}_n, t_n)\}$ such that $(\hat{u}_n, t_n) \in p^{-1}(u_n)$, and for any $(\hat{u}, t)$ with $(\hat{u}, t) \in p^{-1}(u)$, $t_n = t$ eventually and $E^t\hat{u}_n \to E^t\hat{u}$ uniformly.

Conversely, if $\{(\hat{u}_n, t_n)\}$ is a sequence such that $t_n = t$ eventually and $E^t\hat{u}_n \to E^t\hat{u}$ uniformly, then $p(\hat{u}_n, t_n) \to p(\hat{u}, t)$.

Next, assume that R_+ has the usual topology. The above arguemnt may be repeated with the modification that

$$\{p^{-1}p(\theta_u \times N_t)\} = \{(v, t_v) \mid t_v \in N_t \text{ and } d(E^{t_v}\hat{v}, E^{t_v}\hat{u}) < \delta\}$$

where N_t is some open interval about t. In this case it may be verified that $t_n \to t$ and $u_n \to u$ uniformly on any <u>closed</u> <u>interval</u> $[0,s] \subsetneq [0,t)$. Unfortunately the converse does not go through. However, by weakening the topology on $\hat{U}$ to that of <u>pointwise</u> convergence (thus insisting that a metric must exist in the space U) it is possible to remedy this. If this is done, we have

Theorem 1.2.3

Let $\hat{U}$ have the topology of pointwise convergence and R_+ the usual topology.

Suppose $u_n \to u$. Then for any sequence $\{(\hat{u}_n,t_n)\}$ such that $(\hat{u}_n,t_n) \varepsilon\ p^{-1}(u_n)$, and for any $(\hat{u},t)$ with $(\hat{u},t) \varepsilon\ p^{-1}(u)$, $t_n \to t$ and $E^t\hat{u}_n \to E^t\hat{u}$ pointwise.

Conversely, if $\{(\hat{u}_n,t_n)\}$ is a sequence such that $t_n \to t$ and $E^t\hat{u}_n \to E^t\hat{u}$ pointwise, then $p(\hat{u}_n,t_n)\ \varepsilon\ p(\hat{u},t)$.

Before proving the theorem some remarks are in order. With the topology of pointwise convergence on $\hat{U}$ and the usual topology on R_+ theorem 1.2.1 must be re-established. To do this requires that we look at the images of under p of basic open sets of the form $\phi = \theta \times (t-\varepsilon,t+\varepsilon)$, with $\theta = \bigcap_{i=1}^{n} [[t_i,\theta_i]]$, where $[[t_i,\theta_i]] = \{\hat{u}\ \varepsilon\ \hat{U} | \hat{u}(t_i)\ \varepsilon\ \theta_i, \theta_i$ open in U$\}$.

Thus $\phi = \{(\hat{u},t_u) | \hat{u}(t_i)\ \varepsilon\ \theta_i, 1 < i < n$ and $t_u \varepsilon (t-\varepsilon,t+\varepsilon)\}$.

We claim that

$$p^{-1}p(\phi) = \{(\hat{v},t') | t' \varepsilon (t-\varepsilon,t+\varepsilon) \text{ and } \hat{v}(t_i)\ \varepsilon\ \theta_i \text{ for } t_i < t')\}$$

$$= \bigcap_{t_i<t'} [[t_i,\theta_i]] \times (t-\varepsilon,t+\varepsilon)$$

which is open, so that p is open. To prove equality of the two sets, first observe that if $p(\hat{u}_1,t_1) = p(\hat{u}_2,t_2)$ then necessarily $t_1 = t_2$. So if $(\hat{v},t')\ \varepsilon\ p^{-1}p(\phi)$ it must be that $t'\ \varepsilon\ (t-\varepsilon,t+\varepsilon)$. Also, $\hat{v}(t_i) \varepsilon\ \theta_i$ for $t_i < t'$, for supposing the contrary we have $\hat{v}(t_j) \not\varepsilon\ \theta_j$ for some $t_j < t'$. By assumption $p(\hat{v},t')\ \varepsilon\ p(\phi)$, so $p(\hat{v},t') = p(\hat{u},t_u)$ for some $(\hat{u},t_u)\ \varepsilon\ \phi$. Then by observation above $t_u = t'$, and $E^{t'}\hat{u} = E^{t'}\hat{v}$, giving $\hat{v}(t_j)\ \varepsilon\ \theta_j$, a contradiction. Thus the lefthand set is included in the right. The reverse inclusion is easily verified, so its demonstration is omitted.

It is clear that U is first countable under these circumstances.

Proof of theorem:

Let $u_n \to u$, $\{(\hat{u}_n,t_n)\}$ be a sequence such that $(\hat{u}_n,t_n) \in p^{-1}(u_n)$, and $(\hat{u},t) \in p^{-1}(u)$. Let $\phi = \theta \times (t-\varepsilon,t+\varepsilon)$ be basic open set about $(\hat{u},\hat{t})$. By the result above p is open, so $p(\phi)$ is open in U. $u_n \to u$ now implies that u_n is eventually in this open set, i.e., $(\hat{u}_n,t_n)$ is eventually in $p^{-1}p(\phi)$. Recall that $\theta = \bigcap_{i=1}^{n} [[r_i,\theta_i]]$. Claim: $t_n \to t$. Suppose otherwise: then there is $\delta > 0$ and a subsequence $\{t_m\}$ such that $|t_m - t| > \delta$. Since

$$p^{-1}p(\phi) = \{(\hat{v},t_v)|t_v \in (t-\varepsilon,t+\varepsilon) \text{ and } \hat{v}(r_i) \in \theta_i\ r_i < t_v\}$$

choosing $\varepsilon = \delta$ gives a contradiction. So $t_n \to t$. Claim: $E^t\hat{u}_n \to E^t\hat{u}$ pointwise. Again suppose the contrary. Then there exist $t_0 \in [0,t)$, $\alpha > 0$, and a subsequence $\{\hat{u}_m\}$ such that $d(\hat{u}_m(t_0),\hat{u}(t_0)) < \alpha$ for each m. Choosing $r_i = t_0$ and $\theta_1 = \{x \in U\ d(x,u(t_0)) < \alpha\}$ yields the required contradiction because $\hat{u}_n(t_0)$ must eventually be in θ_i.

For the converse part, let the hypotheses hold. Suppose $p(\hat{u}_n,t_n) \not\to p(\hat{u},t)$. Then there is an open set θ_u about $p(\hat{u},t) = u$ and a subsequence $(\hat{u}_m,t_m)$ such that $p(\hat{u}_m,t_m) \notin \theta_u$ for all m. There is a basic open set $\hat{\theta}_u$ about $(\hat{u},t)$ such that $P(\hat{\theta}_u) \subset \theta_u$, $\hat{\theta}_u = \bigcap_{i=1}^{n} [[r_i,U_i]] \times (t-\varepsilon,t+\varepsilon)$, and $(\hat{u}_m,t_m) \in \theta_u$ for all m. Choose $r_i<t$, $1<i<n$. This implies either (i) $t_m \not\to t$ or (ii) $t_m \to t$ but $u_m(r_i) \notin U_i$ for each m, or equivalently $\hat{u}_m(r_i) \not\to \hat{u}/r_i)$. (i) contradicts $t_n \to t$ and (ii) contradicts $E^t\hat{u}_n \to E^t\hat{u}$ pointwise. Thus the conclusions of the converse are proved.

1.3 U as a topological monoid

It was earlier noted that $(U,\circ)$ is a monoid. In fact it is a cancellative monoid. When U has topological structures as given in the previous section it is possible to say more.

Definition 1.3.1

(S,∘) as a right(left) semitopological monoid if S is both a topological space and a monoid, and $\circ : S \times S \to S$ is continuous on the right (left) component. If it is continuous on both right and left it is called semitopological monoid, while if it is jointly continuous on both components it is called a topological monoid.

The following proposition is easy to prove:

Theorem 1.3.1

$(\mathcal{U},\circ)$ is a topological monoid when $\mathcal{U}$ has the topology of uniform convergence and R_+ has the discrete topology.

In the case when $\mathcal{U}$ has the pointwise convergence topology and R_+ the usual topology things are not quite as simple. However the following is true:

Theorem 1.3.2

$(\mathcal{U},\circ)$ is a right semitopological monoid when $\mathcal{U}$ has the topology of pointwise convergence and R_+ has the usual topology.

The distinction between the last two theorems reflects the difference in the power of the two modes of convergence. Without restrictions on $\hat{\mathcal{U}}$ theorem 1.3.2 cannot be strengthened. (See remarks on p. 25.)

In the sequel we shall call the topology in theorem 1.3.1 the u-topology and that in theorem 1.3.2 the p-topology for short.

In a sense the u-topology is rather more fundamental. If the system under consideration is defined on discrete time, hence taking the time set to be the integers Z, it is natural to insist that a sequence of segments converging to a given segment have domains of definition that are eventually equal, i.e., the characteristic of the u-topology. Admittedly, "naturalness" is somewhat subjective, but in attempting to bend the p-topology to yield results of the same power as the u-topology we have to resort to a measure theoretic device.

Topological semigroups have an extensive literature, a good bibliography being contained in [D3]. We shall have occasion to use only the most elementary properties of these objects.

In the next chapter these theorems will be sometimes invoked in a rather explicit form. For reference these are stated as

Corollary .1.3.1

With the u-topology, if $(u_n,v_n) \to (u,v)$ then $u_n \circ v_n \to u \circ v$.

Corollary 1.3.2

With the p-topology if $v_n \to v$ then $u \circ v_n \to u \circ v$.

1.4 Functions on $\mathcal{U}$

Every causal function $\hat{f}:\hat{\mathcal{U}} \to \hat{\mathcal{Y}}$ induces a function $f:\mathcal{U} \to \mathcal{Y}$ in the obvious manner, i.e., if $u \in \mathcal{U}$ and $\mathcal{D}(u) = [0,t)$, then $f(u) = E^t\hat{f}(u)$. For simplicity we shall hence forth denote the length of the domain of u by $|u|$, so that in this case $|u| = t$.

It should be clear that for such functions the output is synchronized with the input as a matter of convention, but this can be easily relaxed.

Lemma 1.4.1

f is continuous if $\hat{f}$ is continuous.

We shall have reason to modify $\hat{\mathcal{U}}$ to let it be the space of functions which are identified if they are equal almost everywhere (a.e.).

In such a case segments may be defined over closed or open intervals without affecting any of the previous results but allowing a modification of the function f. For any $u \in \mathcal{U}$, $|u| = t$, the function $\hat{f}$ cannot depend on the value to the point t, so that in fact the function f induced by $\hat{f}$ may in this instance be defined by either $f:\mathcal{U} \to Y$ or $f:\mathcal{U} \to \mathcal{Y}$, recalling that Y is the value space of segments in $\mathcal{Y}$. These are equivalent definitions when $\hat{\mathcal{U}}$ is a quotient space of functions identified if they are equal a.e.

One significance of the identification of functions equal a.e. is that instances of the proof scheme with this provision are the various L_p spaces which pervade continuous systems theory. More to the point is that in this case theorem 1.3.2 may be strengthened to coincide with theorem 1.3.1, and hence Corollary 1.3.2 will also coincide with Corollary 1.3.1. To see this we rely on the above observation that segments may now be defined over closed intervals, the values at the end points being irrelevant. If $\{u_n\}$ and $\{v_n\}$ are two sequences of segments converging in the modified p-topology to u and v respectively, the original difficulty of establishing pointwise convergence of $\{u_n \circ v_n\}$ to $u \circ v$ arose from the difficulty of establishing convergence at the time point $t = |u|$. Clearly this difficulty vanishes if the values at endpoints are irrelevant.

Henceforth when no explicit mention is made of the choice of topologies for $\mathcal{U}$ it is to be understood that if a conclusion is true for the u-topology, it is also true for the modified p-topology.

We can distinguish two IO system definitions. If $f:\mathcal{U} \to \mathcal{Y}$ is taken as basic we have a Mealy type system: if $f:\mathcal{U} \to Y$ is taken as basic we have a Moore type system.

Chapter 2

The theory of realizations of time-varying systems is complex and unsettled. Several elegant attempts to place the problems in perspective have been made ([B2],[Z6]), but as noted by Kalman [K1] and others these attempts have not successfully solved the problem of canonical realizations. In this chapter we shall address ourselves to the problem of time-invariant, causal systems, for which the algebraic problem of canonical realizations in the discrete-time case has well-known solutions. For completeness we state without proof some known results which are needed. We shall consider time-varying systems in Chapter 6, but not from a state space perspective.

2.1 Factorizations

By the state of a system at time t we mean sufficient information at that time which will permit the determination of future outputs given future inputs. It should be clear that only causal systems can be endowed with state sets. Causality in itself is an extremely general concept, but we shall keep to our specialization in section 1.1. The concept of a state set as employed in practice embodies an additional constraint which may be described as determinism. Determinism is the property of unique transitions from state to state, given an input, or equivalently it is coextant with the existence of a map from inputs to the function space of state-to-state functions. In automata theory this function space is known as the semigroup of the automaton, and we shall examine it later from our perspective.

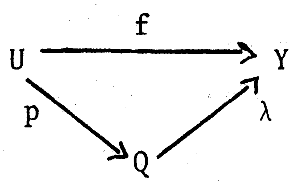
Definition 2.1.1

Given a system $f:\mathcal{U} \to \mathcal{Y}$, the triple (p,Q,λ) is a factorization of f if $p:\mathcal{U} \to Q$, $\lambda:Q \times \mathcal{U} \to \mathcal{Y}$, and $\lambda(p(u),v) = E_{|u|}f(u \circ v)$ for each $p(u) \in R(p)$, $v \in \mathcal{U}$. Q is called the <u>state set</u>. $p(\Lambda)$ is denoted by q_o and called the <u>initial state</u>.

It is noted that $\lambda(p(\Lambda),v) = f(v)$. In the subsequent development we shall need the notion of <u>output functions from a state</u>, denoted $\beta_q:\mathcal{U} \to \mathcal{Y}$, and defined for $p(u) \in R(p)$ by $\beta_{p(u)} = S_{-|u|} \circ \lambda$, so that $\beta_{q_o} = f$. It is not an accident that the Mealy formulation of an automaton was chosen here. The formalisms of chapter 1 in a sense forces this choice. However, when the p-topology is chosen for $\mathcal{U}$, the Moore formulation will prove acceptable. The reason for this is that in this instance $\mathcal{U}$ is in fact a quotient space where segments equal a.e. are identified, so that we may specify $f:\mathcal{U} \to \mathcal{Y}$ by alternatively specifying $f:\mathcal{U} \to Y$ when f is <u>causal</u>. We omit a formal demonstration of this because it is simple but tedious. Hence for this instance we may modify the last definition.

<u>Definition 2.1.2</u>

Given a system $f:\mathcal{U} \to Y$ the triple (p,Q,λ) is a <u>factorization</u> of f if $p:\mathcal{U} \to Q$, $\lambda:Q \to Y$, and the diagram



commutes. Q is called the <u>state-set</u>.

In this case again $\beta_{q_o} = f$.

We shall not take pains to distinguish between Moore and Mealy formulations unless the circumstances warrant it. It will be understood that whenever the p-topology is used for $\mathcal{U}$ we shall be at liberty to use the former.

The existence of factorizations in themselves is not sufficient to guarantee the existence of state sets with the property of determinism. For this we need p to have the characteristics of memory.

Definition 2.1.3 [A3]

$p:U \to Q$ is a memory function if $p(u_1) = p(u_2) \Rightarrow p(u_1 \circ u) = p(u_2 \circ u)$.

An alternative characterization of this is:

Lemma 2.1.1

Let M be the equivalence relation induced on $\mathcal{U} \times \mathcal{U}$ by $u_1 M u_2 \Leftrightarrow p(u_1) = p(u_2)$. Then p is a memory function if and only if M is a right congruence.

Proof:

Clear.

We formalize the concept of determinism as

Definition 2.1.4

A factorization (p,Q,λ) of $f:U \to y$ is deterministic if there exists a function $\delta:Q \times \mathcal{U} \to Q$ such that $\delta(p(u),v) = p(u \circ v)$ for all $p(u) \in R(p)$ and $v \in U$.

Thus in this terminology a distinction is made between causality and determinism. The former refers to a strictly input-output characteristic. Every deterministic system is necessarily causal, but not vice-versa. What this implies is that a causal system may be non-deterministic to the extent that δ is a non-deterministic "map" to equivalent (in the automaton-theoretic sense) states.

Lemma 2.1.2

(p,Q,λ) is a deterministic factorization if and only if p is a memory function.

Proof:

(=>) Let $p(u_1) = p(u_2)$. Then $p(u_1 \circ v) = \delta(p(u_1),v) = \delta(p(u_2),v) = p(u_2 \circ v)$ for all $v \in U$.

(<=) Define $\delta(p(u),v) = p(u \circ v)$. Clearly δ is a map, for $p(u_1) = p(u_2)$, $p(u_1 \circ v) = p(u_2 \circ v)$ for all $v \in \mathcal{U}$ by hypothesis. For $q \in Q/R(p)$, any arbitrary definition will suffice, say, the identity action.

Suppose we begin with a right congruence M on $\mathcal{U} \times \mathcal{U}$ and ask when does it induce a deterministic factorization? To answer this requires two approaches, one for the Mealy and one for the Moore formulation.

For $\varepsilon > 0$ let M be the relation on $\mathcal{U} \times \mathcal{U}$ defined by $uM_\varepsilon v \iff E_{|u|} f(u \circ z) = E_{|v|} f(v \circ Z) \quad Z \in U_\varepsilon$ where $U_\varepsilon = \{u \in U \mid |u| < \varepsilon\}$. It is clear that M_ε is an equivalence relation. Further, if $\varepsilon_1 > \varepsilon_2$ we have that $M_{\varepsilon_1} \sqsubseteq M_{\varepsilon_2}$ where $\sqsubseteq$ is the relation "**refines**".

Then $\{M_\varepsilon, \sqsubseteq\}_{\varepsilon>0}$ is a linear ordering with infimum $N = \bigcap_{\varepsilon>0} M_\varepsilon$ and supremum $M_f = \bigcup_{\varepsilon>0} M_\varepsilon$. For, $uNv \iff uM_\varepsilon v$ for each $\varepsilon>0$, and N is an equivalence relation; $uM_f v \iff uM_\varepsilon v$ for some $\varepsilon>0$, and M_f is an equivalence relation. Moreover N is a fact a right congruence relation because

$$uNv \iff \forall \varepsilon>0 \quad uM_\varepsilon v \Rightarrow \varepsilon>0 \quad u \circ w \circ Z \Rightarrow \forall \varepsilon>0 \quad u \circ w M_\varepsilon v \circ w \Rightarrow u \circ w N v \circ w.$$

In fact N is the Nerode equivalence which is to be explicitly defined in Definition 2.2.2. That N is the infimum follows from the fact that every M_ε is refined by N. Similarly, since for every $\varepsilon>0$ $M_\varepsilon \sqsubseteq M_f$, M_f is the supremum of the ordering.

When a Moore formulation is permissible the above discussion may be simplified. In this case $f:\mathcal{U} \to Y$ and M_f reduces to M_0, that is $\bigcup_{\varepsilon>0} M_\varepsilon = M_0$, because the relation $uM_f v = uM_0 v \iff f(u) = f(v)$ is refined by every M_ε.

Lemma 2.1.3

A right congruence M induces a deterministic factorization $(p_m, \mathcal{U}/M, \lambda_M)$ if M refines M_f.

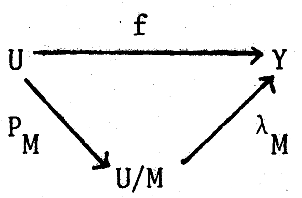
Proof:

(i) The Moore formulation:

Let $p_m : \mathcal{U} \to \mathcal{U}/M$, $\quad \lambda_m : \mathcal{U}/M \to Y$

$$u \to [u]_m \qquad [u]_m \to f(u)$$

We have to show



commutes, and that $\delta_M : \mathcal{U}/M \times \mathcal{U} \to \mathcal{U}_M$ is well-defined.

$$([u]_M, v) \to [u \circ v]_M$$

First, λ_M is well-defined because $u_1 M u_2 \Rightarrow f(u_1) = f(u_2)$ since M refines M_f. Next, δ_M is well-defined because M is a right congruence. Finally, let $u \varepsilon U$. Then $\lambda_M \cdot p_M(u) = \lambda_M [u]_M = f(u)$, so the diagram commutes.

(ii) The Mealy formulation:

Let $p_m : \mathcal{U} \to \mathcal{U}/M$, $\qquad \lambda_m : \mathcal{U}/M \times \mathcal{U} \to Y$

$$u \to [u]_m \qquad ([u]_m, v) \to E_{|u|} f(u\ v)$$

and $\delta_m : \mathcal{U}/M \times \mathcal{U} \to \mathcal{U}/M$

$$([u]_m, v) \to [u \circ v]_m$$

It has to be shown that each map is well-defined.

λ_m is well-defined because: $u_1 M u_2 \Rightarrow u_1 M_f u_2$, and since M is a right congruence, $u_1 \circ v M u_2 \circ v$. So $u_1 \circ v M_f u_2 \circ v$, giving $E_{|u_1 \circ v|} f(u_1 \circ v \circ z)$ for $z \varepsilon U_k$, $k > 0$. The certainly $E_{|u_1|} f(u_1 \circ v) = E_{|u_2|} f(u_2 \circ v)$.

δ_m is well-defined because M is a right congruence.

Every right congruence M on $\mathcal{U} \times \mathcal{U}$ associated with a deterministic factorization (p,Q,λ) of f certainly refines M_f. For the Moore case this is so because $p(u_1) = p(u_2) \Rightarrow \lambda p(u_1) = \lambda p(u_2) \Rightarrow f(u_1) = f(u_2)$. For the Mealy case, $(p(u_2),v)$ for all

$$v = E_{|u_1|} f(u_1 \circ v) = E_{|u_2|} f(u_2 \circ v)$$

for all $v \Rightarrow M \subseteq M_f$. Combining this with the previous lemma yields the following theorem.

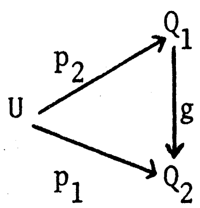
Theorem 2.1.1

A right congruence M on $U \times U$ is associated with a deterministic factorization if and only if M refines M_f.

Henceforth all factorizations will be assumed to be deterministic unless otherwise indicated.

2.2 Algebraic Theory

A given system $f: U \to Y$ may have many factorizations, and among these factorizations there may be some which are in some sense distinguished. A partial ordering $(G, \leqslant)$ of the set g of memory functions (of factorizations) of the system is laid down by saying $p_1 \leqslant p_2$ if the diagram



commutes for some surjective $g: Q_1 \to Q_2$. Call $p \varepsilon g$ minimal if it is minimal with respect to $\leqslant$.

Lemma 2.2.1

When $p_1 \leqslant p_2$ and $p_2 \leqslant p_1$, the two factorizations (p_1, Q_1, λ_1) may be completely identified.

Proof:

Since G is a partial ordering, when $p_1 \leq p_2$ and $p_2 \leq p_1$ we have $p_1 = p_2$. This equality is in fact a statement about the equipotency of Q_1 and Q_2, because $p_2 \leq p_1$ => there is a surjection $g:Q_1 \to Q_2$ => Q_2 is equipotent to a subset of Q_1, and conversely, $p_2 \geq p_1 \Rightarrow Q_1$ is equipotent to a subset of Q_2, hence by the Schroeder-Bernstein Theorem Q_1 is equipotent to Q_2. Thus we may completely identify (p_1,Q_1,λ_1) with (p_2,Q_2,λ_2).

Corollary 2.2.1

Equality in the partial ordering $(G,\leq)$ is an equivalence relation on factorizations.

Definition 2.2.1

A (p,Q,λ) factorization is reachable if p is surjective and observable if the map $\psi:Q \to y^U$ given by $\psi(q) = \beta_q$ is injective. It is canonical if it is both reachable and observable.

In the Moore formulation the observability criterion reduces to one concerning $\psi:Q \to Y^U$. When $p_1 = p_2$ and either is reachable Lemma 2.2.1 states that the other is reachable.

Theorem 2.2.1

A factorization is minimal if and only if it is canonical. Thus it is unique.

Proof:

See ([K1],[A3]).

Definition 2.2.2

$u_1, u_2 \varepsilon \mathcal{U}$ are Nerode-equivalent, written $u_1 N u_2$, if

$$S_{-|u_1|} E_{|u_1|} f(u_1 \circ v) = S_{-|u_2|} E_{|u_2|} F(u_2 \circ v)$$

for each $v \varepsilon \mathcal{U}$.

Theorem 2.2.2

N is a right congruence relation on $\mathcal{U}$.

Proof:

Clear.

Theorem 2.2.3

The Nerode factorization $(p, \mathcal{U}/N, \lambda)$ of $f: \mathcal{U} \to \mathcal{Y}$ given by $p: \mathcal{U} \to \mathcal{U}/N$, $\lambda: \mathcal{U}/N \times \mathcal{U} \to \mathcal{Y}$, where $p(u) = [u]$ and $\lambda([u], v) = E_{|u|} f(u \circ v)$, is canonical and deterministic.

Proof:

See [A3].

The notation $\mathcal{U}/N$ for a quotient class is standard. λ is well-defined because time-invariance allows the identification of all segments related by a shift. The above theorem gives in essence a reduced automaton in the Mealy sense with state transition function

$$\delta: \mathcal{U}/N \times \mathcal{U} \to \mathcal{U}/N$$

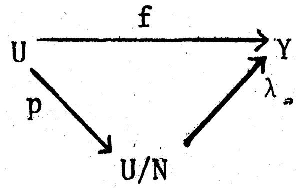
$$([u], v) \to [u \circ v]$$

which is well-defined because N is a right congruence.

For the Moore formulation Theorem 2.2.3 may be restated as

Theorem 2.2.4

The Nerode factorization $(p, \mathcal{U}/N, \lambda)$ of $f:\mathcal{U} \to Y$ given by $p:\mathcal{U} \to \mathcal{U}/N, \lambda:\mathcal{U}_N \to Y$ where $p(u) = [u]$ and $\lambda([u]) = f(u)$ is canonical and deterministic. In this case the diagram



commutes.

Example 2.2.1

$\hat{f}:\ell_2 \to \ell_2$ is defined by

$(x_1, x_2, \ldots) \longrightarrow (y_1, y_2, \ldots)$

where $y_n^2 = \sum_{i=1}^{n} x_i^2/r^n, \quad r > 1$

$$\sum_{n=1}^{\infty} y_n^2 = x_1^2(1/r + 1/r^2 \ldots) + x_2^2(1/r^2 + 1/r^3 \ldots) + \ldots.$$

(where rearrangement was justified by the possitivity of the terms)

$$\leqslant (x_1^2 + x_2^2 \ldots)/(r-1)$$

so that in fact the range of $\hat{f}$ is in ℓ_2.

Let $x = (x_1, x_2, \ldots, x_k, 0, 0, \ldots)$

$y = (y_1, y_2, \ldots, y_\ell, 0, 0, \ldots)$

$$xNy \Rightarrow \sum_{i=1}^{k} x_i^2/r^k = \sum_{i=1}^{\ell} y_i^2/r^\ell$$

Suffixing a segment $(1,0,0,\ldots)$ to x,y, $xNy \Rightarrow \left(\sum_{i=1}^{k} x_i^2 + 1\right)/r^{k+1} = \left(\sum_{i=1}^{\ell} y_i^2 + 1\right)/r^{\ell+1}$, showing that $k = \ell$ and the canonical state space is $R \times Z$.

Theorem 2.2.5

Let (p_1, Q_1, λ_1) and (p_2, Q_2, λ_2) be two factorizations such that $p_2 \leqslant p_1$, $g: Q_1 \to Q_2$, and let M_1, M_2 be the respective associated right congruences. Then M_1 refines M_2.

Proof:

Let $u_1 M_1 u_2$, so that $p_1(u_1) = p_1(u_2)$. Then $p_2(u_2) = gp_1(u_1) = gp_1(u_2) = p_2(u_2)$, or $u_1 M_2 u_2$.

Corollary 2.2.2

If M is the right congruence associated with a factorization then refines N (Nerode equivalence).

Proof:

If (p_N, Q_N, λ_N) is a Nerode factorization it is minimal by Theorems 2.2.1 and 2.2.3. Then p_N is minimal in $(g, \leqslant)$, and the result follows by the above theorem.

Let H = {M: M is a right convergence on UxU corresponding to a factorization of f}

Let $\sqsubseteq$ denote the relation "refines" on H. That $(H, \sqsubseteq)$ is a partial ordering is clear. However a stronger assertion is true.

Theorem 2.2.6

$(H, \sqsubseteq)$ is a complete lattice with inf H = N, and sup H = the discrete partition on UxU.

Proof:

For any subset $\{M\alpha\}_{\alpha \varepsilon A}$ of H $M_\alpha \sqsubseteq M_f$ for each α. then $\bigcap_{\alpha \varepsilon A} M_\alpha$ = inf $\{M_\alpha\}_{\alpha \varepsilon A}$. Clearly $\bigcap_{\alpha \varepsilon A} M_\alpha$ is a right congruence which refines M_α. Thus arbitrary infina exist, hence arbitrary suprema. The remainder of the assertions are easily verified.

By Theorem 2.2.4 it is evident that $(\mathcal{H},\subseteq)$ is anti-isomorphic to $(G,\leq)$ as lattices, revealing the theorem as a statement about a duality between alternative characterizations.

In this lattice of factorizations it is easily seen that if $p:\mathcal{U} \to Q_N$ is open, then $p_M:\mathcal{U} \to \mathcal{U}/M$ is open for each M refined by N.

2.3 Topological Theory

The algebraic theory of factorizations may be reinforced with the introduction of topologies on the state set. Suppose (p,Q,λ) is a factorization of $f:\mathcal{U} \to \mathcal{Y}$. We know from theorem 2.1.1 that there is a right congruence relation M associated with this factorization, so that $\mathcal{U}/M \subset Q$, and $\mathcal{U}/M = Q$ if and only if p is surjective. The natural topology to place on $\mathcal{U}/M$ is the quotient topology through p, that is, a set θ is open in $\mathcal{U}/M$ if $p^{-1}(\theta)$ is open in $\mathcal{U}$. Then $p:\mathcal{U} \to \mathcal{U}/M$ is continuous. Because $\mathcal{U}/M \subset Q$ for $p:\mathcal{U} \to Q$ to be continuous it is sufficient to say that the quotient topology on $\mathcal{U}/M$ is the relative topology of Q restricted to $\mathcal{U}/M$. This will be assumed throughout.

Example 2.3.1

In example 2.2.1 the quotient topology for the Nerode state space may be inferred as follows. A basic open set in R x Z with the natural topology is of the form θ x {m} where is an open interval and m is an integer. All preimages of this must be segments of length m. It can be shown that this set is pulled back into an open set in the segment space, and hence p relative to these topologies is continuous. Conversely, a basic open set in the segment space is mapped by p to some open set in R x Z, so in fact p is open. Hence the natural topology is the quotient topology.

Lemma 2.3.1

With the u-topology, if f is continuous

(i) M is closed

(ii) M[u] is closed for each $u \in \mathcal{U}$.

Proof:

(i) Let $(u,v) \in \overline{M}$. Then there is a sequence $(u_n,v_n) \in M$ such that $(u_n,v_n) \to (u,v) \circ E_{|z|} f(u_n \circ z) = E_{|z|}(v_n \circ z)$ for each $z \in \mathcal{U}$. By continuity of monoid composition $u_n \to u$, $v_n \to v \Rightarrow u_n \circ z \to v_n \circ z$ for each z. $E_{|z|}$ and f continuous now imply $E_{|z|} f(u_n \circ z) \to E_{|z|} f(u \circ z)$ and $E_{|z|} f(v_n \circ z)$ $E_{|z|} f(v \circ z)$, hence $(u,v) \in M$, and M is closed.

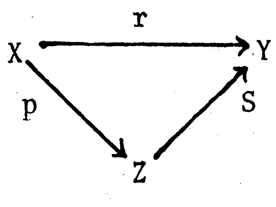
(ii) Let $v \in \overline{M[u]}$. Then there is a sequence $\{v_n\} \subset M[u]$, $v_n \to v$. So $v_n \circ z \to v \circ z$ for each $z \in \mathcal{U}$. But $E_{|z|} f(v_n \circ z) = E_{|z|} f(v \circ z)$ so that $E_{|z|} f(v \circ z) = E_{|z|} f(v_n \circ z) \to E_{|z|} f(v \circ z)$, showing $v \in M[u]$.

With the p-topology the lemma does not hold.

For use in later proofs we record the following easy

Lemma 2.3.2

Let X,Y,Z be topological spaces such that



commutes, r is continuous and Z has the quotient topology. Then S is continuous.

By $\mathrm{Hom}_R(\mathcal{U})$ is meant the monoid of all maps $\mathcal{U} \to \mathcal{U}$ with right composition serving as monoid composition. The following scheme yields the right regular representation Φ for a monoid $\mathcal{U}$.

For each $z \varepsilon \mathcal{U}$ define a map $\mu_z \; \varepsilon \; \mathrm{Hom}_R(\mathcal{U})$ by

$$\mu_z : \mathcal{U} \to \mathcal{U} : u \to u \circ z$$

Then the map $\Phi : \mathcal{U} \to \mathrm{Hom}_R(\mathcal{U}) : z \to \mu_z$ is a monoid anti-monomorphism. Thus the representation is <u>faithful.</u>

In a topological context more can be said. With suitable topologies on $\mathcal{U}$ and $\mathrm{Hom}_R(\mathcal{U})$, Φ may be continuous, and indeed a homeomorphism. A homeomorphism which is also an isomorphism is called an iseomorphism.

Since the u-topology for $\mathcal{U}$ is metrizable a <u>generalized metric</u> for $\mathrm{Hom}_R(\mathcal{U})$ is given by

$$\begin{aligned} d(\mu_z, \mu_w) &= \sup_u \; d_{\mathcal{U}}(\mu_z(u), \mu_w(u)) \\ &= \sup_u \; d_{\mathcal{U}}(u \circ z, u \circ w) \\ &= d_{\mathcal{U}}(z, w) \end{aligned}$$

Thus, Φ is actually even stronger than an iseomorphism in this case. We record this as

<u>Theorem 2.3.1</u>

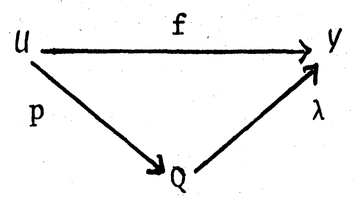
With the u-topology Φ is an isometric anti-monomorphism from into $\mathrm{Hom}_R(\mathcal{U})$.

In a more general case an anti-mono morphism may be asseted. It turns out that this weaker version of the theorem just proved is valid for the p-topology on $\mathcal{U}$ if the pointwise convergence topology is chosen for $\mathrm{Hom}_R(\mathcal{U})$. To see this, let $z_n \to z$. Then $\mu_{z_n}(u) = u \circ z_n$, which by Corollary 1.3.2 converges to $u \circ z$. Thus Φ is continuous. Conversely, let $\mu_{z_n} \to \mu_z$ pointwise. Then $\mu_{z_n}(u) = u \circ z \quad u \circ z = \mu_z(u)$ for each u. In particular, let $u = \Lambda$ the monoid identity. Then $z_n \to z$. Hence Φ^{-1} is continuous, showing that Φ is indeed an anti-moneomorphism. This is stated as a theorem.

Theorem 2.3.2

With the p-topology Φ is an anti-moneomorphism from U into $\mathrm{Hom}_R(U)$.

For a given factorization (p,Q,λ) of f, the continuity of p is assured by the way Q is topologized. What about the continuity of λ? In the Moore formulation the commutativity of the diagram



assures that λ is continuous if and only if f is continuous (lemma 2.3.2). In the Mealy formulation things are not quite as straightforward. Here $\lambda : Q \times U \to Y$, because $\lambda([u],v) = E_{|u|}f(u \circ v)$. The notion of convergence in $U/M \subset Q$ has to be carefully considered. These remarks shall serve as motivation for investigating the topological properties of state spaces.

Some standard results which will prove useful are the following.

Lemma 2.3.3

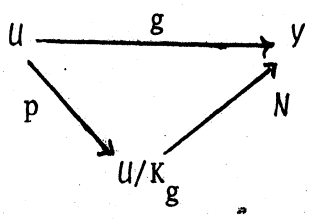
Let M_1 and M_2 be right congruences (or equivalences) on U such that $M_1 \sqsubset M_2$. Then the natural projections p_1, p_2 in

$$U \xrightarrow{p^1} U/M_2 \xrightarrow{p^2} U/M_1$$

are continuous.

Lemma 2.3.4

Let $g: \mathcal{U} \rightarrow \mathcal{Y}$, and K_g denote the kernel equivalence relation on $\mathcal{U}$ defined by $uK_g v \iff g(u) = g(v)$. Then



commutes, and g is continuous if and only if N is continuous.

For proofs of these lemmas any standard text on point-set topology may be consulted. A good reference is Dugundji [D4].

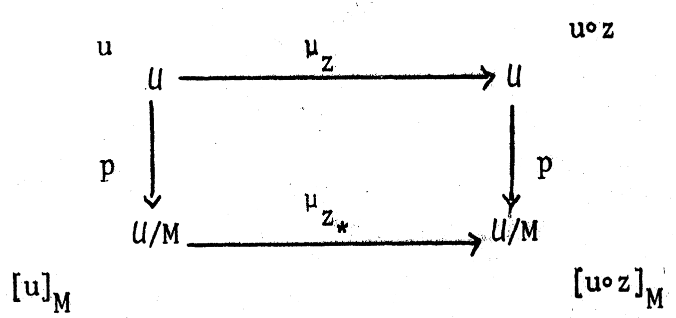
2.4 Topological Automata

Using the theory developed in the previous sections a theory of topological automata will be presented. For simplicity most of the results depend on the use of the u-topology, but applies to the modified p-topology as explained in the previous chapter.

The distinction between the Moore and Mealy formulation is displayed here. It will be noticed that the proofs in the latter case are almost invariably more lengthy although not much more conceptually difficult.

Lemma 2.4.1

The diagram



commutes and μ_{z*} is continuous. Consequently $\forall_z \varepsilon \mathcal{U}$ $[u_n] \to [u]$ => $[u_n \circ z] \to [u \circ z]$.

Proof:

Define μ_{z*} by $\mu_{z*}[u] = [u \circ z]$. This is well-defined because M is a right congruence, namely, if $u_1, u_2 \; \varepsilon \; p^{-1}[u]$, then $u_1 M u_2$ so that $u_1 \cdot z M u_2 \cdot z$ and hence $[u_1 \cdot z] = [u_2 \cdot z]$. The diagram commutes under this definition. To show continuity of μ_{z*} it is sufficient by lemma 2.3.2 to show that $p \cdot \mu_z$ is continuous. But this is true because p is continuous by definition and μ_z is continuous by corollary 2.3.1.

It should be observed that because of the weaker conclusion of corollary 1.3.2 this is not true for the strict p-topology.

In the terminology of factorizations it is seen that the mapping μ_{z*} acts as follows:

$$\mu_{z*}([u]) = \delta([u],z)$$

Thus, the above lemma is in fact a statement that $\delta : \mathcal{U}/M \times \mathcal{U} \to \mathcal{U}/M$ is continuous in $\mathcal{U}/M$.

Lemma 2.4.2

$[u_1] = [u_2]$ => $[u_1 \cdot z] = [u_2 \cdot z]$ for all $z \varepsilon \mathcal{U}$.

Lemma 2.4.3

Let $z_n \to z$. Then $[u \cdot z_n] \to [u \cdot z]$ and is independent of the class representative u.

Proof:

$z_n \to z$ => $v \cdot z_n \to v \cdot z$ by corollary 1.3.1 (or 1.3.2). By continuity of p, $[v \cdot z_n] \to [v \cdot z]$. Now let $u \; \varepsilon \; p^{-1}[v]$. Then $[u] = [v]$, and by the previous lemma $[u \cdot z] = [v \cdot z]$ and $[u \cdot z_n] = [v \cdot z_n]$. Thus the convergence is independent of class representative.

Combining lemmas 2.4.1 and 2.4.3 leads to the following result.

<u>Theorem 2.4.1</u>

In any factorization $(p,\mathcal{U}/M,\lambda)$ of a system f the <u>state transition</u> function $\delta:\mathcal{U}/M \times \mathcal{U} \to \mathcal{U}/M$ is continuous in each component.

<u>Proof</u>:

Only continuity in $\mathcal{U}$ need be demonstrated, but this is precisely lemma 2.4.3.

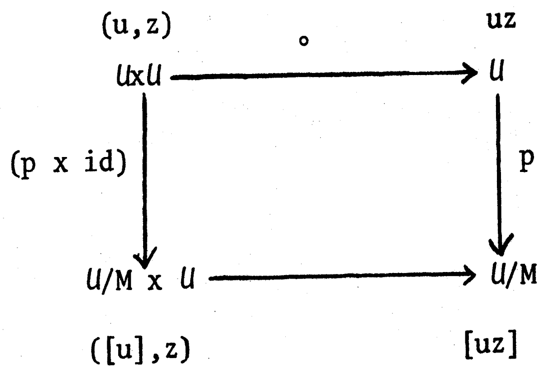
This theorem is strengthened when p is open.

<u>Theorem 2.4.2</u>

If p is open, $\delta:\mathcal{U}/M \times \mathcal{U} \to \mathcal{U}/M$ is continuous.

<u>Proof</u>:

Consider the diagram:



It is clearly commutative. Let $\phi_{[uz]}$ be an (open) neighborhood of [uz] in $\mathcal{U}/M$. By the continuity of p and 0 the set $\Phi = 0^{-1}p^{-1}(\theta_{[uz]})$ is open. Let (u,z) be a preimage of uz in Φ. Then there is a basic open set $\Phi_u \times \Phi_z \subset \Phi$, with $u \in \Phi_u$ and $z \in \Phi_z$, and Φ_u, Φ_z open in $\mathcal{U}$. $(p \times id)(\Phi_u \times \Phi_z) = p\Phi_u \times \Phi_z$ which is open since p is open, hence ([u],z) is an interior point of $p\Phi_u \times \Phi_z$. Thus interior points are mapped by (p x id) to interior points showing that (p x id) is open. Φ is open and (p x id) is open now imply $(p \times id) = \delta^{-1}(\theta_{[uz]})$ is open, so that δ is continuous.

Again this proof does not go through for the p-topology.

We observe that theorems 2.4.1 and 2.4.2 do <u>not</u> assume that f is continuous. This being so it is clear that we may now answer the question posed after theorems 2.3.2.

<u>Theorem 2.4.3</u>

In the Mealy formulation the output function $\lambda: Q \times \mathcal{U} \to \mathcal{Y}$ is continuous in each component if and only if $f: \mathcal{U} \to \mathcal{Y}$ is continuous.

<u>Proof</u>:

(=>) Because $f(u) = \lambda(p(\Lambda), u)$.

(<=) For some sequence $u_n \to u$ and v, $v \circ u_n \to v \circ u$ by corollary 1.3.1, so $f(v \circ u_n) \to f(v \circ u)$, hence $\lambda([v], u_n) = E_{|v|} f(v \circ u_n) \to E_{|v|} f(v \circ u)$. Therefore λ is continuous in $\mathcal{U}$.

For some sequence $[u_n] \to [u]$ and any $z \in \mathcal{U}$, by lemma 2.4.1 $[u_n \circ z] \to [u \circ z]$.

Define a map $g: \mathcal{U} \to \mathcal{Y}$ by $g(u) = E_{|u|} f(u \circ z)$ and a relation M_z on $\mathcal{U}$ such that $u M_z v \iff g(u) = g(v)$, that is, M_z is the kernel equivalence of g. It is clear that M refines M_z. Now apply lemma 2.3.4 with $K_g = M_z$. g is continuous because (i) right composition by z is continuous (ii) f is continuous, and (iii) $E_{|u|}$ is continuous. Thus η is continuous, where $\eta[u]_{M_z} = E_{|u|} f(u \circ z)$.

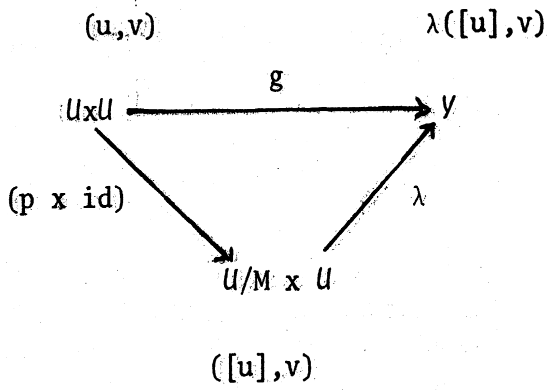
$[u_n \circ z] \to [u \circ z]$ now imply (by lemma 2.4.1) that $[u_n \circ z]_{M_z} \to [u \circ z]_{M_z}$. Hence $\lambda([u_n], z) = E_{|u_n|} f(u_n \circ z) = \eta[u_n]_{M_z} \to \eta\, [u]_{M_z} = E_{|u|} f(u \circ z) = ([u], z)$. Hence λ is continuous in $\mathcal{U}/M$.

<u>Theorem 2.4.4</u>

If p is open, $\lambda: \mathcal{U}/M \times \mathcal{U} \to \mathcal{Y}$ is continuous if and only if f is continuous.

Proof: (<=)

Consider the diagram



where g is defined by $g:(u,v) \to E_{|u|}f(u \circ v)$. The diagram commutes, g is continuous since f is continuous, and p x id is open. Therefore λ is continuous.

(=>) $f(u) = \lambda(p(\Lambda),v)$.

The above proofs could be considerably simplified if the space U has the following property: If M is a right congruence (or equivalence) relation on U, and if a sequence $\{[u_n]\}$ in the quotient space U/M converges to [v], then there exists a sequence $\{x_n\}$ in U converging to x in U with $x_n \in [u_n]$ (regarded as a set) and $x \in [v]$ (regarded as a set). In fact a weaker condition may suffice, viz., that x can be replaced by a sequence $\{y_n\}$, each $y_n \in [v]$, and $d(x_n,y_n) \to 0$. One example where such a property is indeed present is when U is a linear space. We discuss this in the next chapter.

Definition:

A semigroup act is a continuous function $S \times X \to X$ where S is a topological semigroup and X is a Hausdorff space.

Wallace [D3] and his students have investigated semigroups from the viewpoint of semigroup acts. In the context of the foregoing results the state $\mathcal{U}/M$ (or Q) is a semigroup act if $\mathcal{U}/M$ is Hausdorff and δ is continuous.

2.5 State Spaces

When p is open δ is continuous, so satisfying part of the condition for semigroup acts. The other part deals with $\mathcal{U}/M$, specifically with its separation properties. Here several sufficient conditions for $\mathcal{U}/M$ to be Hausdorff are given.

Lemma 2.5.1

p is open, and $\mathcal{U}$ Hausdorff and M is closed => $\mathcal{U}/M$ is Hausdorff.

Proof:

See [D4], where the proof is given for M an equivalence relation, and observe that the same proof holds for M a right congruence relation.

Corollary 2.5.1

p is open => $\delta:\mathcal{U} \times \mathcal{U}/M \to \mathcal{U}/M$ is a monoid act.

Proof:

By lemmas 2.3.1, theorem 2.4.2, and the fact that $\mathcal{U}$ is Hausdorff.

Lemma 2.5.2

If f is continuous and $\mathcal{Y}$ is Hausdorff then $\mathcal{U}/N$ is Hausdorff.

Proof:

Let $[u_1] \neq [u_2]$ in U/N. Then there is a z in U such that $([u_1],z) \neq \lambda([u_2],z)$. Consider the map

$$\lambda_z : U/N \rightarrow Y$$

$$[u] \rightarrow \lambda([u],z)$$

which is certainly well-defined. Since f is continuous, by theorem 2.4.3 λ_z is continuous. The result now follows immediately from the assumption that Y is Hausdorff.

Corollary 2.5.2

f is continuous and Y Hausdorff =>

$$\delta : U \times U/N \rightarrow U/N \text{ is a monoid act.}$$

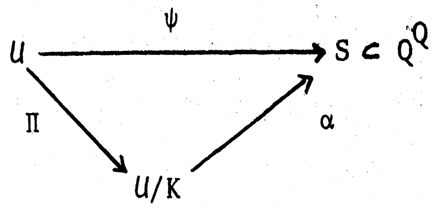
Stronger separation properties are obtainable only with stronger assumptions on U, p, or f. When U has a certain structure (e.g. a Banach space) the state space U/M may have more or less structure (e.g. a Hilbert space or an Abelian group) depending on f. For examples of these situations consult the next chapter.

2.6 Automaton Semigroups

The semigroup of our automaton is the semigroup S of functions $\{\delta_s\}_{s \in U}$ which map $Q \rightarrow Q$, defined by $\delta_s(q) = \delta(q,s)$, and having as composition $\delta_{s_1} \circ \delta_{s_2} = \delta_{s_1 \circ s_2}$. The homomorphic map $\psi : U \rightarrow Q^Q$ with $\psi(u) = \delta_u$ gives rise to a kernel equivalence K on U which is in fact a congruence relation called Myhill equivalence. So, algebraically, U/K is isomorphic to S. Unfortunately, without further assumptions it is not possible to assert that U/K and S are iseomorphic in the topological context. Standard results [D4] are:

<u>Lemma 2.6.1</u>

With the above notation, the diagram



commutes, and α is an iseomorphism if and only if ψ is an identification (i.e., if S has the quotient topology).

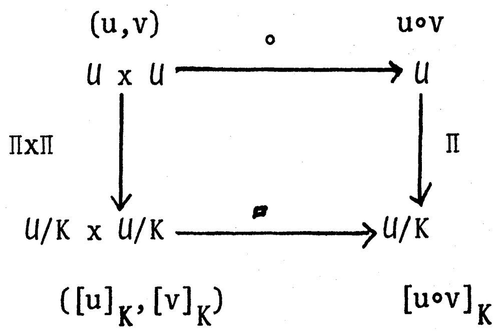
<u>Lemma 2.6.2</u>

If U/K is compact and S is Hausdorff, then α is an iseomorphism.

<u>Theorem 2.6.1</u>

If Π is open δ is continuous and Q is Hausdorff, $U/K \times Q \to Q$ is a semigroup act.

<u>Proof</u>: Consider the diagram:



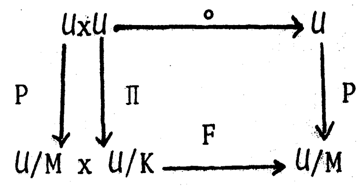
It is commutative because K is a congruence relation. The composition ∘ is continuous since U is a topological monoid, and Π is continuous by definition. Because Π is open it follows that Π x Π is open, and by diagram chasing it is seen that the composition ■ is continuous. Therefore $(U/K, ■)$ is a topological semigroup.

Natural topologies for Q^Q are the compact-open and point-open topologies. See [D4] for details on these topologies. There is a simple relation between these topologies for S (having the relative topology) and that of U/K.

Lemma 2.6.3

Let $p:\mathcal{U} \to \mathcal{U}/M \subset Q$ and $\Pi:\mathcal{U} \to \mathcal{U}/K$ be open. Then the map $F:\mathcal{U}/M \times \mathcal{U}/K \to \mathcal{U}/M$ given by $F([u]_M,[v]_K) = [uv]_M$ is jointly continuous. Hence the quotient topology for $\mathcal{U}/K$ is stronger than the compact-open topology for $S(=\mathcal{U}/K$ algebraically).

Proof: Consider the diagram



which commutes. F is well-defined because K refines M, that is, $v_1Kv_2 \Rightarrow \delta(q,v_1) = \delta(q,v_2)$ for $q \in Q \Rightarrow p(v_1) = p(v_2) \Rightarrow v_1Mv_2$.

By the familiar argument, $p \times \Pi$ is open, so F is continuous.

The last part of the lemma is a statement of the well known fact [D4] that the compact-open topology is the weakest topology for such joint continuity.

Corollary 2.6.1

If Π and p are open and S has the compact open topology then ψ is continuous.

Proof: α is continuous, so ψ is continuous.

2.7 Noncanonical state spaces

Prior to this it was assumed that state spaces are quotients of the input segment space and received the quotient topology. Suppose we begin by selecting an arbitrary state space Q and some topology for it such that $p_1:\mathcal{U} \to Q$ is continuous. The continuity assumption is a very natural one because it is equivalent to insisting that two input segments "near" each other transfer the initial state to final states "near" each other. The reachable portion $p_1(U) \subset Q$ of the state space still posesses the <u>algebraic</u> structure of being a quotient of U with respect to some right congruence M, but in this case it receives its topology from Q, namely the relative topology.

$p_1(U)$ can also be given the quotient topology of U via p_1. Call the relative topology τ_R and the quotient topology τ_F. Because the quotient topology is the strongest such that p_1 is continuous, in general τ_R is weaker than τ_F. If p_1 is also open then $\tau_R = \tau_F$, since in this case τ_R must be the quotient topology.

As the state space of an automaton $p_1(U)$ may be reduced in the standard manner. Denote its reduction by $p_1(U)^R$. Algebraically this is identical to the Nerode state space Q^N. Topologically, when $p_1(U)$ has the topology S and $p_1(U)^R$ is given the quotient topology via $p_R:p_1(U) \to p_1(U)^R$, denoted by S_R, the space $(p_1(U)^R, S_R)$ need not be homeomorphic to (Q^N, S_N) where S_N is the Nerode space quotient topology via $p:U \to Q^N = U/N$. In general S_R is weaker than S_N. The following theorem summarizes the above conclusions and introduces a characterizati[illegible] for topological uniqueness of state spaces.

<u>Theorem 2.7.1</u>

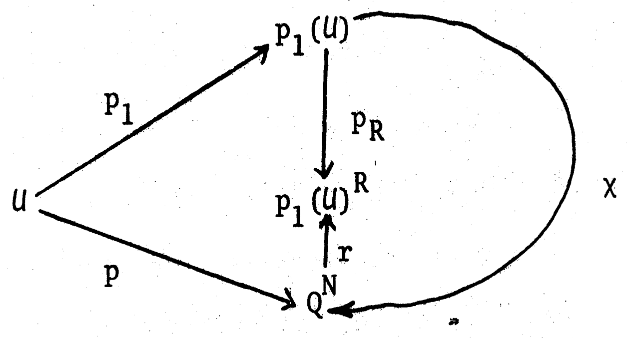
(1) $p_1(U)^R$ has the cardinality as Q^N

(2) Identifying $p_1(U)^R$ with Q^N algebraically, $S_R \subseteq S_N$

(3) If $\chi:p(U) \to Q^N$, then χ is continuous if and only if $p_1(U)^R$ is homeomorphic to Q^N.

(4) χ is continuous if p_1 is also open.

Proof:



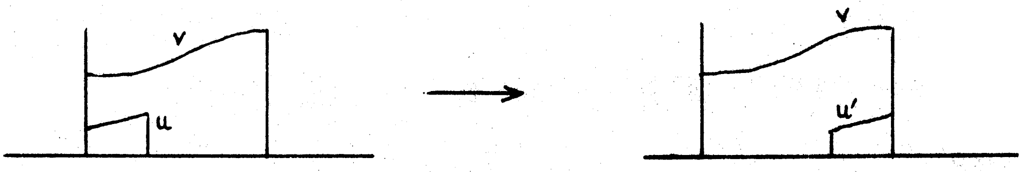
In view of earlier results only (3) remains to be proved. This is easily done by the appropriate pulling back of open sets, using the commutativity of the above diagram.

Note: A preliminary version [F1] of the preceding theory was presented at the 7th Princeton Conference, March 1973. Attention is also drawn to an alternate theory developed by Onat and Geary [O1] which came to my notice after the above investigation was completed.

Chapter 3

3.1 Linear Systems

The theory developed in the preceding two chapters may be specialized to the case when $\hat{f}$ is linear, that is, some linear structure is presumed for $\mathcal{U}$ and $\mathcal{Y}$, and f is a homomorphism of these spaces. For a realization to be possible $\hat{f}$ is assumed to be causal. To keep matters simple we shall assume that $\mathcal{U}$ and $\mathcal{Y}$ are normed linear spaces with the norm arising from norms in linear spaces U and Y in the way specified in Chapter 1. The resulting segment space $\mathcal{U}$ may be made into a linear space by the following device. (Refer to the diagram below.)



For two segments u, v addition $\oplus$ is defined by

$$u \oplus v = u' + v$$

where $u' = S_{|v|-|u|}u$ and u'+v is the segment obtained by pointwise addition, and is defined on $[0,|v|)$. Scalar multiplication is defined in the obvious way. Using this scheme the zero of the space is not well-defined unless we agree to identify every zero-valued segment. If this is done and we factor it out the resulting quotient space of $\mathcal{U}$ is the required linear space. We call it $\mathcal{U}'$.

For convenience we consider the Moore formulation, so $f:\mathcal{U}' \to Y$.

Lemma 3.1.1

$f(u) = f(S_\tau u)$ for all $u \in \mathcal{U}$, and all $\tau \in R_+$.

Proof:

Follows from the causality and linearity of $\hat{f}$ and a standard result in systems theory [N1] which states that the zero state of a causal linear system is unaffected by zero-valued inputs.

Lemma 3.1.2

f is linear on $\mathcal{U}'$.

Proof:

Let $u, v \in \mathcal{U}$ with $|v| > |u|$ and α be a scalar.

$$\begin{aligned} f(u + \alpha v) &= f(u' + \alpha v) && \text{by lemma 3.1.1} \\ &= f(u') + \alpha f(v) && \text{by linearity of } f \\ &= f(u) + \alpha f(v) && \text{by lemma 3.1.1} \end{aligned}$$

The next result is essentially due to Kalman [K1] in a different formulation.

Lemma 3.1.3

$$uNv \quad \text{iff} \quad u \ominus v \in \ker f \qquad \text{for } u,v \in \mathcal{U}'.$$

Proof:

$$\begin{aligned} uNv \quad &\text{iff} \quad f(u \cdot z) = f(v \cdot z) \qquad z \in \mathcal{U} \\ &\text{iff} \quad f(u) + f(S_{|u|} z) = f(v) + f(S_{|v|} z) \\ &\text{iff} \quad f(u) = f(v) && \text{by lemma 3.1.1} \\ &\text{iff} \quad f(u \ominus v) = 0 && \text{by linearity.} \end{aligned}$$

$\mathcal{U}'$ may be given a norm by embedding (and extending by zero) its elements into $\hat{\mathcal{U}}$ and defining the norm of u in $\mathcal{U}'$ to be that of its embedding in $\hat{\mathcal{U}}$. We have to show that this indeed defines a norm. Given u and v in $\mathcal{U}'$, we first note that $||u'|| = ||u||$ since translation is an isometry, and hence

$$||u \oplus v|| = ||u' + v|| \leq ||u'|| + ||v||. \quad \text{Also } ||\alpha u|| = |\alpha| \, ||u||.$$

So indeed this is a norm. The topological theory for linear factorizations can now be derived as a special case of the more general theory. For this we require the assumption that $\hat{f}$ and hence f be continuous.

The next lemma is a standard result from functional analysis [T1] paraphrased for our requirements.

Lemma 3.1.4

Let $\mathcal{U}'$ be a normed linear space and K be a closed subspace. Then $\mathcal{U}'/K$ is a normed linear space with norm

$$||[u]|| = \inf_{k \in K} ||u + k||$$

Wherever $f:\mathcal{U}' \to Y$ is continuous the kernel K of f is closed. Hence $\mathcal{U}'/K$ is a normed linear space. However by lemma 3.1.3 the space $\mathcal{U}'/K$ is algebraically the same as $\mathcal{U}'/N$. The next lemma, again a standard result, shows that they are in fact topologically identical too, so that the theory for topological realizations of linear systems is a specialization of the earlier theory.

Lemma 3.1.5

The norm defined in lemma 3.1.4 generates the quotient topology on $\mathcal{U}'/K$.

Further, it is well known that the natural projection $p:\mathcal{U}' \to \mathcal{U}'/K$ is linear and continuous in this case. Hence $\lambda:\mathcal{U}'/K \to Y$ must also be linear and continuous by virtue of f being so. In fact p is also open [T1]. In this instance the condition on theorem 2.4.2 is satisfied, so the state transition function $\delta:\mathcal{U}'/K \times \mathcal{U}' \to \mathcal{U}'/K$ is continuous. It is easy to verify that δ is actually linear as well. Therefore we have:

Theorem 3.1.1

If $f:\mathcal{U} \to Y$ is linear and continuous then it has a canonical state space which is linear, and the state transition and output functions are also continuous and linear.

3.2 Examples

A criticism of several theories of topological factorizations for causal nonlinear systems ([D1],[S2]) is that a linear space structure is posited for the canonical state space. In the previous section it was shown that this is justifiable when the input-output map is linear. Here we wish to provide some examples to illustrate the theory and show that in general the canonical state space need not be linear. Furthermore it is not clear at this moment how to tell by examining the input-output map itself whether its canonical state-space will have more or less structure than the input or output space. It appears that both possibilities exist.

Example 3.2.1

$\hat{f}:\ell_1 \to \ell_1$ is defined by

$(a_1,a_2,\ldots) \longmapsto (\text{sgn}(a_1),\ \text{sgn}(a_1+a_2),\ldots)$

where $\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ 0 & x = 0 \end{cases}$

Then $f:\mathcal{U} \to R$

where $\mathcal{U}$ are initial segments from ℓ_1.

f is nonlinear, time-invariant, causal and discontinuous.

$$u^{(1)}Nu^{(2)} \iff \sum_{i=1}^{k_1} u_i^{(1)} = \sum_{i=1}^{k_2} u_i^{(2)}$$

where k_1, k_2 are the "lengths" of segments $u^{(1)}$, $u^{(2)}$ respectively.

$$\delta([u],v) = \sum u_i + \sum v_i$$

where the summations run over the "lengths" of u and v, for any $u \in [u]$.

δ is continuous

$\lambda([u]) = f(u)$ and λ is discontinuous.

Remarks: In this example the input space is a normed linear space. The state space is R. We have a state transition function which is continuous and the discontinuities of f are reflected in λ only.

Example 3.2.2

$\hat{f}: \ell_2 \to \ell_2$ is defined by

$(x_1, x_2, \ldots) \to (y_1, y_2, \ldots)$

where $y_n^2 = \sum_{i=1}^{n} x_i^2/r^n, \quad r > 1$

$$\sum_{n=1}^{\infty} y_n^2 = x_1^2(1/r + 1/r^2 \ldots) + x_2^2(1/r^2 + 1/r^3 \ldots) + \ldots.$$

(where rearrangement was justified by the possitivity of the terms)

$$\leqslant (x_1^2 + x_2^2 \ldots)/(r-1)$$

so that in fact the range of $\hat{f}$ is in ℓ_2.

Let $x = (x_1, x_2, \ldots, x_k, 0, 0, \ldots)$

$y = (y_1, y_2, \ldots, y_\ell, 0, 0, \ldots)$

$$xNy = \sum_{i=1}^{k} x_i^2/r^k = \sum_{i=1}^{\ell} y_i^2/r^\ell$$

Suffixing a segment (1,0,0,...) to x,y, $xNy = \left(\sum_{i=1}^{k} x_i^2 + 1\right)/r^{k+1} = \left(\sum_{i=1}^{\ell} y_i^2 + 1\right)/r^{\ell+1}$, showing that $k = \ell$ and the canonical state space is R x Z.

Remarks: In this example the input space is a Hilbert space but the canonical state space is not even a linear space although it is a Z module.

Example 3.2.3

$\hat{f}:BC[0,\infty) \rightarrow PC[o,\infty)$ is defined by

$$(\hat{f}x)(t) = \begin{cases} \sup\limits_{s<t} x(s) & \text{if there are an even number of zeros of } x \text{ in } [0,t)\cdot \\ \inf\limits_{s<t} x(s) & \text{otherwise.} \end{cases}$$

BC and PC are the linear spaces of bounded continuous and piecewise continuous functions respectively.

Without going through a formal proof we can see that the canonical state space of f is R x R x {0,1}. The first R records the infinum up to t, and the second R records the supremum up to t. The set {0,1} records the condition "an even number of zeros" or "otherwise". Observe that "otherwise" includes the cases of both an infinitely denumerable and an uncountable number of zeros.

Remarks: In the previous two examples the Nerode spaces do retain a linear structure if the continuous time versions were used. Here, however, even with a continuous time structure the Nerode space cannot be given a linear structure since {0,1} is simply a two element set.

Chapter 4

The concept of system homomorphism was explained in Chapter 0 as an idealization of a common technique in modeling and simulation. It was also emphasized that there are many types of homomorphisms. In this chapter we focus on input-output homomorphisms in which the input segment space map from base to lumped model is a semigroup homomorphism. From the definition given below it is seen that this corresponds to the common procedure of batching events in discrete simulation languages. For convenience we shall work with Moore-type specifications, the extension to the Mealy-type specifications being quite clear from preceding developments.

4.1 Semigroup Input-Output Homomorphisms

A base system $f:\mathcal{U} \rightarrow Y$ and a lumped system $f_1:\mathcal{U}_1 \rightarrow Y_1$ are presumed to exist in a certain relationship which is to be defined. The notation is that of the previous chapters.

Definition 4.1.1

A semigroup input-output homomorphism (sioh) from f to f_1 is a pair (h,k) of maps $h:\mathcal{U} \rightarrow \mathcal{U}_1$, $k:Y \rightarrow Y_1$ where h is a monoid homomorphism, h,k are surjective, and $kf = f_1 h$.

Observe that if $h:\mathcal{U} \rightarrow \mathcal{U}_1$ is a monoid homomorphism the kernel equivalence induced by h is a congruence relation, and $\mathcal{U}$ modulo this convergence relation is isomorphic to $\mathcal{U}_1$.

From the results of the previous chapters we know that state spaces exist for f and f_1. Given state spaces Q and Q_1 for these systems respectively, and with the standard notation for state transition and output functions, denote the corresponding system representations by $(\mathcal{U},Y,Q,\delta,\lambda)$ and $(\mathcal{U},Y_1,Q_1,\delta_1,\lambda_1)$.

Definition 4.1.2

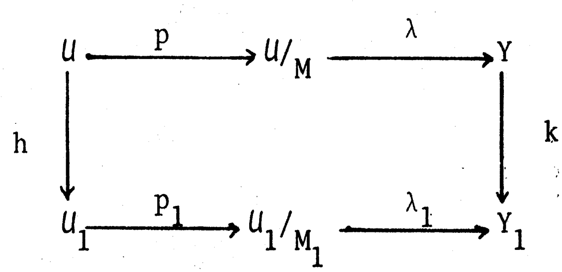
A homomorphism from (U,Y,Q,δ,λ) to $(U_1,Y_1,Q_1,\delta_1,\lambda_1)$ is a triple (h,g,k) where (h,k) is a sioh, $g:Q \to Q_1$ is surjective, and $\delta_1(g(q), h(u)) = h(\delta(q,u))$ and $\lambda_1(g(q)) = k(\lambda(q))$ for all $q\varepsilon Q$, $u\varepsilon U$.

Historically a version of this type of homomorphism was used extensively in algebraic structure theory of automata [H1].

The connection between sioh and homomorphisms was essentially observed by Zeigler [Z2] in a different context where sioh are analogous to function homomorphisms. For completeness we shall briefly outline that development and proceed to extend it to display very satisfying algebraic and topological properties of the collection M of all sioh images of a given system.

In accordance with the earlier results let us assume that Q and Q_1 arise from right congruences on U and U_1, and hence that these state spaces are reachable. We have already seen in section 2.2 that the state realizations associated with a given system f form a complete lattice. Here a similar result will be derived concerning M.

The present situation is:



where $f = \lambda\cdot p$, $f_1 = \lambda_1\cdot p_1$ and M,M_1 are right congruences on U and U_1.

Lemma 4.1.1

The composition of two memory functions is a memory function.

Proof:

Clear.

<u>Corollary 4.1.1</u>

$p_1 \cdot h$ is memory function.

<u>Proof</u>:

Semigroup homomorphisms are memory functions.

This suggests that there ought to be a relationship between $\mathcal{U}/_M$ and $\mathcal{U}_1/_{M_1}$ under suitable circumstances. Indeed there is such a relationship when M_1 is the Nerode-equivalence of f_1. Before proceeding to elucidate it let us see what has to be shown if it is asserted that there is a map $g:\mathcal{U}/_M \to \mathcal{U}_1/_{M_1}$ which satisfies the conditions of definition 4.1.2. First, it must be the case that if uMv for $u,v \varepsilon \mathcal{U}$ then $h(u)M_1h(v)$. With this it follows that g may be defined by $g([u]_M) = p_1(hu) = [hu]_{M_1}$. Also $k\lambda = \lambda_1 g$, for $k\lambda[u]_M = kf(u) = \lambda_1 p_1(hu) = \lambda_1 g[u]_M$. The state transition relation is also satisfied, for $g\delta([u]_M,v) = g[uv]_M = [h(uv)]_{M_1} = [(hu)(hv)]_{M_1} = \delta_1([hu]_{M_1},hv) = \delta_1(g[u]_M,hv)$. Thus, to establish a homomorphism given a sioh it is sufficient to show that uMv => $h(u)M_1h(v)$. The next lemma gives a sufficient condition for this to be the case.

<u>Lemma 4.1.2</u>

If $M_1 = N_1$ the Nerode equivalence on $\mathcal{U}_1$ induced by f_1 then uMv => $h(u)M_1h(v)$.

<u>Proof</u>:

$$
\begin{aligned}
uMv &\Rightarrow f(u\omega) = f(v\omega) \qquad \text{all } \omega \varepsilon \mathcal{U} \\
&\Rightarrow kf(u\omega) = kf(v\omega) \\
&\Rightarrow f_1h(u\omega) = f_1h(v\omega) \\
&\Rightarrow f_1[h(u)h(\omega)] = f_1[h(v)h(\omega)]
\end{aligned}
$$

So $h(u)N_1h(v)$.

Corollary 4.1.2

If (h,k) is a sioh from f to f_1 there exists a map $g: \mathcal{U}/_M \to \mathcal{U}/_{N_1}$, where $\mathcal{U}/_{N_1}$ is the canonical state space of f_1, which makes (h,g,k) a homomorphism.

The connection between these results and those of section 2.2 should be clear when h,k are taken to be the identity maps. The complete lattice of section 2.2 is in fact a lattice of homomorphisms in the special case where the input spaces are identified, and the output spaces are identified.

4.2 The Lattice of Homomorphisms

The remarks concerning the lattice of homomorphisms may be generalized. It turns out that given a system $f: \mathcal{U} \to Y$ the class $\mathcal{M}$ of all its I-O homomorphic images forms a complete lattice under appropriate circumstances. When this is the case corollary 4.1.2 says that if consideration is restricted to the canonical state factorization of systems, $\mathcal{M}$ is in fact a complete lattice of homomorphic images of $(p, \mathcal{U}/_N, \lambda)$, the canonical factorization of f. f is the base system and its sioh images are the corresponding lumped systems.

If f_1 and f_2 are lumped systems then a relation $\sqsubseteq$ is defined by saying that $f_1 \sqsubseteq f_2$ if f_2 is a sioh image of f_1. If the collection of lumped systems of f is denoted by $\mathcal{C}$, $(\mathcal{C}, \sqsubseteq)$ is a partial ordering.

Let $\mathcal{R}$ be the complete lattice of congruence relations on $\mathcal{U}$ and $\mathcal{S}$ be the complete lattice of equivalence relations on Y, both ordered by refinement denoted by $\leqslant$. Define a subset $\mathcal{D}$ of $\mathcal{R} \times \mathcal{S}$ by

$$(R,S) \varepsilon \mathcal{D} \iff \{uRv \Rightarrow f(u)Sf(v) \; \forall u,v \varepsilon \mathcal{U}\}$$

Lemma 4.2.1

$\mathcal{D}$ is a sublattice of $\mathcal{R} \times \mathcal{S}$ isomorphic to $\mathcal{C}$.

Proof:

Given f_1 in C there exist maps h,k and kernel congruence R and equivalence S induced by them respectively. If uRv then h(u) = h(v), so $f_1h(u) = f_1h(v)$. Hence kf(u) = kf(v), and f(u)Sf(v). Thus $(R,S)\varepsilon\mathcal{D}$. Conversely, every pair $(R,S)\varepsilon\mathcal{R} \times \mathcal{S}$ defines natural maps h,k such that (h,k) is an sioh to $f_1:\mathcal{U}/_R \to Y/S$. Given $f_2 \subsetneq f_1$ there exist pairs (R_1,S_1) and (R_2,S_2) corresponding to (h_1,k_1) and (h_2,k_2). Because f_1 is a homomorphic image of f_2 it is clear that $(R_2,S_2) \leqslant (R_1,S_1)$.

Define the index set I by

$$I = \{\alpha:S_\alpha \ \varepsilon \ \Pi_2 \ (\mathcal{D})\}$$

where Π_2 is the projection onto the second component. If y_1,y_2 are in range f, say that $\mathcal{D}$ is <u>f-closed</u> if $y_1S_\alpha y_2$ for all α in I implies $y_1 \bigcup\limits_\alpha S_\alpha y_2$.

The importance of this lies in the fact that it is a condition for completeness. That $\mathcal{D}$ is a sublattice of $\mathcal{R} \times \mathcal{S}$ is in effect a restatement in input-output terms of a fact known about states in SP partitions [H1]. However, without additional hypotheses $\mathcal{D}$ is not a complete lattice even though $\mathcal{R} \times \mathcal{S}$ is complete.

Theorem 4.2.1

If $\mathcal{D}$ is f-closed it is complete.

Proof:

For a collection $\{(R_\alpha,S_\alpha)\}$, $\alpha\varepsilon I$ an index set, such that (R_α,S_α) is in $\mathcal{D}$, we claim that $(\bigcup\limits_\alpha R_\alpha,\bigcup\limits_\alpha S_\alpha)$ is the supremum of the collection and is in $\mathcal{D}$. For, $uR_\alpha v$ for each $\mathcal{D}$ implies $f(u)S_\alpha f(v)$ for each α, hence by f-closure $f(u)\bigcup\limits_\alpha S_\alpha f(v)$. Since suprema exist for arbitrary subsets, so do infima.

f-closure is automatically satisfied under some conditions. A typical one is the following: consider the subset $\mathcal{D}_o$ of $\mathcal{D}$ defined by

$$\mathcal{D}_o = \{(R,S) : (R,S)\varepsilon\mathcal{D} \text{ and } S \text{ is a congruence on } Y\}$$

This assumes that Y has at least the structure of a semigroup. The

Theorem 4.2.2

$\mathcal{D}_o$ is f-closed.

$\mathcal{D}_o$ is in fact isomorphic to a sublattice $\mathcal{C}_o$ of C where the output value map $k:Y \to Y_1$ is a homomorphism. Combining the above results we have the statements:

Corollary 4.2.1

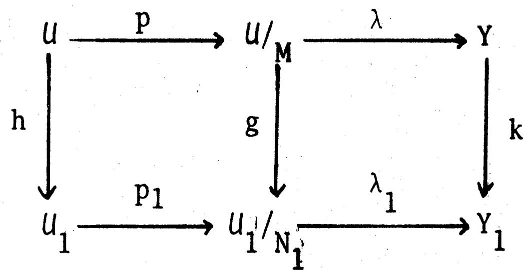
C_o is a complete lattice.

Each element of C_o may be identified with its canonical (state) realization. By corollary 4.1.2 each of these realizations is a homomorphic image of the realization of f, that is, has associated with it a homomorphism (h,g,k). The partial ordering on C_o is directly transferable to that on homomorphisms, so that if C_o' denotes the partially ordered set of homomorphic images of the realization of f, corollary 4.2.1 says that C_o' is a complete lattice.

4.3 Topological Results

Suppose that in definitions 4.1.1 and 4.1.2 every map was required to be continuous. An obvious question to ask is whether the preceding results still hold. This depends on the choice of topologies but if conventions chosen in the earlier chapters are adhered to the answer is yes. The proof of this is essentially an exercise in diagram chasing, and we exhibit just one sample of it, viz., corollary 4.1.2. Refering to the figure below, justified algebraically by the corollary, we have to prove that g is indeed continuous.

Note: Part of the results reported in sections 4.2 and 4.4 were presented in preliminary form [F2] at the 8th Princeton Conference, March 1974.



An open set in $U_1/_{N_1}$ is pulled back via p_1 and h to an open set in U because p_1 and h are continuous. g pulls back this same open set to a set θ in $U/_M$, but because $p^{-1}(\theta)$ is open and $U/_M$ has the quotient topology, θ is open. Hence g is continuous. Thus the corollary is also topologically valid. When λ is continuous it is not necessarily the case that λ_1 is continuous (even though h,k are by assumption). However, it is easily verified that when U_1 has the quotient topology via h, λ_1 must be continuous. This is the case when h is also open. Since by 2.4.4 λ is continuous if and only if f is continuous, the preceding remark gives a condition for the continuity of f_1 when f is continuous.

4.4 Linearity

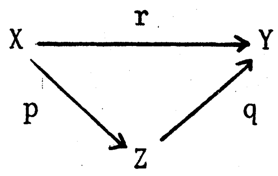
In the previous chapter it was shown that a linear f gives rise to linear transition functions and a linear state space. Suppose now we insist on sioh (h,k) where h,k are also linear. A simple result whose proof is omitted is recorded as the next lemma.

Lemma 4.4.1

f linear and (h,k) linear imply f_1 linear.

More interesting is the fact that in proceeding to a factorization with the above assumptions, $g:U/_M \to U_1/_{N_1}$ turns out to be linear also.

Lemma 4.4.2



If r,p are linear, then q is linear.

Proof:

Suppose $z_1, z_2 \varepsilon Z$, $x_1, x_2 \varepsilon X$ such that $p(x_1) = z_1$, $p(x_2) = z_2$. Then $q(z_1) + q(z_2) = qp(x_1) + qp(x_2) = r(x_1) + r(x_2) = r(x_1 + x_2) = qp(x_1 + x_2) = q(z_1 + z_2)$. Also, if α is any scalar $q(\alpha z_1) = qp(\alpha x_1) = r(\alpha x_1) = \alpha r(x_1) = \alpha qp(x_1) = \alpha q(z_1)$.

Corollary 4.4.1

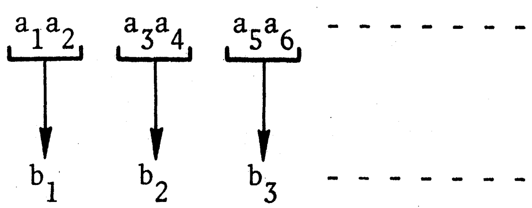
Under the above assumptions $g: U/_M \to U_1/_{N_1}$ is linear.

Proof:

In the commutative diagram for sioh with the linearity assumptions, p, $p_1 \cdot h$ are linear, and g forms the third part of the mapping triangle in the lemma.

Corollary 4.1.2 is therefore valid in the case of continuous linear maps.

An example of linear semigroup homomorphism h is now informally described.



In the representation above a sequence of real numbers is mapped into another sequence such that $b_1 = a_1 + a_2$, $b_2 = a_3 + a_4$, etc. The "batching" of consecutive pairs to singletons gives rise to a semigroup homomorphism.

Within the batch the assignment is a summation, i.e., linear. While this description can be formalized and generalized it is not considered to be terribly enlightening to do so. This technique is typical of a host of dynamic simplification methods used in modeling theory.

We now turn our attention to a particular instance of homomorphism which is called <u>aggregation</u> in the literature. Using Aoki's [A1] formulation we shall examine the case where a linear map $g:Q \to Q_1$ exists between two finite dimensional state spaces of linear machines (A,B,C) and (A_1,B_1,C_1). In particular it is desired to examine the relation between the spectrum of A, denoted $\sigma(A)$, and the spectrum of A_1. The situation is (partially) depicted by:

$$\begin{array}{ccc} Q & \xrightarrow{A} & Q \\ g\downarrow & & \downarrow g \\ Q_1 & \xrightarrow{A_1} & Q_1 \end{array}$$

<u>Theorem 4.4.1 (Aoki)</u>

Let $\lambda \varepsilon \sigma(A)$. If the eigenvectors corresponding to λ are not in the kernel of g then $\lambda \varepsilon \sigma(A_1)$.

<u>Proof</u>:

Suppose $\lambda \varepsilon \sigma(A)$ and q is an eigenvector corresponding to λ, $q \notin \ker g$. Then $Aq = \lambda q$, so $gAq = \lambda gq \neq 0$. Hence $A_1 gq = \lambda gq$, and $\lambda \varepsilon \sigma(A_1)$. Moreover, gq is an eigenvector corresponding to λ.

A partial converse to this result relies on the Jordan reduction for A_1. Call an eigenvalue simple if it occurs in a Jordan block without off-diagonal terms.

<u>Theorem 4.4.2</u>

If $\lambda \varepsilon \sigma(A_1)$ and λ is simple then $\lambda \varepsilon \sigma(A)$.

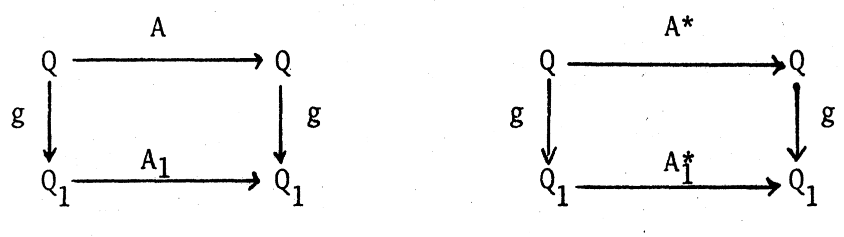
Proof:

From $gA = A_1 g$ we have $A^T g^T = g^T A_1^T$. Without loss of generality assume that A_1 is in Jordan form (otherwise reduce it and modify g accordingly). If A_1 is diagonal the truth of the lemma is obvious, the rows of g being eigenvectors of A. If A_1 is not diagonal a similar argument yields the result for simple λ.

The use of this theorem, besides being a partial converse inference, is to delimit the types of base systems which have known lumped systems, provided they all reside within the class of linear finite dimensional systems, and have linear homomorphisms as relations. Suppose linear systems $A_1, A_2, \ldots, A_n$ are known to be linear homomorphic images of an unknown A (we identify systems with their state transition matrices). Then by the lemma above we can give a lower bound on the dimension of A by counting up the number of simple eigenvalues in $A_1, \ldots, A_n$, disregarding repetitions.

We conclude this section by examining the property of A being normal. Recall that A is normal if it commutes with its adjoint A*, i.e., AA* = A*A. Self-adjointness is a special case of normality. The result in the finite dimensional case is given as:

Theorem 4.4.3



Let Q, Q_1 be finite dimensional unitary spaces, A is normal and linear, g is linear, and assume that left commutative diagram holds (i.e. g is a homomorphism from A to A_1). Then A_1 is normal if and only if the right commutative diagram holds.

Proof:

(<=) Suppose the diagram commutes. Then $gA = A_1 g$ and $gA^* = A_1^* g$. It suffices to show that

$$\langle A_1 A_1^* g_1, q_1 \rangle = \langle A_1^* A_1 q_1, q_1 \rangle \qquad \text{for all } q_1 \varepsilon Q_1.$$

$$\begin{aligned} \text{The LHS} &= \langle A_1 A_1^* gq, q_1 \rangle && \text{where } gq = q_1 \\ &= \langle A_1 gA^* q, q_1 \rangle \\ &= \langle gAA^* q, q_1 \rangle \\ &= \langle gA^* Aq, q_1 \rangle && \text{by normality of A} \\ &= \langle A_1^* gAq, q_1 \rangle \\ &= \langle A_1^* A_1 gq, q_1 \rangle \\ &= \langle A_1^* A_1 q_1, q_1 \rangle \end{aligned}$$

(=>) Suppose A_1 is normal. Use the spectral theorem [T1] to obtain the spectral decompositions of A_1 and A, i.e.

$$A_1 = \sum_{i=1}^{m} \mu_i P_i \qquad A = \sum_{i=1}^{n} \lambda_i R_i$$

From $gA = A_1 g$ we obtain $gf(A) = f(A_1)g$ for analytic f. Choosing $f(v) = \bar{v}$ we get $f(A) = A^*$ and $f(A_1) = A_1^*$, hence $gA^* = A_1^* g$.

The extension of the above result to the infinite dimensional case will have to invoke compactness.

4.5 Input-output Homomorphisms

The definition of a sioh (h,k) is predicated upon the assumption of causality and preservation of concatenation. The causality assumption is implicit in the definition of a system being a map from a segment space $\mathcal{U}$ to a value space Y. In many modeling situations causality is not necessary, and neither is the semigroup property. Thus we may relax definition 4.1.1 to that of an input-output homomorphism (i-o homomorphism) which requires the following framework. A system is an operator $K:X \to Y$ where X and Y are spaces of functions defined on a time set (e.g., R, Z, an ordered group, etc.), and given two systems K_1, K_2, an i-o homomorphism is a pair (h,k) such that $hK_1 = K_2k$. Often we will identify X and Y for convenience. The rest of the development will be concentrating on i-o homomorphisms of this type. We hasten to add that the scheme discussed in the previous section, although it is a state homomorphism, falls within this framework, merely by letting K be identified with A, and X,Y, be identified with Q. Hence, whenever a result is stated for i-o homomorphisms it is understood that a corresponding result holds for state homomorphisms.

In this section we shall consider various typical system theoretic properties which are preserved under i-o homomorphisms. Stability and time-variance have been postponed to subsequent chapters for more detailed investigation. For the rest of this section by homomorphism we shall mean i-o homomorphism.

A system $K:X \to X$ has a <u>fixed point</u> x_o if $Kx_o = x_o$. The existence of fixed points is an important property in both control and computation.

<u>Theorem 4.5.1</u>

If h is a homomorphism from $K:X \to X$ to $K_1:X_1 \to X_1$, then if x_o is a fixed point of K, hx_o is a fixed point of K_1.

<u>Proof</u>:

$$Kx_o = x_o \Rightarrow hKx_o = hx_o$$
$$\Rightarrow K_1hx_o = hx_o$$

The converse is generally untrue.

Given a system $K:X \to Y$, we say it has a (function-theoretic) property P if K has that property as a function space map. The following results are elementary and their proofs are omitted. In each case $h:X \to X_1$ is the homomorphism from K to K_1, input and output spaces being identified, the extension to an output map $k:Y \to Y_1$ being easy.

Theorem 4.5.2

If K is linear and h is linear, then K_1 is linear.

Theorem 4.5.3

If K is continuous (open) and H is continuous and open, K_1 is continuous (open).

This condition on h imposes the quotient topology on X_1.

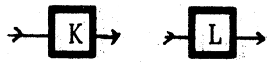
Corollary 4.5.1

If K is continuous (open), X and X_1 are complete, and h is continuous and linear, then K_1 is continuous (open).

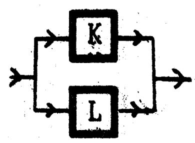
This follows from the open mapping theorem [T1].

In closing we consider a problem in the interconnection of systems. In doing so we shall anticipate some of the development in the next chapter. The problem consists in the question: if local linear i-o homomorphisms are known to exist between two isomorphic networks of systems, is this sufficient to assert that a global linear i-o homomorphism also exists? We shall make this more precise later. Initially the primitive composition operators are defined, and it is shown that they preserve linear homomorphisms. For simplicity we assume that the systems in a given network are defined over a single input-output space. The relaxation of this condition is not difficult except that consistency conditions on interconnection must be met.

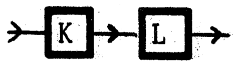
The three primitive composition operators are called series, parallel, and feedback compositions, abbreviated respectively S,P and FB. All three are binary. Their mode of action is best described by the schematic diagrams:



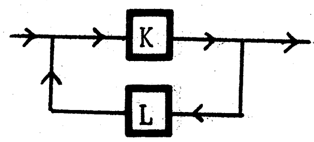
(K,L)



P(K,L)



S(K,L)



FB(K,L)

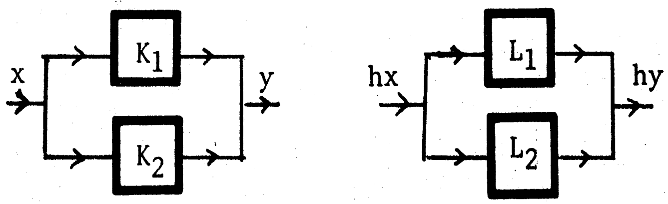
Whenever two arrows meet a summation is intended, and whenever they separate they denote quantities identical to the sum which is effective for that node.

An interconnection of systems is obtained by a finite number of applications of these operators. A precise statement of our problem can now be made. Let G_1, G_2 be two directed, labelled, possibly multiple edged graphs which are isomorphic. Assume that every edge in G_1 has a system and the corresponding edge in G_2 has a system linearly homomorphic to it. Then we claim that every choice of global input and output nodes will result in a linear homomorphism of the overall interconnected systems. (It is evident that the type of graphs described capture essentially the interconnections permitted.) The demonstration of this proceeds via three lemmas. A notational convenience is gained by letting $K \xrightarrow{h} L$ mean a homomorphism exists from K to L.

Lemma 4.5.1

If $K_1 \xrightarrow{h} L_1$ and $K_2 \xrightarrow{h} L_2$, then $P(K_1,K_2) \xrightarrow{h} P(L_1,L_2)$.

Proof:



It is required to show that

$$hy = (L_1 + L_2)hx$$

Since $y = (K_1 + K_2)x$

$$\begin{aligned} hy &= h(K_1 + K_2)x \\ &= (hK_1 + hK_2)x \\ &= (L_1h + L_2h)x \\ &= (L_1 + L_2)hx \end{aligned}$$

Lemma 4.5.2

If $K_1 \xrightarrow{h} L_1$ and $K_2 \xrightarrow{h} L_2$, then $S(K_1,K_2) \xrightarrow{h} S(L_1,L_2)$.

Proof:

It is required to show that

$$hy = L_2L_1hx$$

Since $y = K_2K_1x$

$$\begin{aligned} hy &= hK_2K_1x \\ &= L_2hK_1x \\ &= L_2L_1hx \end{aligned}$$

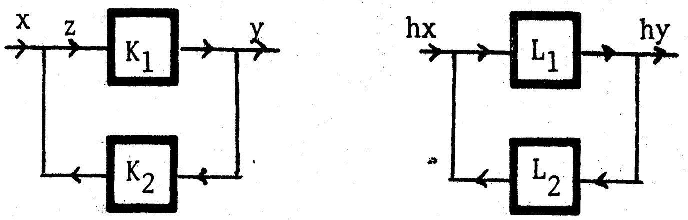
Observe that linearity of h is not used in this case.

Before proceeding to the next lemma we have to discuss a point which is reiterated in the next chapter. In a feedback composition of K_1,K_2, for the composed system to be well defined, $(I-K_2K_1)$ must be invertible.

Lemma 4.5.3

If $K_1 \xrightarrow{h} L_1$ and $K_2 \xrightarrow{h} L_2$, then $FB(K_1,K_2) \xrightarrow{h} FB(L_1,L_2)$.

Proof:



It is required to show that

$$hy = L_1(I-L_2L_1)^{-1}hx \ .$$

Because $x = (I-K_2K_1)z$

$$hx = h(I-K_2K_1)z$$
$$= (I-L_2L_1)hz$$

so $hz = (I-L_2L_1)^{-1}hx$

since $y = K_1z$

$$hy = hK_1z$$
$$= L_1hz$$
$$= L_1(I-L_2L_1)^{-1}hx.$$

Using these three lemmas we complete the proof by induction on the number of operators. So assume the conclusion is true for all interconnections generated by less than or equal to n operators. (The case for one operator is in lemmas 4.5.1, 4.5.2, and 4.5.3.) Observe that a network generated by n (or less) operators is itself a system, with i-o nodes where a new connection is to be made to another system. Applying lemmas 4.5.1, 4.5.2 and 4.5.3 to these two systems completes the induction.

Topological consequences of this result may be obtained by appealing to corollary 4.5.1.

4.6 Discussion

The results on the lattice of homomorphisms is interpretable as statements about the hierarchy of models of a given system, which in turn may reside in some other hierarchy. The connection to decomposable systems (e.g. Ando, et al. [A2]) is not immediate. An element of our lattice need not be a component in a decomposition. A more correct analogy is that the lattice represents hierarchical descriptions in terms of relative micro- and macro-structure. However, in special cases, for instance in considering the homomorphic images of linear systems whose state transition matrices (possibly infinite dimensional) are near diagonal, some elements of the lattice may in fact be component subsystems.

We will now restate some of the results of this chapter from the perspective of system predicate validity of homomorphisms.

Corollary 4.1.2 has the form of a justifying condition as explained in chapter 0. Stated in predicate terms it is

$$\mathrm{REDUCED}(S') \wedge \mathrm{I\text{-}O\_HOM}(S,S') \supset \mathrm{STATE\_HOM}(S,S')$$

or $$\mathrm{TOP\_REDUCED}(S') \wedge \mathrm{CONT\_I\text{-}O\_HOM}(S,S') \supset \mathrm{CONT\_STATE\_HOM}(S,S')$$

or $$\mathrm{LIN\_REDUCED}(S') \wedge \mathrm{LIN\_I\text{-}O\_HOM}(S,S') \supset \mathrm{LIN\_STATE\_HOM}(S,S')$$

or $$\mathrm{TOP\_LIN\_REDUCED}(S') \wedge \mathrm{CONT\_LIN\_I\text{-}O\_HOM}(S,S') \supset \mathrm{CON\_LIN\_STATE\_HOM}(S,S')$$

where TOP, CONT, LIN, HOM stand respectively for topological, continuous, linear, and homomorphism, and the predicate names speak for themselves. The various forms of corollary 4.1.2 are different instances of the same scheme for i-o homomorphisms and (state) homomorphisms.

In section 4.4 validity of sioh was established with respect to the LIN predicate where (h,k) were linear;

$$\mathrm{LIN\_I\text{-}O\_HOM}(S,S') \supset (\mathrm{LIN}(S) \supset \mathrm{LIN}(S'))$$

Also shown was

$$\mathrm{CONT\_OPEN\_HOM}(S,S') \supset (\mathrm{CONT}(S) \supset \mathrm{CONT}(S'))$$

Proceeding to the examination of spectra for state transition matrices of finite dimensional systems we stated Aoki's lemma and proved lemma 4.4.3. Although these may be stated in predicate form the restrictive conditions make it cumbersome to do so.

In section 4.5 we looked at (not necessarily causal) i-o homomorphisms. Let the FIXPOINT (K,x) be a formula free in K and x. where the notation is as explained in section 0.3. If FIXPOINT (K,x_o) is satisfied we mean that x_o is a fixed point of K. It was shown that

$$\mathrm{I\text{-}0\_HOM}(S,S') \supset (\mathrm{FIXPOINT}(S,x_o) \supset \mathrm{FIXPOINT}(S',hx_o))$$

where h is the i-o map.

Also shown was

$$\mathrm{LIN\_CONT\_I\text{-}0\_HOM}(S,S') \wedge \mathrm{COMPLETE}(X) \supset \mathrm{COMPLETE}(X') \supset (\mathrm{CONT}(S) \supset \mathrm{CONT}(S')) \ .$$

The next result concerns an extension theorem for local homomorphisms. We state for one of the lemmas what form the predicate validity takes and indicate the form of the overall statement. For the parallel operation we have

$$\mathrm{I\text{-}0\_HOM}(S_1,S_1') \wedge \mathrm{I\text{-}0\_HOM}(S_2,S_2') \supset (\mathrm{PAR}(S_1,S_2) \supset \mathrm{PAR}(S_1',S_2'))$$

and also

$$\mathrm{I\text{-}0\_HOM}(S_1,S_1') \wedge \mathrm{I\text{-}0\_HOM}(S_2,S_2') \supset \mathrm{I\text{-}0\_HOM}(\mathrm{PAR}(S_1,S_2),\ \mathrm{PAR}(S_1',S_2'))$$

where PAR is the binary predicate indicating parallel composition. The first assertion is embedded in the proof of lemma 4.5.1 and the second is the homomorphism extension theorm for parallel composition. The general statement for extensions can be made recursively, having as basis two other assertions like the second one above, one for the SER (for series) and one for the FB (for feedback) predicate.

Chapter 5

In this chapter we take up the issue of validity of homomorphisms with respect to stability predicates. There are numerous definitions of system stability but we shall consider only a few of them. We omit with some reluctance the treatment of structural stability, hoping to investigate it in future research. To begin with we shall consider autonomous state systems and Lyapunov and asymptotic stability. We conclude with an examination of feedback stability. In considering the latter it turns out that several other interesting predicates are relevant.

5.1 Lyapunov Stability

Lyapunov stability was defined originally for systems governed by differential equations. The development of dynamical systems theory has allowed the definition to be considerably broadened. Here we take Lyapunov stability to presume the existence of a system with state space Q which is a normed linear space, and a state transition function $\delta(t;t_o):Q \to Q$ where t,t_o are the initial and final times. Observe that the system is autonomous.

It is presumed that the system under consideration has the property that it has a zero solution, that is, q = 0 for all t is a valid trajectory of the system. In discussions of Lyapunov stability it is the stability of this trajectory that is investigated. The question asked is whether the system is sensitive to small perturbations of this trajectory.

Definition 5.1.1

(The zero "solution" of) a system is Lyapunov stable (LS) if for each $t_o > 0$ and each $\varepsilon>0$ there exists a γ such that

$$||q_o|| < \gamma \Rightarrow ||\delta(t,t_o)q_o|| < \varepsilon$$

for all $t > t_o$.

Theorem 5.1.1

If $h: Q \to Q_1$ is a homomorphism from system δ to system δ_1 such that $h(0) = 0$, and h is continuous and open, then δ is LS implies δ_1 is LS.

Proof:

Let $B(\varepsilon)$ denote the sphere $\{q \in Q: ||q|| < \varepsilon\}$.

Let $\varepsilon > 0$ be given. If we can show there exists $\gamma > 0$ such that for each t_o

$$\delta_1(t,t_o)B_1(\gamma) \subset B_1(\varepsilon)$$

for all $t > t_o$, the proof is complete.

Because $h(0) = 0$ and h is continuous, given $B_1(\varepsilon)$ there exists a sphere $B(\varepsilon')$ such that

$$B(\varepsilon') \subset h^{-1}B_1(\varepsilon).$$

δ is LS now implies there is $\gamma' > 0$ such that

$$\delta(t,t_o)\ B(\gamma') \subset B(\varepsilon') \qquad \text{for all } t > t_o$$

i.e.
$$\begin{aligned} B(\gamma') &\subset \delta^{-1}(t,t_o)\ B(\varepsilon') \\ &\subset \delta^{-1}(t,t_o)h^{-1}B_1(\varepsilon) \\ &= h^{-1}\delta_1^{-1}(t,t_o)\ B_1(\varepsilon) \end{aligned}$$

so
$$hB(\gamma') \subset \delta_1^{-1}(t,t_o)\ B_1(\varepsilon) \quad .$$

But h is open, $h(0) = 0$, so there exists a sphere $B_1(\gamma)$ such that

$$\begin{aligned} B_1(\gamma) &\subset hB(\gamma') \\ &\subset \delta_1^{-1}(t,t_o)B_1(\varepsilon) \end{aligned}$$

i.e.
$$\delta_1(t,t_o)B_1(\gamma) \subset B_1(\varepsilon) \qquad \text{for all } t > t_o \ .$$

Corollary 5.1.1

If $h: Q \to Q_1$ is a continuous linear homomorphism from δ to δ_1 such that Q and Q_1 are complete then δ is LS implies δ_1 is LS.

Proof:

By the open mapping theorem [T1] h is also open in this instance.

5.2 Asymptotic Stability

A common feature of many autonomous systems is the long-term decay of the state to the zero point. If this is accompanied by Lyapunov stability we have asymptotic stability.

Definition 5.2.1

A system is asymptotically stable (AS) if it is LS and for all $q \in Q$, $t_o \in R$

$$\lim_{t\to\infty} \delta(t,t_o)q = 0$$

Theorem 5.2.1

If $h:Q \to Q_1$ is a homomorphism from system δ to system δ_1 such that $h(0) = 0$, and h is continuous and open then δ is AS implies δ_1 is AS.

Proof:

δ is AS $\Rightarrow$ δ is LS

$\Rightarrow$ δ_1 is LS by theorem 5.1.1

Hence all that is required is to show that for all q_1 and t_o

$$\lim_{t\to\infty} \delta_1(t,t_o)q_1 = 0 \quad .$$

$$\delta_1(t,t_o)q_1 = \delta_1(t,t_o)hq \qquad \text{for some } q \in Q$$

$$= h\delta(t,t_o)q$$

so

$$\lim_{t\to\infty} \delta_1(t,t_o)q_1 = \lim_{t\to\infty} h\delta(t,t_o)q$$

$$= h \lim_{t\to\infty} \delta(t,t_o)q \qquad \text{since h is continuous}$$

$$= h(0)$$

$$= 0.$$

Corollary 5.2.1

If $h:Q \to Q_1$ is a continuous linear homomorphism from δ to δ_1 such that Q and Q_1 are complete then δ is AS implies δ_1 in AS.

Proof:

As in corollary 5.1.1.

5.3 Bounded-input bounded-output stability

Bounded sets about the origin in the normed linear space X or X_1 play an important role in some types of stability as well as in the problem of converse validity. In keeping with the earlier notation a ball of radius M about the origin is denoted by B(M) if it is X and by $B_1(M)$ if it is in X_1. The use of the symbols X and X_1 is deliberate because we will use this discussion to examine bounded-input bounded-output stability, when X will stand for the input (as well as the output) space as in the previous chapter. If $h:X \to X_1$ is such that $h(0) = 0$, by the boundedness of h we mean

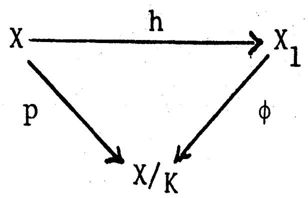
$$||h|| = \sup_{x\varepsilon X} \frac{||hx||}{||x||} < \infty \quad .$$

The point to set function h^{-1} implicitly defined is the proof of theorem 5.1.1 has two obvious senses of boundedness. If h^{-1} takes bounded sets to bounded sets we say it is (strongly) bounded. If h^{-1} applied to some bounded set U_1 yields a set which contains a bounded subset U such that $h(U) = U_1$ then we say that it is weakly bounded. Weak boundedness is relatively easy to satisfy. An instance of this is given in the next lemma.

Lemma 5.3.1

If $h:X \to X_1$ is a continuous linear surjection and X,X_1 are Banach spaces then h^{-1} is weakly bounded.

Proof:



Refering to the commutative diagram, the following assertions are consequences of standard results in functional analysis [T1]. K, the kernel of h, is closed, and hence X/K is a normed linear space, in fact a Banach space. ϕ is an isomorphism, and is continuous because of the open mapping theorem. Given a bounded set U_1 in X_1 with a bound M, $\phi(U_1)$ is bounded by $M||\phi||$. If $x+K \varepsilon \phi(U_1)$ by the definition of the quotient norm in X/K, there exists $k_x \varepsilon K$ such that $||x+k_x|| < M||\phi||+\varepsilon$ and ε is some fixed positive number. Now, consider the set $U = \{x+k_x: x \varepsilon h^{-1}(U_1)\}$. If $y \varepsilon U$, $y = x+k_x$ for some $x \varepsilon h^{-1}(U_1)$, so $h(y) = h(x)$ hence $h(U) \subset U_1$. Conversely, if $x_1 \varepsilon U_1$ there is $x+k_x \varepsilon U$ such that $h(x+k_x) = h(x) = x_1$, so $U_1 \subset h(U)$. Thus $h(U) = U_1$. By construction U is bounded by $M||\phi||+\varepsilon$.

Definition 5.3.1

If $K:X \rightarrow X$ is an i-o system K is bounded-input bounded-output (BIBO) stable if bounded input sets are mapped into bounded output sets.

Theorem 5.3.1

If $h:X \rightarrow X_1$ is an i-o homomorphism from system K to system K_1, h is bounded and h^{-1} is weakly bounded, then K is BIBO stable imples K_1 is BIBO stable.

Proof:

Let U_1 be a bounded set in X_1. h^{-1} is weakly bounded implies there exists a bounded set U in X such that $h(U) = U_1$. Hence $K_1U = K_1h(U) = hK(U)$. But by BIBO stability of K, K(U) is bounded, thus by boundedness of h, hK(U) is bounded, hence K_1U is bounded.

Corollary 5.3.1

If $h:X \to X_1$ is an i-o homomorphism from system K to system K_1, X and X_1 are complete, and h is linear and continuous, then K is BIBO stable imples K_1 is BIBO stable.

Proof:

Immediate from lemma 5.3.1 and theorem 5.3.1.

Other forms of bounded stability may be treated in analogous fashion. (Strong) boundedness is not as easy to achieve. In fact if h is linear h^{-1} cannot be bounded because ker h is not bounded. If boundedness for h^{-1} holds the converse inference for BIBO stability can be made. One naive example of when h^{-1} is bounded is the following: let $h:R \times R \to R$ be given by

$$h(x,y) = \begin{cases} x^2 + y & \text{if } y > 0 \\ x^2 - y & \text{if } y < 0 \end{cases}$$

Theorem 5.3.2

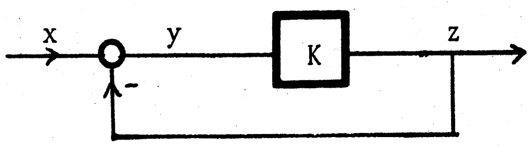
If $h:X \to X_1$ is an i-0 homomorphism from system K to system K_1 such that h is bounded and h^{-1} is bounded then K is BIBO stable if and only if K_1 is BIBO stable.

Proof:

Similar to the proof of theorem 5.3.1.

5.4 Feedback stability

Another aspect of stability theory concerns the stability of non-linear feedback systems. The literature in this area is extensive (See Damborg and Nayor [D2] for a bibliography) but our attention will be directed towards the problem of homomorphic models of such systems. The point of view we shall take in regard to feedback systems is that outlined in reference [D2], primarily because normed linear spaces are much easier to deal with than the extended linear spaces of some other references. A brief discussion of the simple nonlinear feedback system is included, but for motivation the references should be consulted.



A schematic representation of the feedback system is shown above where $K:H \rightarrow H$ is a nonlinear operator and an appropriate (normed) linear space H is chosen to represent inputs and outputs. For reasons of simplicity, and because it is easy to recover z from y (z=Ky), we shall look at the interaction between x and y, viz., $(I + K)y = x$.

Here we make a slight change in convention from that in chapter 4. In deference to the vast literature in feedback systems the feedback loop is considered to be negative in sign. This change is actually quite immaterial but convenient for purposes of cross referencing.

<u>Definition 5.4.1 [D2]</u>

The feedback system of the Figure is <u>feedback stable</u> (FB) if (i) I + K is injective, (ii) I + K is surjective, (iii) $(I + K)^{-1}$ is causal, continous and maps bounded sets into bounded sets.

As noted in the reference, this definition divides feedback stability into five sub-issues, each of which may be examined separately (once $(I + K)^{-1}$ exists).

In the modelling situation the simplification process is normally an attempt to capture the essential features of K. Often, the internal structure of K may give rise to a state description which may be simplified to yield a state homomorphism. If this is too difficult recourse may be had to input-output homomorphisms of the type encountered earlier. The general i-o homomorphism is not adequate in the sense that in the modelling of a feedback system as in the figure, the simplification process is clearly one of capturing the essential features of K, using a $K_1 : X_1 \to X_1$ in a new normed linear space. What this amounts to is that we would insist upon a structural preservation of the feedback loop as well. The preservation of structure was formalised as an insistence on the isomorphism of labelled directed graphs in the previous chapter. In the interest of readability some of that development is briefly repeated here.

Suppose we regard the i-o homomorphism as going from the subsystem K to the subsystem K_1. Taken in isolation, this says $K_1 h = hK$, where $h : X \to X_1$. Now, we claim that this gives rise to an overall i-o homomorphism from the first feedback system (S) to the second (S_1), given that (I+K) and ($I+K_1$) are both invertible. The overall system operator for S is $K(I+K)^{-1}$ and for S1 it is $K_1(I+K_1)^{-1}$. To say h is an overall i-o homomorphism is to say that $hK(I+K)^{-1} = K_1(I+K_1)^{-1}h$. This is a consequence of $hK = K_1 h$ because of the following chain of implications:

$$
\begin{aligned}
K_1 h = hK \quad &\Rightarrow \quad h+K_1 h = h+hK \\
&\Rightarrow \quad h(I+K)^{-1} = (I+K)^{-1}h \\
&\Rightarrow \quad K_1 h(I+K)^{-1} = K_1(I+K_1)^{-1}h
\end{aligned}
$$

whence, by $K_1 h = hK$ again, we have

$$hK(I+K)^{-1} = K_1(I+K_1)^{-1}h.$$

The point to observe is that once it can be established that $hy=y_1$ at the inputs to K and K_1, the overall i-o homomorphism is assured. The above formal derivation may obscure this fact, so we examine it more closely.

Given x, x_1 such that $hx=x_1$ at the overall inputs, it is clear that

$$hy = h(I+K)^{-1}x$$

and $\quad y_1 = (I+K)^{-1}x_1 = (I+K_1)^{-1}hx$

Thus $\quad hy = y_1$ if $h(I+K)^{-1} = (I+K_1)^{-1}h$

i.e. $\quad hK = K_1h.$

The converse problem of beginning with an i-o homomorphism from S to S_1 and deducing conditions under which this induces an i-o homomorphism from K to K_1 is difficult in the general case. When h is linear, however, the problem is almost trivial, and no conditions on the invertibility of (I+K) or $(I+K_1)$ need be imposed.
It is simply observed that

$$\begin{aligned} hy &= h(x-z) \\ &= hx-hz \\ &= x_1-z_1 \\ &= y_1 \end{aligned}$$

So

$$\begin{aligned} hKy &= hz \\ &= z_1 \\ &= K_1y_1 \\ &= K_1hy \end{aligned}$$

and we have $hK = K_1h$.

Invertibility assumptions have been invoked rather freely. Because it is desired to investigate the conditions under which stability is preserved under homomorphisms, it is a hypothesis that S is stable, so by assumption I+K is invertible.

Lemma 5.4.1

If h is an i-o homomorphism from K to K_1 then $I+K_1$ is surjective if I+K is surjective.

Proof:

Recall that h is a surjection by definition. Then $K_1h = hK \Rightarrow (I+K_1)h = h(I+K)$, so $(I+K_1)h$ is surjective, hence $I+K_1$ is surjective.

This is one half of the invertibility condition. The other is injectivity. It is therefore relevant to see when injectivity is preserved under homomorphisms.

Lemma 5.4.2

Let h be an i-0 homomorphism from $F:X \to X$ to $F_1:X_1 \to X_1$, and F is invertible, then $hx = hy \Rightarrow hF^{-1}(x) = hF^{-1}(y)$ is a necessary and sufficient condition for F_1 to be invertible.

Proof:

Since F is surjective, so is F_1. It suffices to show that the condition is equivalent to injectivity of F_1.

Sufficiency: Suppose $hx \neq hy$, but $F_1hx=F_1hy$. Since $F_1h=hF$, $hFx=hFy$. Let the hypothesis of the condition be $hFx=hFy$, then because F^{-1} exists the conclusion says $hx=hF^{-1}F_x=hF^{-1}Fy=hy$, a contradiction.

Necessity: Let $hx=hy$. Because F^{-1} exists let $x'=F^{-1}x$ and $y'=F^{-1}y$. It is clear that $F_1hx'=F_1hy'$ since this is equivalent to $hFx'=hFy'$, which is simply $hx-hy$. But F_1 is injective by assumption, so $hx'=hy'$, that is $hF^{-1}x=hF^{-1}y$.

Corollary 5.4.1

If h is an i-o homomorphism from K to K_1, I+K is invertible, then a necessary and sufficient condition for $I+K_1$ to be invertible is

$$hx=hy \Rightarrow h(I+K)^{-1}x = h(I+K)^{-1}y.$$

Proof:

h is an i-o homomorphism from I+K to $I+K_1$.

Another condition is the following for Hilbert spaces.

Lemma 5.4.3

Suppose X is a Hilbert space. If h is an I-O homomorphism from K to K_1 then a sufficient condition for $I+K_1$ to be invertible is

$$|\langle hKx-hKy,\ hx-hy\rangle| \leqslant ||hKx-hKy||\ \ ||hx-hy|| \text{ for all } x,\ y\varepsilon X.$$

Proof:

It is easy to show [D2] that $I+K_1$ is injective if and only if

$$||K_1x_1-K_1y_1||^2 + 2\langle K_1x_1-K_1y_1,\ x_1-y_1\rangle + ||x_1-y_1||^2 \geqslant 0 \quad \text{for all } x_1,y_1\varepsilon H_1$$

i.e. $||hKx-hKy||^2 + 2\langle hKx-hKy,\ hx-hy\rangle + ||hx-hy||^2 \geqslant 0$ for all $x,y\varepsilon H$

$$0 \leqslant (||hKx-hKy|| - ||hx-hy||)^2$$

$$= ||hKx-hKy||^2 - 2||hKx-hKy||\ ||hx-hy|| + ||hx-hy||^2$$

$$\leqslant ||hKx-hKy||^2 + 2\langle hKx-hKy,\ hx-hy\rangle + ||hx-hy||^2$$

if the condition is satisfied.

This condition is interpretable geometrically as a prohibition of hKx-hKy and hx-hy being parallel, and does not depend on I+K being invertible.

When all of h,K,K_1 are linear, and K,K_1 satisfy conditions appropriate to the spectral theorems ([N1], [T1]) then using the relation $hK = K_1h$ to show $hp(K) = p(K_1)h$ for all polynomials p, it is possible to deduce invertibility of $I+K_1$ from that of I+K.

The next property to consider is causality. Strictly, the discussion should take place within a framework similar to the Resolution Spaces of Porter and Saeks [P2], [S2]. It suffices for our purposes to consider the projection operator E^t which acts as follows:

$$(E^t x)(s) = \begin{cases} x(s) & \text{if } s \leq t \\ 0 & \text{if } x > t \end{cases}$$

Then recall that $F:X \to X$ is <u>causal</u> if $E^t x = E^t y \Rightarrow E^t Fx = E^t Fy$ for $x, y \varepsilon X$. By assumption $(I+K)^{-1}$ is causal, and it is desired to find conditions for the causality of $(I+K)^{-1}$, given K_1 is a homomorphic image of K.

Lemma 5.4.4

If h is an i-o homomorphism from $F:X \to X$ to $F_1:X_1 \to X_1$, F is causal, h is causal, and for every x_1, $y_1 \varepsilon X_1$

$$E^t x_1 = E^t y_1 \Rightarrow \exists x, y \varepsilon X,\ hx = x_1,\ hy = y_1 \text{ and } E^t x = E^t y$$

then F_1 is causal.

Proof:

$$\begin{aligned} \text{If so, } E^t Fx &= E^t Fy && \text{as F is causal} \\ E^t hFx &= E^t hFy && \text{as h is causal} \\ E^t F_1 hx &= E^t F_1 hy \\ E^t F_1 x_1 &= E^t F_1 y_1 \end{aligned}$$

When both F and F_1 are linear, a simple criterion for causality is obtained as follows:

Let X_T be the set of functions with support contained in (T,∞). Then linear $K:X \to X$ is causal if and only if $KX_T \subseteq X_T$ for all T. See [N1].

Lemma 5.4.5

If h is an i-o homomorphism from $F:X \to X$ to $F_1:X_1 \to X_1$, F and F_1 are linear, and F is causal, a sufficient condition for F_1 to be causal is that $hX_T = X_{1_T}$.

Proof:

$$\begin{aligned} L \text{ causal and linear } &\Rightarrow LX_T \subseteq X_T \\ &\Rightarrow hLX_T \subseteq hX_T \\ &\Rightarrow KhX_T \subseteq hX_T \\ &\Rightarrow KX_{1_T} \subseteq X_{1_T} \\ &\Rightarrow K \text{ is causal (since it is linear).} \end{aligned}$$

Lemma 5.4.6

Under the same assumptions as in the previous lemma, if $h^{-1}X_{1_T} = X_T$, then F is causal if and only if F_1 is causal.

Proof:

($\Rightarrow$) Follows from the previous lemma by $h^{-1}X_{1_T} = X_T \Rightarrow hX_T = X_{1_T}$

($\Leftarrow$) Suppose F is not causal. Then there exists $x \varepsilon X_T$, $Fx \notin X_T$. Then $hx \varepsilon X_{1_T}$, so $F_1 hx \varepsilon X_{1_T}$ since F_1 is causal. Hence $hFx \varepsilon X_{1_T}$. This says $h^{-1}X_{1_T} \neq X_T$, a contradiction.

When h is in addition time-invariant, the previous conditions reduce to $hX_0 = X_{1_0}$ and $h^{-1}X_{1_0} = X_0$. An example of when $hX_0 = X_{1_0}$ is satisfied is when h is multiplicative, i.e., $(hx)(t) = g(t)\,x(t)$ where $g(t): V \to V$ in the case where H consists of functions from R to V. It is easy to see that h multiplicative implies h memoryless, but not vice-versa.

Continuity and boundedness predicates were discussed in sections 4.5 and 5.3 respectively. More specifically in section 4.5 we showed that if $h: X \to X_1$ is both continuous and open then $F: X \to X$ is continuous, and that if X and X_1 are complete and h is linear, the contintuity alone of h is sufficient for this to be true. In section 5.3 it was shown that if h is bounded and h^{-1} is weakly bounded then F is bounded implies

F_1 is bounded, although the terminology was somewhat different in that context.

In our current discussion F is to be identified with $(I+K)^{-1}$ and F_1 with $(I+K_1)^{-1}$.

We are now in a position to collect together the information in the preceding lemmas to state theorems under which FB stability is preserved through i-o homomorphisms. Because of the several conditions on each subissue these theorems will be tedious to enumerate, so we have arbitrarily selected one possibility to give the flavor of the results. Other statements may be easily culled from the lemmas.

Theorem 5.4.1

If $h:X \rightarrow X_1$ is linear and satisfies the following conditions

(i) $hx=hy \Rightarrow h(I+K)^{-1}x = h(I+K)^{-1}y$, (ii) $E^t x_1 = E^t y_1 \Rightarrow \exists x,y \varepsilon X$, $hx=x_1$, $hy=y_1$ and $E^t x = E^t y$ (iii) h is continuous and open, (iv) h^{-1} is weakly bounded; then S is FB stable implies S1 is FB stable.

The point to note about the FB predicate is that rather strong conditions on homomorphisms have to be invoked in order that it be preserved. This is not altogether unexpected because the introduction of feedback can radically alter the behavior of a system, and further, feedback stability is critically dependant on the nature of the feedback. This is an instance when simple homomorphisms do not appear to suffice.

5.5 Discussion

We summarize the results on validity of homomorphic simplification obtained in this chapter. The abbreviations used are self-explanatory. In section 5.1 it was shown that

$$\text{CONT\_OPEN\_I-O\_HOM}(S,S') \supset (LS(S) \supset LS(S'))$$

and

$$\text{CONT\_LIN\_I-O\_HOM}(S,S') \wedge \text{COMPLETE}(Q) \wedge \text{COMPLETE}(Q') \supset (\text{LS}(S) \supset \text{LS}(S'))$$

In 5.2 it was shown that

$$\text{CONT\_OPEN\_I-O\_HOM}(S,S') \supset (\text{AS}(S) \supset \text{AS}(S'))$$

and

$$\text{CONT\_LIN\_I-O\_HOM}(S,S') \wedge \text{COMPLETE}(Q) \wedge \text{COMPLETE}(Q') \supset (\text{AS}(S) \supset \text{AS}(S')) .$$

Next, we proved that

$$\text{BDD\_h\_WEAK-BDD\_h}^{-1}\text{\_HOM}(S,S') \supset (\text{BIBO}(S) \supset \text{BIBO}(S'))$$

$$\text{CONT\_LIN\_I-O\_HOM}(S,S') \wedge \text{COMPLETE}(X) \wedge \text{COMPLETE}(X') \supset (\text{BIBO}(S) \supset \text{BIBO}(S')$$

and

$$\text{BDD\_h\_BDD\_h}^{-1}\text{\_HOM}(S,S') \supset (\text{BIBO}(S) \leftrightarrow \text{BIBO}(S'))$$

The final results given in section 5.4 can be cast into predicate form, one for each lemma. The predicates are however too tedious to write in complete form. In particular the conditions in the final theorem are not particularly pleasant to restate.

Chapter 6

In systems theory the concept of slowly time varying systems has led to the modelling of such systems by their time-invariant counterparts. The question of whether there are meaningful measures of degrees of time-variance in systems is, in this context, both interesting and important. Such a measure is proposed and its consequences examined in this chapter.

We focus on the viewpoint that a system is an operator from a function space to another, the function spaces consisting of functions from a time set to a linear space. For simplicity, the domain or input space is assumed to be the same as the range or output space, and is a normed linear space. The alternative viewpoint that a system is associated with a state space, implying causality of the system operator, will be only incidental in this chapter.

Although our development is most concerned with deterministic systems, the theory developed may be adapted to stochastic systems. For instance, time-invariance corresponds to the stochastic concept of stationarity.

6.1 Time-invariance

The framework within which we will work is reiterated below.

An input-output (i-0) system is a list <K,X,Y> where X is a normed linear space of functions $x:R \to Y$, Y itself being a normed linear space, and $K:X \to X$ is the system operator.

The shift operator $S(\tau)$ acts on X as follows:

$$(S(\tau)x)(t) = x(t-\tau).$$

It is clear that $S(\tau)$ is a linear operator, in fact an isometric isomorphism on X.

Recall that the norm of an operator G on X is defined as $||G|| = \sup_{x \neq 0} \frac{||Gx||}{||x||}$. It will always be assumed that although any operator G may be nonlinear, $G(0) = 0$.

An i-0 system $K:X \to X$ is time-invariant if $KS(\tau) = S(\tau)K$ for all $\tau \varepsilon R$.

The difference $KS(\tau) - S(\tau)K$ appears in quantum physics as the commutator of K and $S(\tau)$, written $[K,S(\tau)]$. For convenience we follow this convention. Thus a system K is time-invariant if and only if $[K,S(\tau)]$ for all τ.

A system is time-varying if it is not time-invariant.

Denote the space of all functions from X to X by $[X \to X]$.

We denote the linear space of all bounded operators on X by $\mathcal{B}(X)$, and the subspace of $\mathcal{B}(X)$ consisting of the time-invariant operators by $\mathcal{C}(X)$. Thus $\mathcal{C}(X) \subset \mathcal{B}(X) \subset [X \to X]$. A standard result [P2], [W2] is stated as the following theorem.

Theorem 1.1

$\mathcal{C}(X)$ is a closed left-distributive subalgebra of $\mathcal{B}(X)$.

We shall drop the qualification "left-distributive" and use the term "algebra" to mean a "left-distributive algebra" in the sequel. Of course when the systems are all linear the algebras become both left- and right-distributive. Recall that an algebra is a Banach algebra when it is closed under some topology, in this case the norm topology of the operator space; in other words the linear space structure of the algebra is a Banach space.

When X is a Banach space then $\mathcal{B}(X)$ is a Banach algebra , and therefore so is $\mathcal{C}(X)$. The multiplicative operation is operator composition. However $C(X)$ is not an ideal of $\mathcal{B}(X)$. Despite this it is true that the quotient space $\mathcal{B}(X)/C(X)$ is a normed linear space (although not a normed algebra) since $C(X)$ is a closed linear subspace.

6.2 A first order measure

Beginning with the observation that $[K,S(\tau)]$ is not zero for some τ in a time-varying system K, it is natural to regard the norm of $[K,S(\tau)]$ as some indication of the deviation from time-invariance for time shift τ.

A slight but useful modification will sometimes be made to our definition of the space X. Occasionally X will be taken to be the space of functions defined over some closed interval of R or a semi-infinite interval of R. A justification for this is that often in practice two systems are compared for their time-varying characteristics only over some finite duration of an experiment, and many systems are defined for nonnegative time only.

Consider the map $\phi(K):R \to [X \to X]$ given by $\tau \to [K,S(\tau)]$. Often this map is continuous, but continuity is assumed only when needed. Observe the map $\tau \to S(\tau)$ is sometimes not continuous in the uniform topology for $\mathcal{B}(X)$, although it is practically always strongly continuous. This does not preclude the continuity of $\phi(K)$ where $[X \to X]$ has the uniform topology.

Example 6.2.1

If the map $\tau \to S(t)$ is continuous then the map $\phi(K)$ is continuous, because

$$||KS(\tau+\Delta\tau)-S(\tau+\Delta\tau)K-KS(\tau)+S(\tau)K||<2||K||\ ||S(\tau+\Delta\tau)-S(\tau)||.$$

Example 6.2.2

Let the linear operator $K:L_1(-\infty,\infty)\to L_1(-\infty,\infty)$ be defined by the singular Volterra equation

$$(K\ )(t) = \int_{-\infty}^{t} k(s,t)u(s)ds$$

where k(s,t) is Lipshitz continuous with Lipshitz constant c in each of its arguments. Then

$$(KS(\tau)u)(t) = \int_{-\infty}^{t} k(s,t)u(s-\tau)ds$$

$$= \int_{-\infty}^{t-\infty} k(s+\tau,t)u(s)ds$$

$$(S(\tau)Ku)(t) = \int_{-\infty}^{t-\tau} k(s,t-\tau)u(s)ds$$

So

$$[(S(\tau)K-KS(\tau) - S(\tau+\Delta\tau)K + KS(\tau+\Delta\tau))u](t)$$

$$= \int_{-\infty}^{t-\tau} [k(s+\tau,t) - k(s,t-\tau)]u(s)ds$$

$$\int_{-\infty}^{t-\tau-\Delta\tau} [k(s+\tau+\Delta\tau,t) - k(s,t-\tau-\Delta\tau)]u(s)ds$$

$$= \int_{-\infty}^{t-\tau-\Delta\tau} [k(s+\tau,t) - k(s+\tau+\Delta\tau,t) - K(s,t-\tau) + k(s,t-\tau-\Delta\tau)]u(s)ds$$

$$+ \int_{t-\tau-\Delta\tau}^{t-\tau} [k(s+\tau,t) - k(s,t-\tau)]u(s)ds$$

Therefore,

$$||S(\tau)K - KS(\tau) - S(\tau+\Delta\tau)K + KS(\tau+\Delta\tau)||$$

$$= \sup_{||u||=1} ||S(\tau)Ku - KS(\tau)u - S(\tau+\Delta\tau)Ku + KS(\tau+\Delta\tau)u||$$

$$\leq 2c\Delta\tau \int_{-\infty}^{t-\tau-\Delta\tau} |u(s)|ds + 2c\Delta\tau \int_{t-\tau-\Delta\tau}^{t-\tau} |u(s)|ds$$

$$\leq 4c\Delta\tau \to 0 \quad \text{uniformly.}$$

Thus, the map $\phi(K)$ is continuous.

Note: The common Lipshitz constant in both arguments can be relaxed to two different Lipshitz constants.

Lemma 6.2.1

The map $\psi(\tau):\mathcal{B}(X) \to R$ given by $\psi(\tau)K = ||[K,S(\tau)]||$ is a seminorm, and is an even function.

Proof:

$$\begin{aligned}\psi(\tau)(\alpha K+L) &= ||[\alpha K+L,S(\tau)]|| \\ &= ||[\alpha K,S(\tau)] + [L,S(\tau)]|| \\ &\leqslant |\alpha|||[K,S(\tau)]|| + ||[L,S(\tau)]|| \\ &= |\alpha|\phi(\tau)K + \psi(\tau)L .\end{aligned}$$

Evenness is clear and its proof is omitted.

Lemma 6.2.2

$\psi(\tau)$ is a norm on $\mathcal{B}(X)/\mathcal{C}(X)$.

Proof:

$\psi(\tau)$ acts on a co-set $K + \mathcal{C}(X)$ by

$$\psi(\tau)(K+\mathcal{C}(X)) = \psi(\tau)K.$$

Thus $\psi(\tau)(K+\mathcal{C}(X)) = 0 \iff \psi(\tau)K = 0 \iff K\varepsilon\mathcal{C}(X)$.

Denote by $\psi_{[a,b]}$ the map $\mathcal{B}(X) \to R$ induced as follows:

$$\psi_{[a,b]}K = \sup_{\tau\varepsilon[a,b]} \psi(\tau)K.$$

Then it is easily verified that $\psi_{[a,b]}$ is also a seminorm on $\mathcal{B}(X)$ and a norm on $\mathcal{B}(X)/\mathcal{C}(X)$. In particular, denote by ψ by the induced map ψ_R. Strictly speaking ψ may not be finite on $[X \to X]$ so that it may sometimes be necessary to say that ψ is a generalized seminorm or norm, and this will be assumed on occasion.

ψ may serve as a measure of time-invariance in systems. Suppose K,L are two systems such that $\psi K = \psi L$. Then we may conclude that their long-term deviation from time-invariance is the same. However ψ alone will mask some important differences between K and L in most cases.

Example 6.2.3

Let K and L be in $\mathcal{B}(X)$ where

$X = \mathcal{C}(-\infty,\infty)$

$(Kx)(t) = \cos t \cdot x(t)$

$(Lx)(t) = \cos 2t \cdot x(t)$

Then $\psi K = \psi L = 2$ but clearly K and L are periodic with different periods. This will be made more precise in the next section.

Example 6.2.4

Let K and L map $\mathcal{C}(-\infty,\infty)$ into itself by

$(Kx)(t) = tx(t)$

$(Lx)(t) = t^2 x(t)$

Then $\psi K = \psi L = \infty$, but for each interval [a,b] it is seen that $\psi_{[a,b]}L > \psi_{[a,b]}K$.

The above examples suggest that ψ alone is not an adequate characterization of time-variation. This is remedied in section 6.4. But first, a few remarks about ψ itself are in order.

Both $\psi(\tau)$ and ψ are (ordinary) norms on $\mathcal{B}(X)/\mathcal{C}(X)$, and they are generalized norms only on $[X \to X]/\mathcal{C}(X)$. This is easily seen because

$$||KS(\tau) - S(\tau)K|| < ||KS(\tau)||+||S(\tau)K|| < 2||K||$$

since it is clear that $S(\tau)$ is an isometry for each τ.

When $\mathcal{B}(X)/\mathcal{C}(X)$ or $[X \to X]/\mathcal{C}(X)$ is given the quotient norm relative to the operator norm, that is,

$$||K + \mathcal{C}(X)|| = \inf_{C \in \mathcal{C}(x)} || K+C||$$

it is well-known that this in fact yields the quotient topology for the quotient space. The relation of this norm to that induced by $\psi_{[a,b]}$ or ψ is as follows.

Suppose $\{K_n + C(X)\}$ is a sequence converging in the quotient norm to $\{K + C(X)\}$, i.e.,

$$\inf_{C\varepsilon C(x)} || K_n - K + C|| \to 0.$$

We have for each τ and each $C\varepsilon C(X)$:

$$\begin{aligned} ||(K_n-K)S(\tau) - S(\tau)(K_n-K)|| &= ||(K_n-K+C)S(\tau) - S(\tau)(K_n-K+C)|| \\ &\leq 2||S(\tau)|| \; ||K_n-C|| \\ &\leq 2||K_n-K+C|| \end{aligned}$$

So
$$\begin{aligned} \psi(K_n-K+C(X)) &= \sup_{\tau\varepsilon R} ||[K_n-K,S(\tau)]|| \\ &\leq 2\inf_{C\varepsilon C(x)} || K_n-K+C|| \to 0 \end{aligned}$$

giving the result:

Lemma 6.2.3

The quotient norm on $B(X)/C(X)$ or $[X \to X]/C(X)$ is stronger than the ψ-norm.

This leads immediately, by a corollary to the open mapping theorem [T1] on the equivalence of comparable norms on Banach spaces to the following result.

Theorem 6.2.1

If $B(X)$ is a Banach space, then the ψ-norm is equivalent to the quotient norm. In particular, this is true when X is a Banach space.

So in a predominantly interesting class of systems the first measure of time-variance is in fact the natural norm of the quotient space. It clearly remains true that $\psi_{[a,b]_{a,b \in R}}$ is a net defined on the lattice of intervals in R, so induce a lattice of topologies on systems. The upper bound is ψ and the lower bound is the trivial topology.

6.3 Periodic and almost periodic systems

Before moving to more refined measures of time-variance in systems we make a small digression into an important class of systems. In linear system theory the class of systems which are described in the usual notation by

$$\frac{dx}{dt} = A(t)x$$

where $A(t) = A(t+\sigma)$ for each $t \in R$, is called periodic with period σ. The properties of such systems are well-known through the Floquet theory [P3]. The theory has been developed primarily from an internal viewpoint, that is, a state system is postulated. It would be gratifying to have an input-output characterization of a system periodicity in the same way that other system-theoretic properties like causality, passivity and time-invariance are characterized, postulating only an i-0 system. One possible approach is given here.

Definition 6.3.1

A system K is periodic if there exists $\sigma>0$ such that $S(\sigma)K = KS(\sigma)$. σ is called a period of the system.

Observe that this is equivalent, in our notation, to saying $[K,S(\sigma)] = 0$.

It will follow from the next theorem that if a system is periodic in σ, it is also periodic in $k\sigma$, where k is any integer. Also, time-invariant systems are trivially periodic in the sense that the periodicity condition is satisfied by all $\alpha\varepsilon R$. We call periodic systems which are not time-invariant <u>strictly periodic</u>. In such cases an algebraic characterization discussed after the next theorem says that there is a minimum period which divides all other periods.

<u>Theorem 6.3.1</u>

The following are equivalent

(1) K is periodic with period σ

(2) $S(\sigma)S(\alpha)K = S(\alpha)KS(\sigma)$ for some $\alpha \varepsilon R$

(3) $S(\sigma)KS(\alpha) = KS(\alpha)S(\sigma)$ for some $\alpha \varepsilon R$

(4) $S(\sigma)S(\alpha)K = S(\alpha)KS(\sigma)$ for all $\alpha \varepsilon R$

(5) $S(\sigma)KS(\alpha) = KS(\alpha)S(\sigma)$ for all $\alpha \varepsilon R$

(6) $S(p\sigma)K = KS(p\sigma)$ for all $p \varepsilon Z$

<u>Proof</u>:

(1) => (2) : trivially true for $\alpha = 0$.

(2) => (4) : Let (2) be true for some $\alpha\varepsilon R$. Then (4) follows by successively pre-multiplying by $S(-\alpha)$ and $S(\beta)$, where β is arbitrary in R.

(3) => (5) : Similar to above.

(4) => (6) : Let $\alpha = 0$. When $p = 0$ the result is trivially true. An induction argument establishes (6) for non-negative p. The result for negative p follows from this by the obvious pre- and post-multiplications.

(6) => (1) : Let $p = 1$.

It remains to show that (2) => (3) and (5) => (4). These, however, follow by suitable pre- and post-multiplications.

At this point it is interesting to observe the following conclusions. <u>Suppose $\psi(\tau)K$ is continuous in τ, and periodic K has periods which have</u>

<u>some accumulation point in R</u>. <u>Then K is time-invariant</u>. To see this, note that if σ and ν are two periods of K, a simple application of theorem 6.3.1 shows that $|\sigma-\nu|$ is a period of K. Because the collection of periods accumulate, by this result arbitrarily small periods exist. Since multiples of periods are periods, it is now clear that given any real number r there exist periods of K arbitrarily close to r. Hence, by continuity of $\psi(\tau)K$, $\psi(r)K = 0$. Since r was arbitrary this shows that K is time invariant.

As an immediate corollary of this we have the statement that for systems with $\psi(\tau)K$ continuous, it is not possible for time-varying K to have a zero commutator over an interval of R, or equivalently, a system which has zero commutator over an interval of R is necessarily time-invariant.

So, for systems which are smooth enough to have $\psi(\tau)K$ continuous, we may re-define time-invariance as follows:

<u>A system K is time-invariant if</u> $\{\tau:[S(\tau),K] = 0\}$ <u>has an accumulation point</u>.

In fact nothing quite as strong as continuity is required above. It is sufficient that $\psi(\tau)K$ have no discontinuities of the second kind [R1].

It is the above statement which is the justification for experimental i-0 verifications of time invariance in systems by observing that $[S(\tau),K] = 0$ for a large number of τ's within some small interval of R. While a finite number of observations can never establish what is essentially a non-finite property, the above statement does provide a measure of condidence in such an experimental approach toward (partial) verification.

An interesting algebraic characterization of systems may be obtained in the following way.

Let $\Pi_K = \{\tau : [S(\tau), K] = 0\}$ for a system K. Then $\Pi_K = \{0\}$ if and only if K is time varying but not periodic. Next, observe that if α, β are periods of a periodic system K $k\alpha + \ell\beta$ is also a period of K, where k, $\ell \varepsilon Z$ the set of integers. Thus Π_K is a Z-module. More can be said. Π_K is a cyclic nontrivial Z-module if and only if K is strictly periodic (i.e., periodic but not time-invariant). The positive generator of Π_K is then the period (meaning the smallest period) of K. In the forward direction the truth of the implication is obvious. In the backward direction it suffices to show that if Π_K is nontrivial and acyclic, then there is an accumulation point of Π_K. Let $\beta \varepsilon \Pi_K$, and $\beta \neq 0$. Since Π_K is acyclic $\Pi_K \neq \{k\beta : k \varepsilon Z\}$, so there is α in Π_K such that

$$k\beta < \alpha < (k+1)\beta$$

for some $k \varepsilon Z$. Thus $0 < \alpha - k\beta < \beta$, and this shows that given arbitrary $\beta \varepsilon \Pi_K$, there exists $\gamma \varepsilon \Pi_K$ such that $0 < \gamma < \beta$, i.e. there is an accumulation point in $[0, \beta]$.

These observations may be summarized by the following diagram

system K, Π_K

- Π_K is trivial <=> K is time varying and non-periodic
- Π_K is nontrivial
  - Π_K is cyclic <=> K is strictly periodic
  - Π_K is acyclic <=> K is time invariant

Example 6.3.1

For the system K in example 6.2.3,

$$[KS(2\Pi)x] = \cos t \cdot x(t-2\Pi)$$

$$[S(2\Pi)Kx] = \cos\ (t-2\Pi)\cdot x(t-2\Pi)$$

$$= \cos t \cdot x(t-2\Pi)$$

so indeed $KS(2\Pi) = S(2\Pi)K$

and it is easily seen that 2Π is the minimal number for which the above is true. Hence the system is periodic with period 2Π.

Similar reasoning will show that the system L in example 6.2.3 has period Π

More generally, given f a scalar-valued function, define K by

$$(Kx)(t) = f(t)x(t)$$

Then this system is periodic with period σ if and only if f is periodic with period σ. If we restrict consideration to a point in the input space X which is a constant function, the above system recaptures the classical function-theoretic meaning of periodicity, for in this case, say, $X_0(t) \equiv 1$, so

$$(KX_0)(t) = f(t)x(t)$$

$$(KS(\sigma)X_0)(t) = (KX_0)(t) = f(t)$$

$$(S(\sigma)KX_0)(t) = (S(\sigma)f)(t) = f(t-\sigma)$$

and indeed K is periodic if and only if f is periodic.

Example 3.2

In the standard treatment of linear periodic systems the input-output description yields a $K:C(-\infty,\infty) \to C(-\infty,\infty)$

$$(Ku)(t) = \int_{-\infty}^{t} P^{-1}(t)e^{R(t-s)}P(s)u(s)ds$$

where P is a matrix periodic with period σ, and R is a constant matrix. See, for example [P3]. The assumption here is that the system begins in the 0-state.

Now,

$$[S(\sigma)K_u](t) = \int_{-\infty}^{t} P^{-1}(t-\sigma)e^{R(t-\tau)}P(\tau-\sigma)u(\tau-\sigma)d\tau$$

$$= \int_{-\infty}^{t} P^{-1}(t)e^{R(t-\tau)}P(\tau)u(\tau-\sigma)d\tau$$

$$[KS(\sigma)u](t) = \int_{-\infty}^{t} P^{-1}(t)e^{R(t-s)}P(s-\sigma)u(s-\sigma)ds$$

$$= \int_{-\infty}^{t} P^{-1}(t)e^{R(t-s)}P(\ s\ )u(s-\sigma)ds$$

$$= [S(\sigma)K_u](t)$$

thus vindicating the reasonableness of our definition.

Beginning with this i-0 definition of periodicity, assuming that a state system exists, it is conjectured that a Floquet-like state transition function must exist, and if so the definition is, in a sense, complete.

Corollary 6.3.1

If K is periodic with period σ then $[K,S(\tau)]$ regarded as a system is periodic.

Proof:

By expanding $S(\sigma)[K,S(\tau)]$ and $[K,S(\tau)]S(\sigma)$ and appealing to conditions (4) and (5) of the theorem. (See the proof of the next lemma.)

A partial converse to the above corollary is the next result.

Corollary 6.3.2

If $[K,S(\tau)]$ is periodic with period σ then $[K,S(\sigma)]$ regarded as a system is time-invariant.

Proof:

$[K,S(\tau)]$ periodic with period $\sigma \Leftrightarrow S(\sigma)KS(\tau) - S(\sigma)S(\tau)K = KS(\tau)S(\sigma) - S(\tau)KS(\sigma)$ for all $\tau \in R$ by theorem 6.3.1.

Rearranging terms gives

$$(SK(\sigma) - S(\sigma)K)S(\tau) = S(\tau)(KS(\sigma) - S(\sigma)K \quad \text{for all } \tau \in R.$$

Since time-invariant operators form the 0 of the space $[X \to X]/C(X)$, in this quotient space the previous two corollaries would in fact be full converses if $[K,S(\tau)]$ is independent of coset representative. That this is indeed so may be seen from the definition of $\psi(\tau)$, but reiterated below for completeness.

Let K_1,K_2 be elements of the coset $K+C(X)$. Then $K_1-K_2 = C$ for some time-invariant C. Hence $(K_1-K_2)S(\tau) - S(\tau)(K_1-K_2) = 0 \;\forall \tau$, giving $K_1S(\tau) - S(\tau)K_1 = K_2S(\tau) - S(\tau)K_2 \forall \tau$, i.e., $[K_1,S(\tau)] = [K_2S(\tau)]$.

The above argument also shows that within the same coset all elements have the same period, if any exists. This suggests the question: Does the set $P(X)$ of all periodic systems form a linear subspace of $[X \to X]$? For, if so, by the statement above on coset properties of periods, it follows that the partition on systems induced by $C(X)$ must refine that induced by $P(X)$.

That this is so is the substance of the next theorem,

Theorem 6.3.2

$P(X)$ is a subalgebra of $[X \to X]$, and $[X \to X]/_{P(X)}$ is refined by $[X \to X]/C(X)$.

Proof:

If K,L are periodic with periods σ and η, then it is easily shown that $K+L$ and $K\cdot L$ are periodic with period the lowest common multiple of

σ and η.

It is clear that αK is periodic for any scalar α.

The above theorem is a purely algebraic statement. The corresponding topological statement cannot be made unless the space $P(X)$ is closed under the operator norm topology. It turns out that $P(X)$ is in fact not closed under the operator norm topology. However, a slightly weaker class of systems, which includes $P(X)$, is indeed closed. To investigate this enlarged class an input-output characterization is given.

Definition 6.3.2

A system K is called almost periodic if the following two conditions are satisfied:

(1) For each $\varepsilon > 0$ there exists σ_ε such that

$$||S(\sigma_\varepsilon)K - KS(\sigma_\varepsilon)|| < \varepsilon$$

(2) For each $\varepsilon > 0$ there is $\ell(\varepsilon) > 0$ such that every interval of length $\ell(\varepsilon)$ contains a σ_ε satisfying the above inequality. Following the terminology of almost periodic functions we call σ_ε a shift number.

The theory of periodic and almost periodic systems closely parallels that for scalar and Banach space valued periodic and almost periodic functions. The class of almost-periodic systems will be denoted by $A(X)$. It is clear that $P(X) \subset A(X)$.

The difference between the above definition for almost periodicity and that for a scalar valued almost periodic function f is that condition (1) is, in the scalar function case,

$$|f(t) - f(t+\sigma_\varepsilon)| < \varepsilon .$$

It turns out that K is almost periodic if and only if $\psi(\tau)K$ is almost periodic as a scalar function. This is recorded later as corollary 6.3.4.

Precisely the same remarks may be made about recovering the classical meaning of almost-periodicity as were made about periodicity in the discussion following example 6.3.1.

To explore some properties of $P(X)$ and $\mathcal{A}(X)$ it will prove expedient to use the commutator and its norm.

<u>Theorem 6.3.3</u>

If K is periodic (resp. almost periodic) then $\psi(\tau)K$ is periodic (resp. almost periodic) as a scalar function of τ.

<u>Proof</u>:

When K is periodic the result follows from Corollary 6.3.1.

To prove it for almost periodic K we have to exhibit, for each $\varepsilon > 0$, a shift number σ_ε such that $|\psi(\tau)K - \psi(\tau+\sigma_\varepsilon)K| < \varepsilon$ for all τ. The demonstration of this is best separated out as the next lemma.

<u>Lemma 6.3.1</u>

If $||S(\sigma)K - KS(\sigma)|| < \varepsilon$ then

$$\big|\, ||S(\sigma+\tau)K - KS(\sigma)S(\tau)|| - ||S(\tau) - KS(\tau)||\, \big| < \varepsilon$$

for all τ .

<u>Proof</u>:

$$KS(\tau)S(\sigma) - S(\sigma)S(\tau)K$$

$$= KS(\tau)S(\sigma) - S(\tau)KS(\sigma) + S(\tau)KS(\sigma) - S(\sigma)S(\tau)K$$

So

$$0 \leq ||KS(\tau)S(\sigma) - S(\sigma)S(\tau)K|| < ||KS(\tau) - S(\tau)K||\ ||S(\sigma)|| + ||S(\tau)||\varepsilon$$

$$= ||KS(\tau) - S(\tau)K|| + \varepsilon$$

Conversely,

$$KS(\tau)S(\sigma) - S(\tau)KS(\sigma)$$

$$= KS(\tau)S(\sigma) - S(\sigma)S(\tau)K + S(\sigma)S(\tau)K - S(\tau)KS(\sigma)$$

giving

$$0 \leqslant ||KS(\tau) - S(\tau)K|| \leqslant ||KS(\tau)S(\sigma) - S(\sigma)S(\tau)K|| + \varepsilon$$

and the two inequalities prove the result.

It is evident that if $\psi(\tau)K = 0$ for some $\tau_0 \neq 0$ then K is necessarily periodic, since the commutator $[K,S(\tau_0)] = 0$ in this case. The simple converse of the Theorem 6.3.3 is clearly untrue since it is possible that $\psi(\tau)K > 0$ and yet is periodic, so K is not periodic. Thus a necessary and sufficient characterization for K to be periodic is that $\psi(\tau)K= 0$ for some nonzero τ. This is recorded as a corollary.

Corollary 6.3.3

K is periodic if and only if $\psi(\tau)K = 0$ for some non-zero τ.

The situation for almost periodic systems is similar, but not quite as simple. Whenever $\psi(\tau)K$ is almost periodic we have that given $\varepsilon > 0$ there is $\ell(\varepsilon)$ such that for every interval of length $\ell(\varepsilon)$ there is a shift number σ_ε, i.e., for all τ

$$|\psi(\tau+\sigma_\varepsilon)K - \psi(\tau)K| < \varepsilon$$

In particular, for $\tau = 0$ we have

$$||S(\sigma_\varepsilon)K - KS(\sigma_\varepsilon)|| < \varepsilon$$

showing that K is almost periodic. We state this result as:

Corollary 6.3.4

K is almost periodic if and only if $\psi(\tau)K$ is almost periodic.

The condition analogous to that in Corollary 6.3.3 for this case turns out to be not quite strong enough for sufficiency. This analogous condition is in fact $\lim_{\tau\to\infty} \inf \psi(\tau)K = 0$. It is clearly a necessary condition for K to be almost periodic.

For the sequel we shall need some results about almost periodic functions. These are stated in the next lemma without proof in condensed form in terms of algebras. A good reference is Amerio [A5].

Lemma 6.3.1

The set of (continuous) almost periodic functions is a closed algebra.

Remark:

Continuity is an assumption which is really too strong for the truth of the above. If we consider almost periodic functions in the sense of Stepanov (see reference cited) then we may consider systems which are more general. However we defer this more complete development to a later report and concentrate here on more immediate issues.

Lemma 6.3.2

The set $\mathcal{B}_c(X)$ of bounded systems for which $\phi(K)$ is continuous in τ is a Banach subalgebra of $[X \to X]$.

Proof:

First observe that

$$[S(\tau),\ K+K] = \alpha[S(\tau),K] + [S(\tau),L]$$

$$[S(\tau),K\cdot L] = [S(\tau),K]\cdot L + K\cdot [S(\tau),L]$$

showing that this set is a subalgebra. That it is closed follows from the fact that uniform limits of continuous functions are continuous, and $\phi(K):R \to \mathcal{B}(X)$.

Theorem 6.3.4

The subset $A_c(X)$ of almost periodic systems in $\mathcal{B}_c(X)$ is a Banach subalgebra.

Proof:

Let K, L be almost periodic. It is evident that αK is almost periodic. We wish to show K+L and K·L are almost periodic. By Corollary 6.3.4 it suffices to show that $\psi(\tau)(K+L)$ and $\psi(\tau)(K\cdot L)$ are almost periodic Also, by the same corollary and the hypotheses above, $\psi(\tau)K$ and $\psi(\tau)L$ are almost periodic.

$$||S(\tau)(K+L) - (K+L)S(\tau)|| < ||S(\tau)K - KS(\tau)|| + ||S(\tau)L - LS(\tau)||$$

and the right hand side is a sum of two almost periodic functions, hence by lemma 6.3.1 is almost periodic; call it η. Observe that $\psi(0)(K+L) = 0 = \eta(0)$. Since η is almost periodic, for every $\varepsilon > 0$ there is $\ell(\varepsilon)$ such that for every interval of length $\ell(\varepsilon)$ there is a shift number σ_ε with

$$|\eta(\sigma_\varepsilon+\tau) - \eta(\tau)| < \varepsilon \quad \text{for all } \tau.$$

Choosing $\tau = 0$ we have the result

$$||S(\sigma_\varepsilon)(K+L) - (K+L)S(\sigma_\varepsilon)|| < \varepsilon$$

showing that K+L is almost periodic.

Next, observe that

$$||S(\tau)K\cdot L - K\cdot LS(\tau)|| \leq ||S(\tau)K - KS(\tau)||\ ||L|| + ||K||\ ||S(\tau)L - LS(\tau)||$$

and the right hand side is almost periodic. A similar argument to the above says that K·L is almost periodic.

Finally, suppose $K_n \to K$, and K_n is almost periodic for every n. Then

$$\begin{aligned} 0 &\leq \big|\ ||S(\tau)K_n - K_nS(\tau)|| - ||S(\tau)K - KS(\tau)||\ \big| \\ &\leq ||S(\tau)K_n - K_nS(\tau) - S(\tau)K + KS(\tau)|| \\ &= ||S(\tau)(K_n-K) + (K-K_n)S(\tau)|| \\ &\leq 2||K-K_n|| \to 0 \end{aligned}$$

So $\psi(\tau)K_n \to \psi(\tau)K$.

By corollary 6.3.4 $\psi(\tau)K_n$ is almost periodic for each n, and so by lemma 6.3.1 $\psi(\tau)K$ is almost periodic. Thus, using corollary 6.3.4 again, K is almost periodic.

Corollary 6.3.5

Regarding $A_c(X)$ as a closed linear subspace of $\mathcal{B}(X)$, $\mathcal{B}(X)/A_c(X)$ is a Banach space.

It should be obvious that $A_c(X)$ is not an ideal of $\mathcal{B}(X)$.

6.4 Homomorphisms

Recall that a homomorphism from a system $K:X \to X$ to $K_1:X \to X_1$ such that $hK = K_1h$, and the system K_1 may be regarded as a model of system K.

Example 6.4.1

Let $K:L_2(-\infty,\infty) \to L_2(-\infty,\infty)$. Choose an orthonormal basis for $L_2(-\infty,\infty)$, calling it $\{e_n\}_{n=1}^{\infty}$. If K is linear, then the map h defined by

$$h: \sum_{n=1}^{\infty} \alpha_n e_n \to \sum_{n=1}^{k} \alpha_n e_n$$

yields a homomorphic system K_1 which is also linear. K_1 is K restricted to the subspace generated by $\{e_n\}_{n=1}^{k}$.

More generally [B1] let $\{e_n\}_{n=1}^{\infty}$ be the nodes of a digraph. The branch a_{ij} exists if and only if $\langle e_j, Kx\rangle \neq 0$ for some x such that $\langle e_i, x\rangle \neq 0$. If the infinite matrix representation of this digraph is upper triangular except for an initial submatrix of finite dimension k, then the h defined above is agian a homomorphism. K is not necessarily linear in this case, and the input-output space is any Hilbert space. The point here is that orthonormal bases may sometimes be chosen so that such a dependancy matrix is obtained. This emphasizes the fact that homomorphic simplification is a basis dependant procedure in practice.

When applied to state systems homomorphisms go by the name aggregation [A1] as discussed for the finite dimensional state case in chapter 4.

The result presented in this section relates to the time-varying characteristics of systems which are homomorphic images of other systems. Given a homomorphism $h:X \to X_1$ of systems $K:X \to X$ and $K_1:X_1 \to X_1$, the following arguments may be applied.

It is first assumed that $hS(\tau) = S(\tau)h$ for all τ, that is, h is time-invariant. This is the case for many homomorphisms schemes of importance, and is true in particular for the example above. Notationally, different $S(\tau)$ should be used for the two sides of the equation because they act on different spaces, but where the context is clear we indulge in small liberties.

Whenever h is linear it is easily shown that

$$[S(\tau)K_1 - K_1S(\tau)]h = h[S(\tau)K-KS(\tau)]$$

using $K_1h = hK$ and h time-invariant. This will be useful later in the chapter. Without this assumption, however, we have:

Theorem 6.4.1

If K_1 is a homomorphic image of K and h is time-invariant then K is time-invariant (resp. periodic) implies K_1 is time-invariant (resp. periodic). Moreover, periods are preserved.

Proof:

It suffices to show this for periodic K with period σ.

Because $K_1 h = hK$

$$S(\sigma)K_1 h - K_1 hS(\sigma) = S(\sigma)hK - hKS(\sigma)$$
$$= h[S(\sigma)K - KS(\sigma)]$$
$$= 0$$

But K_1 is periodic with period σ if

$$[S(\sigma)K_1 - K_1 S(\sigma)]x_1 = 0 \quad \text{for all } x_1 \in X_1$$

and since $x_1 = hx$ for some $x \in X$, this is true if $S(\sigma)K_1 h - k_1 S(\sigma)h = 0$, i.e., if $S(\sigma)K_1 h - K_1 hS(\sigma) = 0$, which is the case.

The converse inference problem is as usual much harder and requires additional assumptions. Without further assumptions, at least the following can be said. Suppose K_1 is periodic. Let σ be a period of K_1. Then the range of $S(\sigma)K - KS(\sigma)$ must be in the kernel of h. This means that the subset of X on which K is not periodic is "eliminated" by h in the passage to X_1, a paraphrase of the fact that the choice of h in a modelling situation is partly to ignore portions of an input-output space which do not interest the modeller. Exactly the same remarks could be made about the time-invariance of K if K_1 is time-invariant.

Almost-periodicity, unlike time-invariance and periodicity, is not a purely algebraic concept. So the norms of the respective commutators have to be considered. In almost all cases of interest h is norm bounded, and this will be assumed. We turn our attention to the validity of homomorphisms with respect to almost-periodicity. Keeping the same notation, we have:

$$||S(\tau)K_1 - K_1S(\tau)|| = \sup_{x_1 \neq 0} \frac{||[S(\tau)K_1 - K_1S(\tau)]hx||}{||x_1||}$$

$$\leqslant \sup_{x_1 \neq 0} \frac{||h|| \; ||S(\tau)K - KS(\tau)|| \; ||x||}{||x_1||}$$

$$\leqslant ||h|| \; ||S(\tau)K - KS(\tau)|| \sup_{x_1 \neq 0} \frac{||x||}{||x_1||}$$

where $x \in h^{-1}(x_1)$.

Observing that the ratio is being supremised over x_1 but is true for any $x \in h^{-1}(x_1)$ we may write

$$||S(\tau)K_1 - K_1S(\tau)|| < ||h|| \; ||S(\tau)K - KS(\tau)|| \; M$$

where $M = \sup_{x_1 \neq 0} \inf_{x \in h^{-1}(x_1)} \frac{||x||}{||x_1||}$.

Now, in cases where M is infinite, nothing can be said about the almost-periodicity of K_1 when K is almost-periodic. In case M is finite, it is clear that K almost periodic implies K_1 almost periodic. It turns out that M is indeed finite for many cases of interest, and in fact a reasonable homomorphism which attempts to capture, along with other properties, the time-varying property of K will probably yield an M which is finite (this is clearly not a necessary condition). In the example discussed at the beginning of this section, M is exactly 1. These conclusions are recorded as

Theorem 6.4.2

If $\sup_{x_1 \neq 0} \inf_{x \in h^{-1}(x_1)} \frac{||x||}{||x_1||} < \infty$ then

K almost periodic => K_1 almost periodic

Corollary 6.4.1

If h is a linear transformation from X to X_1, X_1 a subspace or a quotient space of X, then

K almost periodic => K_1 almost periodic.

Analogous to the converse problem for periodicity, h acts such that for each ε there is a σ_ε and

$$||h[S(\sigma_\varepsilon)K - KS(\sigma_\varepsilon)]x|| < \varepsilon||x||$$

whenever K is almost periodic. This inequality says that even if σ_ε is not a shift number for K, h acts on the range of $S(\sigma_\varepsilon)K - KS(\sigma_\varepsilon)$ to eliminate the appearance of non almost periodicity in the passage to X_1.

We digress a little at this point and consider linear systems and their adjoints. If $K:X \to X$ is a linear system, its adjoint is denoted by $K^*:X^* \to X^*$, where X^* is the dual space of X. When X is a Hilbert space $X = X^*$, and then K and K^* act on the same space. If K is linear and h is linear, K_1 will be linear. Linearity will be assumed for the time being.

Theorem 6.4.3

$$\psi(\tau)K = \psi(\tau)K^* \qquad \text{for all } \tau .$$

Proof:

$S(\tau)^* = S(-\tau)$ is easily verified. The assertion above then follows from the fact that $||A^*|| = ||A||$ for linear operator A, and

$$[S(\tau)K - KS(\tau)]^* = K^*S(-\tau) - S(\tau)K^*$$

Corollary 6.4.2

K is time-invariant (periodic, almost-periodic) if and only if K^* is time-invariant (periodic, almost-periodic).

The relation of homomorphisms to adjoints is the following simple one.

Theorem 6.4.4

If h is a linear homomorphism of $K:X \to X$ to $K_1:X_1 \to X_1$, then h* is a linear homomorphism of $K_1^*:X_1^* \to X_1^*$ to $K|_{h^*(X_1^*)}:X^* \to X^*$.

Proof:

Take the adjoint of $hK = K_1 h$.

The nonsurjectivity of h* is clear fmom the simple case when X is finite dimensional, in which case $\dim(X^*) = \dim(X) > \dim(X_1) = \dim(X_1^*)$ whenever h is not an isomorphism.

The range of h* in X is a subspace X_o^* on which K* is time-invariant (periodic) when K_1^* is time-invariant (periodic), by theorem 6.4.1.

For a system K the subset $\bigcap_{\tau \varepsilon R} \{x\varepsilon X:\psi(\tau)Kx = 0\}$ is a subspace Y_φ, being the intersection of kernels of linear operators, on which K has a time-invariant homomorphic image K_1, as the following example shows.

Example 6.4.2

Let X be the linear space of functions from R to R^2 and X_1 be the linear space of functions from R to R. Let K be defined by

$$(Kx)(t)_1 = 1/2\ x(t)_1 + 1/2\ x(t)_2 + \sin t$$

$$(Kx)(t)_2 = 1/2\ x(t)_1 + 1/2\ x(t)_2 - \sin t$$

where the subscripts denote the projections onto that component. Then it is easily verified that $h:R^2 \to R$ given by $(hx)(t) = x(t)_1 + x(t)_2$ is a homomorphism of K to $K_1 = I$ the identity operator, which is time-invariant, but $Y_o \subset R^2$ the time-invariant subspace of K is the trivial {0} - subspace.

Given $K:X \to X$, and $K^*:X^* \to X^*$, denote the associated time-invariant subspaces by Y_o and Z_o respectively, $Y_o \subseteq X$, $Z_o \subseteq X^*$. A relation between Y_o and Z_o is the following. If $x \notin Y_o$, then for some τ, $[S(\tau)K - KS(\tau)]x \neq 0$. By the Hahn-Banach theorem there exists $x' \varepsilon X^*$ such that $<[S(\tau)K - KS(\tau)]x,x'> \neq 0$ which is equivalent to $<x,[K^*S(-\tau)-S(-\tau)K^*]x'> \neq 0$. So $x' \notin Z_o$. Conversely, given $x' \notin Z_o$, $[K^*S(-\tau) - S(-\tau)K^*]\, x' \neq 0$ for some τ. Then there is $x \neq 0$ such that $0 \neq <x,[K^*S(-\tau) - S(-\tau)K^*]x'> = <[S(\tau)K - KS(\tau)]x,x'>$ so $x \notin Y_o$.

6.5 Higher order measures

It was earlier observed that the measure of time-variance so far developed is not entirely adequate. Higher order measures can be introduced through Gateaux and Frechet derivatives of commutators $[S(\tau),K]$ with respect to τ. For a complete discussion of these derivatives in a normed linear space setting Kantorovich [K2] should be consulted. Here we recall some basic definitions and results relevant to our discussion.

If K is an operator from X to Y, and X,Y are normed linear spaces, the Gateaux (G-) derivative of K at $x_o \varepsilon X$ is denoted by $K'(x_o)$, defined by

$$K'(x_o)x = \lim_{\delta \to 0} \frac{K(x_o+\delta x) - K(x_o)}{\delta}$$

for each $x \varepsilon X$. $K'(x_o)$ is a linear operator from X to Y.

If the convergence above is uniform in X, the derivative is called the Frechet (F-) derivative of K at x_o. The same symbol will be used for for both derivatives when the context is clear. If an F-derivative exists at x_o, then K is said to be differentiable at x_o.

If K happens to be linear, K is differentiable at every point x_o in X and $K'(x_o) = K$. nth derivatives are denoted by superscript n.

Suppose the G-derivative of K exists in some region of X. The derivative may be regarded as a map which associates each point x in the region with a linear operator $K'(x)$ from X to Y. Thus $K':X \to L(X,Y)$

K has G-derivatives throughout X. If K' is continuous at x_o, then $K'(x_o)$ is also an F-derivative.

In the case when X is one-dimensional, so that we can identify X with the scalars, say R, it is a fact that $\mathcal{L}(R,Y)$ is isometrically isomorphic to Y. If $L \varepsilon \mathcal{L}(R,Y)$, $L(r) = rL(1)$, so every element in the range of L is a scalar multiple of $L(1) \varepsilon Y$. The mapping $L \to L(1)$ is evidently the isometric isomorphism sought. Thus we can identify $\mathcal{L}(R,Y)$ with Y. In particular, if $P:R \to Y$ is a (nonlinear) operator $P':R \to \mathcal{L}(R,Y)$, so in fact $P':R \to Y$. The conclusion which is useful here is that if $\{P_\alpha\}$ is a collection of operators from R to Y which also contain their first derivatives, then $\{P_\alpha\}$ must also contain all higher-order derivatives. Stated more succinctly, if $\{P_\alpha\}$ is closed under first derivatives then it is analytic.

The $P:R \to Y$ above will take the form $\phi(K):R \to \mathcal{B}(X)$ in our specialization, recalling that $\phi(K)(t) = [K,S(\tau)]$. Corresponding to $\psi(\tau)K = ||[K,S(\tau)]||$, the quantity $||[K,S(\tau)]^n||$ or equivalently $||\phi(K)^n(\tau)||$ will be denoted by $\psi_n(\tau)K$.

If $K,L \varepsilon \mathcal{B}(X)$ are such that $\phi(K)^n(\tau)$ and $\phi(L)^n(\tau)$ exist (either G- or F- derivatives), then $\phi(K+\alpha L)^n(\tau)$ exists for each scalar α. Therefore the subset of $\mathcal{B}(X)$ consisting of systems K for which $\phi(K)^n(\tau)$ exists is a linear subspace of $\mathcal{B}(X)$. In general this is not a closed subspace.

An explicit form for $\phi(K)^n(\tau)$, the nth derivative of $[K,S(\tau)]$, may be obtained. In fact it is easily verified that $\phi(K)^n(\tau) = [K,S(\tau)D^n]$ where D is the operator on X which yields time-derivatives of functions. D is a linear operator which is closed but not necessarily bounded unless it is onto and X is a Banach space (as a consequence of the closed graph theorem [T1]). To render the subsequent discussion more tractable it will be assumed that X is a Banach space. A simple instance of this is when X is the linear space of scalar-valued functions which have bounded continuous derivatives up to the n-th order, and the norm is chosen to be the maximum of the sup-norms on each derivative. Actually, if X is not a Banach space, all that needs to be done is to replace norms by generalized norms in the discussion that follows.

With the assumption, $\phi(K)^n(\tau)$ is in $\mathcal{B}(X)$, so $\phi(K)^n : R \to \mathcal{B}(X)$.

Lemma 6.5.1

$\psi_n(\tau)$ is a seminorm on $\mathcal{B}(X)$, and is an even function in τ.

Proof:

$$\begin{aligned}[K+\alpha L, S(\tau)]^n &= [K+\alpha L, S(\tau)D^n] \\ &= [K, S(\tau)D^n] + \alpha[L, S(\tau)D^n] \\ &= [K, S(\tau)]^n + \alpha[L, S(\tau)]^n\end{aligned}$$

from which the seminorm property is immediate.

Evenness of ψ_n in τ is a consequence of the time-invariance of D.

Denote the set $\{K : \psi_n(\tau)K = 0\}$ by $\mathcal{D}_n(\tau)$. This is a closed linear subspace of $\mathcal{B}(X)$. Closure follows from the equality

$$\begin{aligned}[K, S(\tau)D^n] = &(K-K_m)S(\tau)D^n + [K_m, S(\tau)D^n] \\ &+ S(\tau)D^n(K_m - K)\end{aligned}$$

for a sequence $\{K_m\}$ in $\mathcal{D}_n(\tau)$ converging to K.

This yields the result:

Corollary 6.5.1

$\psi_n(\tau)$ is a norm on $\mathcal{B}(X)/\mathcal{D}_n(\tau)$.

Paralleling the development for $\psi(\tau)$ ($\psi_0(\tau)$ in the above notation), it is noted that the previous lemma and corollary may be extended to $\psi_n[a,b]$ where

$$\psi_n[a,b]K = \sup_{\tau\in[a,b]} \psi_n(\tau)K$$

whence $\psi_n[a,b]$ is also a seminorm on $\mathcal{B}(X)$, and a norm on $\mathcal{B}(X)/\mathcal{D}_n[a,b]$, and is therefore closed. In particular, denote by ψ_n the map $\psi_n(-\infty,\infty)$. The map ψ discussed easlier corresponds to ψ_o. Because of the assumption that X is a Banach space, D^n is bounded so $[K,S(\tau)]^n = [K,S(\tau)D^n]$ is bounded. Further,

$$||[K,S(\tau)D^n]|| < 2||K|| \; ||D^n||$$

showing that ψ_n is an ordinary seminorm for each n. Denoting the subspace $\mathcal{D}_n(-\infty,\infty)$ simply by $\mathcal{D}_n$ it is clear that ψ_n is a norm on $\mathcal{B}(X)/\mathcal{D}_n$.

Lemma 6.5.2

As subspaces the collection $\{\mathcal{D}_n\}$ forms a chain

$$\mathcal{D}_0 \subseteq \mathcal{D}_1 \ldots\ldots \subseteq \mathcal{D}_n \subsetneq \ldots\ldots$$

Proof:

It suffices to show that if $K\varepsilon\mathcal{D}_n$ then $K\varepsilon\mathcal{D}_{n+1}$ for arbitrary n.

If $K\varepsilon\mathcal{D}_n$ then $[K,S(\tau)]^n = 0$ for all τ and the derivative of the 0 operator is 0.

Corollary 6.5.2

The quotient spaces $\{\mathcal{B}(X)/\mathcal{D}_n\}$ form a reverse chain

$$\mathcal{B}(X)/\mathcal{D}_0 \supsetneq \mathcal{B}(X)/\mathcal{D}_1 \ldots\ldots \supsetneq \mathcal{B}(X)/\mathcal{D}_n \supseteq \ldots\ldots$$

Note that in the earlier notation $\mathcal{C}(X)$ corresponds to $\mathcal{D}_0$.

A natural question arising from the foregoing is whether there is any simple way to combine the measures $\psi_n(\tau)$ and ψ_n. We confine our remarks to ψ_n since they also apply to $\psi_n(\tau)$ with the obvious modifications. The answer to this question is summarized in the next theorem.

Theorem 6.5.1

Let ψ_n exist for n = 1,2,...k (possibly for all integers n).

The collection $\{\phi_n\}_{n=1,\dots,k}$ generates a topology on $\mathcal{B}(X)$ with a base at 0 consisting of the family of all finite intersections

$$r_1 V(\psi_1) \cap r_2 V(\psi_2) \dots \cap r_n V(\psi_n), \quad r_i > 0 \quad n < k$$

where $V(\psi_n) = \{K:\psi_n(K) < 1\}$. With this topology, $\mathcal{B}(X)$ is a locally convex linear topological space.

Proof:

From the foregoing discussion $\{\psi_n\}_{n=1,\dots,k}$ is a collection of seminorms on $\mathcal{B}(X)$. The theorem is a standard result [T1] in functional analysis, given this fact.

It is noted that if $k < \ell$, the locally convex topology generated by $\{\psi_n\}_{n=1,\dots,k}$ is weaker than that generated by $\{\psi_n\}_{n=1,\dots\ell}$. If ψ_n exists for all integers n, and the locally convex topology generated by this family is considered, it is clearly weaker than the norm topology. (It would be interesting to discover under what circumstances it would be equal to the norm topology, but we do not pursue this matter.)

Paraphrasing the last few results it may be said that each measure ψ_n successively refines the classification of systems according to successively higher orders of time-variance. Thus corollary 6.5.2 states that this refinement proceeds in an orderly fashion, with the n^{th} order characteristics factored out before the $(n+1)^{th}$. From the viewpoint of modelling it is the reverse of the refinement process that is relevant. A coset in $\mathcal{B}(X)/\mathcal{D}_n$ is the form $K+\mathcal{D}_n$, and consists of systems which are not distinguished insofar as (n+1)th order time-variance is neglected A canonical map from $\mathcal{B}(X)/\mathcal{D}_n$ to $\mathcal{B}(X)/\mathcal{D}_{n-1}$ coalesces two cosets $K+\mathcal{D}_n$ and $L+\mathcal{D}_n$ whenever nth order differences between K and L can be neglected.

In the previous section it was shown that an i-0 homomorphism h from system K to system K_1, under certain conditions, preserved the 0-th order property of K in the passage to K_1. Higher order properties will be investigated here.

If h is <u>time-invariant</u>, <u>linear</u> and <u>continuous</u>, then

$$hD = h \lim_{\delta\to 0} \frac{S(\delta) - S(0)}{\delta}$$

$$= \lim_{\delta\to 0} h \frac{S(\delta) - S(0)}{\delta} \qquad \text{by continuity}$$

$$= \lim_{\delta\to 0} \frac{S(\delta)h - S(0)h}{\delta} \qquad \text{by time-invariance and linearity}$$

$$= \lim_{\delta\to 0} \frac{S(\delta) - S(0)}{\delta} h$$

$$= Dh$$

<u>Theorem 6.5.2</u>

If h is an i-0 homomorphism which is linear, time-invariant and continuous, $K\varepsilon\mathcal{D}_n \Rightarrow D_1\varepsilon\mathcal{D}_n$.

<u>Proof</u>:

$K\varepsilon\mathcal{D}_n \Leftrightarrow KS(\tau)K^n = S(\tau)D^nK$

The above remarks on h are easily extended to show

$hD^n = D^nh$

Then

$$K_1S(\tau)D^nx_1 = K_1S(\tau)D^nhx, \quad x\varepsilon X$$
$$= K_1S(\tau)hD^nx$$
$$= K_1hS(\tau)D^nx$$
$$= hKS(\tau)D^nx$$
$$= hS(\tau)D^nKx, \quad K\varepsilon\mathcal{D}_n$$
$$= S(\tau)D^nhKx$$
$$= S(\tau)D^nK_1hx$$
$$= S(\tau)D^nK_1x_1$$

so $K_1\varepsilon\mathcal{D}_n$.

For any $K\varepsilon\mathcal{B}(X)$, the elements of the sequence $\{\psi_n(K)\}$ may be zero for some index m, in which case lemma 6.5.2 says $\psi_n(K) = 0$ for $n > m$. Call the least such m the index of K. In these terms the preceding theorem states that linear, continuous, time-invariant i-0 homomorphisms cannot increase the index of systems. This statement generalizes the corresponding statement concerning ψ_o in the pervious section.

In fact the proof of theorem 6.5.2 shows that under the assumptions about h,

$$h[K,S(\tau)]^n = [K_1,S(\tau)]^n h$$

This yields a generalization of theorem 6.4.2.

Theorem 6.5.3

If h is linear, continuous, time-invariant and satisfies

$$M = \sup_{x_1 \neq 0} \inf_{x \epsilon h^{-1}(x_1)} \frac{||x||}{||x_1||} < \infty$$

then

$$||[K_1,S(\tau)]^n|| < M||h|| \; ||[K,S(\tau)]^n||$$

Proof:

Using the equality $h[K,S(\tau)]^n = [K_1,S(\tau)^n h$, mimic the proof of theorem 6.4.2.

This result is paraphrased as saying that h maps a base of the ψ_n-induced seminorm topology on $\mathcal{B}(X)$ to the corresponding one on $\mathcal{B}(X_1)$. Thus if $\{K_n\}$ is a sequence of operators converging in the locally convex topology (of time-varying characteristics) to an operator K, the homomorphic image sequence $\{hK_n\}$ must converge likewise to hK. That the converse is generally untrue is shown by example 6.4.2.

This section is concluded with a few miscellaneous remarks.

K is time-invariant if and only if it has index zero.

If K is periodic then $\psi_n(K)$ is periodic for each n. To prove this, it is sufficient to note that $[K,S(\sigma)] = 0$ implies $S(\sigma)[K,S(\tau)]^n = [K,S(\sigma+\tau)]^n$.

For linear, bounded K it is easy to verify that $([K,S(\tau)]^n)^* = [K^*,S(-\tau)]^n$, hence $\psi_n(K) = \psi_n(K^*)$ for all n.

The topology induced by the ψ_n is not the only one that is meaningful for measuring time-variance. Weighted integrals of the action of the ψ_n may serve the same function, and in some instances may prove to be more meaningful. The purpose of the preceeding theory is to demonstrate that such topologies are indeed available.

6.6 Discussion

The bulk of this chapter consists of an attempt to make precise the fuzzy predicate of "slowly time varying". It should be clear that this kind of fuzziness cannot be removed easily by a strict numerical ordering, for if this is possible the fuzziness would not have arisen in the first place. Our approach was to construct a topological measure of time-variance, thus essentially capturing the idea of a "distance" measure without really introducing a metric. This was the content of theorem 6.5.1 . It was shown in theorem 6.4.1 that

$$\text{TIME-INV\_HOM}(S,S') \supset (\text{TIME-INV}(S) \supset \text{TIME-INV}(S'))$$

and

$$\text{TIME-INV\_HOM}(S,S') \supset (\text{PERIODIC}(S) \supset \text{PERIODIC}(S'))$$

We have no convenient name for the condition placed on h in the premise of theorem 6.4.2. If we did, it is clear how the theorem could be written out as above, i.e., this instance of a homomorphism is ALMOST-PERIODIC valid. In particular Corollary 6.4.1 says that

$$\text{LINEAR\_SUBSPACE\_HOM}(S,S') \supset (\text{ALMOST-PERIODIC}(S) \supset \text{ALMOST-PERIODIC}(S'))$$

With higher order measures of time-variance these results are generalized so as to replace the predicates TIME-INV and PERIODIC by more sensitive predicates for which we do not as yet have appropriate names. The generalization to theorem 6.5.3 has special significance as indicated by the remarks following the theorem. Let us define a predicate CONVERGE such that

CONVERGE $(\{S_n\},S) \iff S_n$ converges to S in the topology for time-variance.

which is an abbreviation of a second order formula of the type discussed in chapter 0. Then theorem 6.5.3 may be paraphrased to say that linear, continuous time-variant homomorphisms are CONVERGE-valid.

Finally, looking at the index of systems as defined in the remarks following theorem 6.5.2, the same instance of a homomorphism was shown to be UB_INDEX valid, where UB_INDEX is a binary predicate such that

UB_INDEX(m,S) <=> m is an upper bound on the index of S.

Chapter 7

In this closing chapter we discuss a few selected open questions motivated by the foregoing results.

In most instances of homomorphism schemes which were tested against systems-theoretic predicates, the particular instantiation gave sufficient conditions for validity with respect to those predicates. It is not known whether a complete characterization of valid instances with respect to arbitrary sets of predicates is possible. This correpsonds to finding necessary and sufficient conditions for preservation of system properties and relations. It is conjectured that this is in general not possible but I have not examined the question in detail.

The noncomputability of system predicates for continuous systems may be mitigated by considering an analog theory of computability, as mentioned in Chapter 0. Developments in this direction could lead to new insights about continuous systems and their realizations.

System realization theory for time-varying systems was not studied because a canonical realization theory is not available. If one relaxes the condition of canonicity, it should be possible to relate the i-0 theory of Chapter 6 to a state-space theory, which will exhibit the time-varying characteristics induced by i-0 functions and vice-versa.

One predicate which was omitted is structural stability. Because this features very prominently in modern topological dynamics it is important that this be studied in the near future. It is anticipated that results in this area will contribute to a theory connecting local to global system homomorphisms.

Structured homomorphisms [Z1] were not pursued in this investigation. Since much realistic modeling is performed with coordinatized sets it would be desirable to develop a topological theory for such homomorphisms paralleling the preceding theory. Deeper results ought to be obtainable because of the much richer structure accompanying these homomorphisms.

Finally, the whole issue of approximate and time-varying homomorphisms and stochastic systems is wide open. Convergence criteria should be established for approximate homomorphisms, and the relation between deterministic and nondeterministic systems via simplification procedures should be investigated. There are many classical problems in the approximation theory of systems which would benefit from being viewed in the perspective of predicate validity. Conversely, new procedures may in fact be suggested by this perspective.

REFERENCES

(A1) Aoki, M. "Control of Large Scale Dynamic Systems by Aggregation," IEEE Trans. on Automatic Control, Vol. AC-13, No. 3, June 1968.

(A2) Ando, A., Fisher, F. M., Simon, H. A. Essays on the Structure of Social Science Models, MIT Press, 1963.

(A3) Arbib, M. A. "Decomposition Theory for Automata and Biological Systems," JACC, Aug. 1971, Systems Structure, IEEE Pub. No. 71C61-CSS.

(A4) ___, "Automata Theory and Control Theory: A Rapprochement," Automatica, 3:161-189, 1966.

(A5) Amerio, L. Almost-Periodic Functions and Functional Equations, Van Nostrand Reinhold, 1971.

(B1) Barto, A. G. Private Communication, 1973.

(B2) Balakrishnan, A. V. "State Space Theory of Linear Time-Varying Systems," in System Theory, ed. Zadeh & Polak, McGraw-Hill, 1969.

(D1) DeSantis, R. M. Causality Structure of Engineering Systems, Ph.D. Thesis, Computer Information and Control Engineering, 1971, University of Michigan.

(D2) Damborg, M. and Naylor, A. "The Fundamental Structure of Input-Output Stability for Feedback Systems," IEEE Trans. System Science and Cybernetics, Vol. SSC-6, No. 2, April 1970.

(D3) Day, J. M. "Semigroup Acts, Algebraic and Topological," Proceedings of the Second Florida Symposium on Automata and Semigroups, 1971, Department of Mathematics, University of Florida, Gainsville.

(D4) Dugundji, J. Topology, Allyn & Bacon, 1966.

(F1) Foo, N. "Topologizing State Construction" (Abstract). Proceedings Seventh Princeton Conference on Information Sciences and Systems, March 1973.

(F2) ___, "Lattices of Homomorphisms" (Abstract) Proceedings Eighth Princeton Conference on Information Sciences and Systems, March 1974.

(H1) Hartmanis, J. and Stearns, R. Algebraic Structure Theory of Sequential Machines, Prentice Hall, 1966.

(K1) Kalman, R. E., Falb, P. L., and Arbib, M. A. Topics in Mathematical System Theory, McGraw-Hill, 1969.

(K2) Kantorovich, L. and Akilov, G. Functional Analysis in Normed Spaces, MacMillan, 1964.

(M1) Mihram, G. A. Simulation: Statistical Foundations and Methodology, Academic Press, 1972.

(M2) ____, "The Modeling Process," IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-2, No. 5, Nov. 1972.

(M3) Mesarovic, Macko & Takahara, Theory of Hierarchical, Multilevel Systems, Academic Press, 1970.

(N1) Naylor, A. W. and Sell, G. Linear Operator Theory in Engineering and Science, Holt, Rinehart & Winston, 1971.

(O1) Onat, E.T. and Geary, J.A. "Representation of a class of nonlinear systems" Proceedings of the First International Symposium on Category Theory applied to Computation and Control, 1974; University of Massachusetts, Amherst.

(P1) Park, D. J. M. "The Hierarchical Structure of Metabolic Networks and the Construction of Efficient Metabolic Simulators," To appear in the Journal of Theoretical Biology.

(P2) Porter, W. A. and Zahm, C. L. Basic Concepts in System Theory, University of Michigan Technical Report 01656-2-T, Department of Electrical Engineering, 1969.

(P3) Porter, W. A. Modern Foundations of Systems Engineering, MacMillan, 1966.

(P4) Pour-El, Marian B. "Abstract Computability versus Analog-generability," Proceedings Cambridge Summer School in Logic, 1971, Springer-Verlag.

(R1) Rudin, W. Principles of Mathematical Analysis, McGraw-Hill, 1964.

(S1) Simon, H. A. "The Architecture of Complexity," Proceedings of the American Philosophical Society, 106:467-482, Dec. 1962.

(S2) Saeks, R. "Causality in Hilbert Space," SIAM Review, Vol. 12, No. 3, July 1970.

(S3) Sandberg, I. W. "Conditions for the Causality of Nonlinear Operators defined on a Linear Space," Quarterly of Applied Math., Vol. 23, No. 1, 1965.

(T1) Taylor, A. E. Introduction to Functional Analysis, John Wiley, 1958.

(W1) Wasow, W. R. Assymptotic Expansions for Ordinary Differential Equations, Interscience, 1965.

(W2) Willens, J. C. The Analysis of Feedback Systems, MIT Press, 1971.

(Z1) Zeigler, B. P. "Towards a Formal Theory of Modeling and Simulation: Structure Preserving Morphisms," JACM, Vol. 19, No. 4, Oct. 1972.

(Z2) ___, "Modeling and Simulation: Structure Preserving Morphisms for Discrete and Continuous Systems," Computers and Automata, Proceedings 21st International Conference, Polytechnic Institute of Brooklyn, N. Y., 1972.

(Z3) ___, "On the Formulation of Problems in Modeling and Simulation within the Framework of Mathematical Systems Theory," Proceedings Sixth International Congress on Cybernetics, Namur, Belgium, 1972.

(Z4) Zeigler, B. P. and Weinberg, R. "System Theoretic Model Analysis: Computer Simulation of a Living Cell," J. Theoret. Biology, 29, 1970.

(Z5) Zeigler, B. P. Theory of Modeling and Simulation, (Forthcoming Book) John Wiley, 1975.

(Z6) Zadeh, L. A. "The Concepts of System, Aggregate, and State in System Theory," in System Theory, ed. Zadeh & Polak, McGraw-Hill, 1969.

(Za1) ___, "Fuzzy Sets," Information and Control, 8, 1965.