RECENT DEVELOPMENTS AND FUTURE
PERSPECTIVES IN NONLINEAR
SYSTEM THEORY

John L. Casti

August 1980
WP-80-123

PREFACE

Results on controllability, observability and realization
of input/output data for linear systems are well-known and
extensively covered in a variety of books and papers.  What
is not so well-known is that substantial progress has been
made in recent years on providing similarly detailed results
for nonlinear processes.  This paper represents a survey of
the most interesting work on nonlinear systems, together with
a discussion of the major obstacles standing in the way of a
comprehensive theory of nonlinear systems.

# 1. Basic Problems and Results in Linear System Theory

The theory of linear dynamical processes has by now been developed to such an extent that it is only a slight exaggeration to term it a branch of applied mathematics, sharing equal rank with more familiar areas such as hydrodynamics, classical and quantum mechanics and electromagnetism, to name but a few. For those who doubt this assessment of linear system theory, a perusal of some of the more advanced recent literature [1-5] should prove to be an enlightening activity, showing how deeply imbedded system-theoretic concepts are in areas such as algebraic geometry, differential topology and Lie algebras. Conversely, the "purer" parts of mathematics have proven to be fruitful sources of inspiration for system theorists seeking more powerful tools with which to analyze and classify broad classes of problems.

Encouraged by the tremendous success in the study of linear processes, system theorists have been increasingly turning their attention and methods to the analysis of the same circle of questions for nonlinear systems. As one would suspect, the jungleland of nonlinearity is not easily tamed and so far no comprehensive theory has emerged capable of treating general nonlinear processes with the detail available in the linear case. Nonetheless, substantial progress has been made on several fronts and part of our story will be to survey some of the more interesting developments.

An equally important part of the picture we wish to present is to outline some of the reasons why a complete theory of nonlinear systems seems remote, at least at our current level of

mathematical sophistication. All current indications point
toward the conclusion that seeking a completely general theory
of nonlinear systems is somewhat akin to the search for the Holy
Grail: a relatively harmless activity full of many pleasant sur-
prises and mild disappointments, but ultimately unrewarding. A
far more profitable path to follow is to concentrate upon special
classes of nonlinear problems, usually motivated by applica-
tions, and to use the structure inherent in these classes as
a guide to useful (i.e., applicable) results. As we go along
in this survey, we shall try to emphasize this approach by
example, as well as by precept.

Before entering into the mainstream of nonlinear system
theory and the problems inherent therein, let us briefly review
the principal questions and results of the linear theory. We
are concerned with a process described by the system of
differential equations

$$\frac{dx}{dt} = Fx(t) + Gu(t) \quad , \qquad x(0) = x_0 \quad , \qquad (\Sigma)$$

$$y(t) = Hx(t) \quad ,$$

where x, u and y are n, m and p-dimensional vector functions,
taking values in the vector spaces X, U and Y, respectively.
For ease of exposition, we assume that the matrices F, G and H
are constant, although the theory extends easily to the time-
varying case at the expense of more delicate notation and
definitions.

The principal questions of mathematical system theory may
be conveniently separated into three categories:

A.  Reachability/Controllability — given an admissible set of input functions $\Omega$, what region $\mathscr{R}$ of the system state space X can be reached from the initial state $x_0$ in some prescribed time T by application of inputs $u \in \Omega$.  If $x_0 \neq 0$ and $\mathscr{R} = 0$, then we have a problem of controllability; otherwise it is a question of reachability.  In the case of constant F and G (the output matrix H plays no role in category A problems), with $\Omega$ = continuous functions on $[0,T]$, the two notions coincide and the basic result is

Theorem 1 [6-8].  A state x is reachable (and controllable) if and only if x is contained in the subspace of X generated by the vectors

$$\{G, FG, F^2G, \ldots, F^{n-1}G\} \quad .$$

The system $\Sigma$ is said to be completely reachable if and only if $\mathscr{R} = X$, i.e., x is reachable for every $x \in X$.  An immediate consequence of Theorem 1 is

Corollary 1.  $\Sigma$ is completely reachable if and only if the $n \times nm$ matrix

$$\mathscr{C} = [G \,|\, FG \,|\, F^2G \,|\, \ldots \,|\, F^{n-1}G]$$

has rank n.

Many variations on the above theme are possible by changing $\Omega, \mathscr{R}$, T and/or admitting time-varying F and G (see [7] for details).  However, the algebraic result given by Theorem 1 and its corollary forms the cornerstone for the study of almost all questions relating to reachability and controllability of linear

- 4 -

systems. As we shall see below, the same type of algebraic
result can also be obtained for large classes of nonlinear
systems at the expense of a more elaborate mathematical
machinery, further emphasizing the underlying algebraic nature
of dynamical systems.

B. Observability/Constructibility — switching attention
from inputs to outputs, we consider the class of questions
centering upon what information can be deduced about the
system state from the measured output. As in category A,
the basic question comes in two forms, depending upon whether
we wish to determine the initial state $x_0$ from knowledge of
future inputs and outputs (observability) or if we wish to
determine the current state x(T) from knowledge of past
inputs and outputs (constructibility). The linearity of the
situation enables us to consider the case of no input (u = 0)
and, as in the controllability/reachability situation, the
two basic concepts of observability and constructibility
coincide if F and H are constant matrices. The main result
for category B questions is

Theorem 2 [6-8]. A state x ε X is unobservable (uncon-
structible) if and only if x is contained in the kernel of
the matrix

$$\theta = \begin{bmatrix} H \\ \hline HF \\ \hline HF^2 \\ \hline \vdots \\ \hline HF^{n-1} \end{bmatrix} .$$

Note that the basic test implicit in Theorem 2 is given in terms of <u>unobservable</u> states. Thus, any initial state $x_0 \neq 0$ may be uniquely determined from the measured output $y(t)$, $0 \leq t \leq T$, $T > 0$, if and only if $x_0 \notin$ kernel $\Theta$. The important corollary to Theorem 2, characterizing <u>complete</u> <u>observability/</u> <u>constructibility</u> is

   <u>Corollary 2</u>. <u>The system $\Sigma$ is completely observable</u> <u>(constructible) if and only if the matrix $\Theta$ has rank n</u>.

The striking similarity in form between Theorems 1 and 2 suggests a duality between the concepts of reachability and observability. This idea can be made mathematically precise through the identifications

$$F \to F' \quad , \qquad G \to H' \quad ,$$

showing that any result concerning reachability may be transcribed into a dual result about observability, and conversely.

   C.  <u>Realizations/Identification</u> — the basic questions subsumed under categories A and B assume for their statement that the system is given in the so-called <u>state-variable</u> form $\Sigma$. The most important of all system-theoretic questions is that of determining "good" state-variable models given only input/output (experimental) data.

Consider the Laplace transform of the system $\Sigma$ and let $\mathscr{U}(s)$ and $\mathscr{Y}(s)$ denote the transforms of the input and output functions, respectively. It is then easy to see that $\mathscr{U}$ and $\mathscr{Y}$ are linearly related as

$$\mathcal{Y}(s) = W(s)\mathcal{U}(s) \quad ,$$

where

$$W(s) = H(sI - F)^{-1} G \quad ,$$

is called the system <u>transfer</u> <u>matrix</u>. If W(s) is a <u>strictly</u> <u>proper</u> <u>rational</u> <u>matrix</u> (i.e., the elements of W are ratios of relatively prime polynomials with the degree of the numerator less than that of the denominator), then we may expand W(·) in a Laurent series about ∞ obtaining

$$W(s) = \sum_{i=1}^{\infty} A_i s^{-i} \quad .$$

The matrix W(s) or, equivalently, the infinite sequence $\{A_1, A_2, A_3, \ldots\}$ will be called the <u>input/output data</u> (or external description) of the system Σ. We can now state the central problem of mathematical system theory:

<u>The Realization Problem</u>: given the input/output data of a linear system Σ, determine a state-variable model Σ such that

i)  the input/output behavior of the model agrees <u>exactly</u> with the given data and

ii)  the model is completely reachable and completely observable, i.e., the model is <u>canonical</u>.

<u>Remark</u>: Condition (ii), that the model be canonical, is mathematically equivalent to requiring that the dimension of the state space X of the model be minimal. However, for purposes of extension to the nonlinear case, where X may not

even be a vector space, it is preferable to state the require-
ment as given above.

Perhaps surprisingly, the Realization Problem for linear
systems has the following definitive solution.

Theorem 3 [6-8]. For each input/output description of a
system there exists a canonical model $\Sigma$, which is unique up
to a choice of coordinate system in the state space X.

A weak form of the Realization Problem occurs when the
dimension of $\Sigma$ is fixed in advance, perhaps by a priori
engineering or physical considerations, and only some of the
components of F, G and H need to be determined from the input/
output data. This is the so-called parameter identification
problem and is tantamount to not only forcing the system upon
the data (by fixing the dimension of X), but also partially
fixing the coordinate system in X (by demanding that certain
elements of F, G and H remain fixed). Nevertheless, much work
has been done on parameter estimation, especially in the case
where there are uncertainties in the data, a situation which
makes the conceptual approach somewhat easier to accept.

It will be noted that the Realization Problem demands all
of the system input/output data before the internal model $\Sigma$
can be chosen. In principle, this involves an infinite data
string. Of somewhat more practical concern is the case in
which only a finite behavior sequence

$$B_N = \{A_1, A_2, \ldots, A_N\}$$

is available. The construction of a canonical model $\Sigma_N$ from

the sequence $B_N$ constitutes the <u>partial</u> <u>realization</u> problem, which has only recently been definitively resolved. While a precise statement of the main result would take us too far afield, the basic conclusion is that each behavior sequence $B_N$ has a canonical realization $\Sigma_N$, which may be unique (modulo a coordinate change in X), or which may contain a certain number of undetermined parameters. Furthermore, it can be shown that as N increases (more data becomes available), the sequence of canonical realizations $\{\Sigma_N\}$ is nested, i.e., the matrices $F_N$, $G_N$, $H_N$ of the realization $\Sigma_N$, appear as submatrices in the realization $\Sigma_{N+k}$, $k \geq 1$. A complete discussion of these matters is given in [9-10].

In addition to the problems of categories A, B and C, two other broad areas are also usually considered to form part of the general field of mathematical system theory: stability theory and optimization. Generations of work on optimal control theory and stability is by now so well covered in the literature that we shall refrain from a discussion of these areas here. For the interested reader, the sources [11-13] can be recommended.

## 2. Linearization

Given a nonlinear internal model

$$\dot{x} = f(x,u) \quad , \qquad x(0) = x_0 \quad , \qquad (N)$$

$$y(t) = h(x) \quad ,$$

the first temptation in analyzing questions of Type A or B is to linearize the process (N) by choosing some nominal input $\bar{u}(t)$

and generating the corresponding reference trajectory $\bar{x}(t)$.
Such a procedure yields the linearized dynamics

$$\dot{z} = F(t) z + G(t) v \quad , \qquad z(0) = x_0 \quad , \qquad (\Sigma_L)$$

$$w(t) = H(t) z \quad ,$$

where

$$z(t) = x(t) - \bar{x}(t), \quad v(t) = u(t) - \bar{u}(t), \quad w(t) = y(t) - \bar{y}(t),$$

with

$$F(t) = \left(\frac{\partial f}{\partial x}\right) \quad , \quad G(t) = \left(\frac{\partial f}{\partial u}\right) \quad , \quad H(t) = \left(\frac{\partial h}{\partial x}\right) \quad ,$$

with $F(.)$, $G(.)$ and $H(.)$ being evaluated at the pair $(\bar{x}(t), \bar{u}(t))$.
The approach to studying reachability/observability issues is
to now employ the time-varying analogues of Theorems 1 and 2
for the analysis of the system $\Sigma_L$. We would clearly like to
be able to conclude something about the controllability prop-
erties of (N) in a neighborhood of $(\bar{x}, \bar{u})$ by studying the
corresponding properties of $\Sigma_L$. A typical result in this
direction is

Theorem 4 [14]. Let the dynamics $f(x,u)$ be $C^1$ in a
neighborhood U of $(\bar{x}, \bar{u})$. Then the system (N) is locally
controllable if the pair $(F(t), G(t))$ is controllable in U.

Here "local controllability" means that for each $x^*$ in some
neighborhood of $\bar{x}$, there exists a piecewise-continuous control
$u^*(t)$, in some neighborhood of $\bar{u}(t)$, $0 \leq t \leq T$, such that $x(T) = 0$.

The problem with the above linearized results is that they
usually provide only sufficient conditions and are inherently

local in character.  As illustration of this point, consider
the 2nd-order nonlinear problem

$$\dot{x}_1 = -x_1 + u \quad ,$$

$$\dot{x}_2 = -x_1^3 - x_2 \quad ,$$

with $|u(t)| \leq 1$.  Let $\bar{x}(t) = 0$, $\bar{u}(t) = 0$, so that the linearized
system is

$$\dot{x} = Fx + Gu \quad ,$$

with

$$F = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \quad , \qquad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad .$$

The pair $(F,G)$ is not controllable since

$$\mathrm{rank} \left[ \begin{array}{c|c} 1 & -1 \\ 0 & 0 \end{array} \right] = 1 < 2 = n \quad .$$

Nevertheless, it can be shown [14] that each initial state
$(x_1^0, x_2^0)$ near $(0,0)$ can be transferred to the origin in finite
time by a control of the above type.  Thus, the system is
locally controllable although the linearized approximation is
not controllable.

Another obvious defect of linearization is the smoothness
requirement on the dynamics $f(x,u)$ and/or the output function
$h(x)$.  In order for the linearization to make sense, these
functions must be at least continuously differentiable in each
argument.  While many practical processes obey this restriction,

systems with switching points in the dynamics or other types
of discontinuities frequently occur and would be outside the
realm of straightforward linearization techniques.

### 3. Nonlinear Processes

The inadequacies of linearization as outlined in the
preceding section are far from the only reasons why we would
like to develop a system theory for truly nonlinear processes.
Some of the reasons are associated with intrinsic features of
nonlinear dynamical processes, while others are more closely
connected with the methods employed in the study of such pro-
cesses. Let us consider the first of these aspects as it is
somewhat more relevant to the issues raised in this survey.

Among the inherent difficulties associated with nonlinear
processes, which are not present in linear phenomena we may
cite nonuniqueness, singularities and critical dependence on
parameters as features worthy of special attention.

Nonuniqueness — the simple scalar process

$$\dot{x} = ax + bx^3 + u \quad , \tag{1}$$

illustrates the fact that a nonlinear process may have multiple
equilibria, even in the presence of no control input ($u = 0$).
In the event a feedback law

$$u = \phi(x) \quad ,$$

is employed, the closed-loop dynamics

$$\dot{x} = ax + bx^3 + \phi(x)$$

may have an infinite (or even uncountable) number of equilibria, depending upon the form of $\phi$. Clearly, this situation is in stark contrast to the linear case where only the equilibrium $x = 0$ can generically occur. Furthermore, no linearized version of (1) can possibly capture the global structure of the system equilibria manifold as a function of a and b.

Singularities — the solutions of many nonlinear systems may develop singularities, even though the systems themselves have smooth coefficients. The simple two-point boundary value problem

$$\ddot{x} + x\dot{x} = 0 \quad ,$$

$$x(0) = 0 \quad , \quad x(T) = 0 \quad ,$$

possesses no solutions without singularities for any $T > 0$.

In a more system-theoretic direction, it can be shown [15] that the system

$$\dot{x}_1 = x_2 \quad , \qquad\qquad\qquad x_1(0) = 1 \quad , \qquad\qquad (2)$$

$$\dot{x}_2 = -x_1 + u(t) x_1 \quad , \qquad\qquad x_2(0) = 0 \quad ,$$

with $|u(t)| \leq \varepsilon \ll 1$, has a reachable set from $x_0$ which is homeomorphic to a disk for T small, but encircles the origin for T large (see Fig. 1).

- 13 -



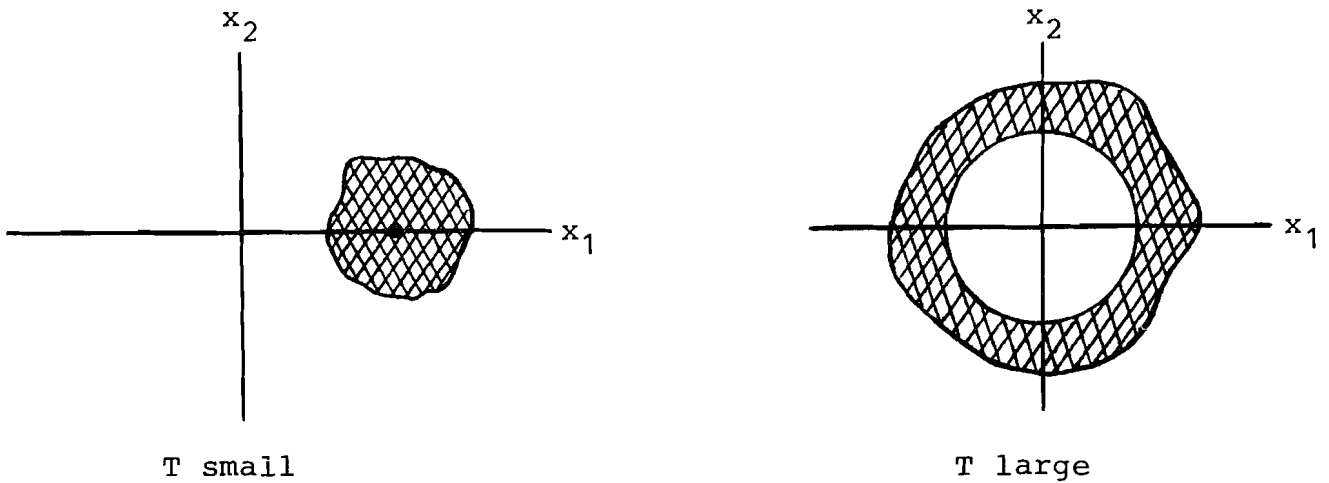T small                                    T large

Figure 1.  The Reachable Set for the System (2)


The situation can be even worse than this as some nonlinear

systems have a reachable set which is not even simply-connected

[15].  In the linear case, of course, Theorem 1 shows that the

reachable set is a subspace of $R^n$, hence, not only simply-

connected but even convex.  Again, no linearized version of the

system (2) can hope to capture the global structure of the

reachable set.

The simple bilinear system

$$\dot{x} = xu \quad ,$$

also shows that a state may not be reachable from the origin

with bounded control.  Thus, a more appropriate state space for

this problem is the "punctured" region $R^n - \{0\}$, rather than

$R^n$ itself.  In general, the "natural" state space for a non-

linear process is no longer the familiar vector space (or

module) of the linear theory, but a much more complicated

mathematical object, usually some type of manifold in a

Euclidean space of high dimension.  For instance, if the

system is multilinear then the state space has the structure
of an abelian variety (= algebraic manifold) [16].   Such facts
account  for the need to employ much more sophisticated ma-
chinery than simple linear algebra to study the structure of
nonlinear processes.

Critical Dependence on Parameters — for the linear dynamical
system

$$\dot{x} = Fx \quad ,$$

there are no parametric changes in the elements of F which can
cause the system to have more than a single solution curve $x(t)$.
However, this is far from the case for nonlinear processes.  For
example, consider the system

$$\ddot{x} + \tfrac{1}{2}\lambda e^{x} = 0 \quad ,$$

$$x(0) = x(1) = 0 \quad .$$

For $\lambda > \beta$ (a certain positive number), the system has no smooth
solution.  For $\lambda = \beta$ there is exactly one smooth solution,
while for $0 < \lambda < \beta$ there are two solutions.  Thus, $\beta$ is a bifur-
cation point in the parameter space at which the character of
the solution set changes radically.

To illustrate another point, consider the system

$$\dot{x}_1 = \mu x_1 + x_2 - x_1^{\,3} \quad ,$$

$$\dot{x}_2 = -x_1 \quad , \qquad\qquad -1 \le \mu \le 1 \quad .$$

For each $\mu$, $-1 \le \mu \le 0$, all solutions tend asymptotically to zero

as $t \to \infty$. As $\mu$ crosses $0$, the system has a unique periodic

solution $p(\mu)$ and the origin becomes a source. For all $\mu$,

$0 < \mu \leq 1$, every nontrivial solution tends to $p(\mu)$ as $t \to \infty$.

Thus, $\mu = 0$ is a bifurcation point at which the equilibrium

at the origin changes suddenly from a sink to a source and

a limit cycle $p(\mu)$ is created. This so-called "Hopf bifur-

cation" is a consequence of the system nonlinearity and has

no counterpart in linear problems.

Finally, consider the equilibria of the nonlinear system

$$\dot{x} = f(x,\alpha) \quad ,$$

where $\alpha$ is an m-dimensional vector of parameters. The equi-

libria $x^*$ for which $f(x^*,\alpha) = 0$ depend upon $\alpha$ and we can define

a map

$$\begin{aligned} \chi : A &\longrightarrow X \quad , \\ \alpha &\longmapsto x^*(\alpha) \end{aligned}$$

where $A \subset R^m$, $X \subset R^n$. Under appropriate hypotheses on the function

$f$, properties of the map $\chi$ can be characterized using Thom's

theory of catastrophes. In particular, it is of interest to

categorize those submanifolds of A for which the map $\chi$ is dis-

continuous, the so-called "catastrophe" manifold. Again, if

$f$ is linear the map $\chi$ is continuous and there is no interesting

structure to analyze. Thus, no linearized version of the problem

will suffice to study the geometry of the equilibrium manifold.

The above examples provide convincing evidence of the need

to develop a nonlinear system theory capable of handling the

same broad array of questions so successfully dealt with by the

linear theory. In succeeding sections, we present some steps in this direction. As will become evident, almost everything remains to be done to complete such a program despite the impressive advances of recent years.

## 4. Reachability and Controllability

### Smooth Systems

Certainly the area in which most progress has been made in understanding the system-theoretic behavior of nonlinear processes is in the effective characterization of reachable sets and in the determination of algebraic criteria for complete reachability. Since the mathematical apparatus involved goes somewhat beyond the elementary linear algebra which suffices for the study of linear systems, we make the following definitions as originally given in [17].

Consider the nonlinear system

$$\dot{x} = f(x,u) \quad , \qquad x(0) = x_0 \quad , \qquad (N)$$

$$y(t) = h(x) \quad ,$$

where $u \in \Omega \subset R^m$, $x \in M$, a $C^\infty$-connected manifold of dimension n and f and h are $C^\infty$ functions of their arguments. To simplify notation, it is assumed that M admits globally defined coordinates $x = (x_1,\ldots,x_n)'$, allowing us to identify the points of M with their coordinate representations and to describe the control system (N) in the usual engineering form above. We also assume that (N) is complete, i.e., for every bounded measurable control u(t) and every $x_0 \in M$, there exists

a solution of $\dot{x} = f(x,u)$ satisfying $x(0) = x_0$, $x(t) \in M$ for all real t.

Definition 1. Given a point $x^* \in M$, we say that $x^*$ is reachable from $x_0$ at T if there exists a bounded measurable control $u(t)$, satisfying $u(t) \in \Omega$, such that the system trajectory satisfies $x(0) = x_0$, $x(T) = x^*$, $x(t) \in M$, $0 \leq t \leq T$. The set of states reachable from $x_0$ is denoted as

$$\mathcal{R}(x_0) = \bigcup_{0 \leq T < \infty} \{x : x \text{ reachable from } x_0 \text{ at time } T\} \; .$$

We say (N) is reachable at $x_0$ if $R(x_0) = M$ and reachable if $\mathcal{R}(x) = M$ for all $x \in M$.

Since it may be necessary to either travel a long distance or a great time to reach points near $x_0$, the property of reachability from $x_0$ is not always of practical use. This fact leads to a local version of reachability.

Definition 2. (N) is locally reachable at $x_0$ if for every neighborhood U of $x_0$, $\mathcal{R}(x_0)$ is also a neighborhood of $x_0$ with the trajectory from $x_0$ to $\mathcal{R}(x_0)$ lying entirely within U. The system (N) is locally reachable if it is locally reachable for every $x \in M$.

The reachability concept detailed in Definition 1 is not symmetric: $x^*$ may be reachable from $x_0$ but not conversely (in contrast to the situation for autonomous linear systems). To remedy this situation, we need a weaker notion of reachability. This is provided by

Definition 3. Two states $x^*$ and $\bar{x}$ are <u>weakly</u> <u>reachable</u> from each other if and only if there exist states $x^0, x^1, \ldots, x^k$ such that $x^0 = x^*$, $x^k = \bar{x}$ and either $x^i$ is reachable from $x^{i-1}$ or $x^{i-1}$ is reachable from $x^i$, $i = 1, 2, \ldots, k$. The system (N) is said to be <u>weakly</u> <u>reachable</u> if it is weakly reachable from every $x \in M$. Since weak reachability is a global concept like reachability, we can define a local version of it in correspondence to Definition 2.

Among the various reachability concepts, we have the following chain of implications

locally reachable $\implies$ reachable
$\Downarrow$                         $\Downarrow$
locally weakly reachable $\implies$ weakly reachable

For autonomous linear systems it can be shown that all four of the above notions coincide.

The advantage of local weak reachability over the other concepts defined above is that it lends itself to a simple algebraic test. For this, however, we need a few additional notions.

Definition 4. Let $p(x)$, $q(x)$ be two $C^\infty$ vector fields on M. Then the <u>Jacobi</u> <u>bracket</u> of p and q, denoted $[p,q]$ is given by

$$[p,q](x) = \left(\frac{\partial q}{\partial x}\right) p - \left(\frac{\partial p}{\partial x}\right) q \bigg|_x \ .$$

The set of all $C^\infty$ vector fields on M is an infinite-dimensional vector space denoted by X(M) and becomes a Lie algebra under the the multiplication defined by the Jacobi bracket.

Each <u>constant</u> control $u \epsilon \Omega$ defines a vector field $f(x,u) \epsilon X(M)$. We let $\mathcal{F}_0$ denote the subset of all such vector fields, i.e., $\mathcal{F}_0$ is the set of all vector fields generated from $f(x,\cdot)$ through use of constant controls. $\mathcal{F}$ denotes the smallest subalgebra of $X(M)$ containing $\mathcal{F}_0$. The elements of $\mathcal{F}$ are linear combinations of elements of the form

$$[f^1[f^2\ldots[f^i,f^{i+1}]\ldots]] \quad ,$$

where $f^i(x) = f(x,u^i)$ for some constant $u^i \epsilon \Omega$. We let $\mathcal{F}(x)$ be the space of tangent vectors spanned by the vector fields of $\mathcal{F}$ at x.

<u>Definition 5.</u> (N) is said to satisfy the <u>reachability rank condition at</u> $\underline{x}_0$ if the dimension of $\mathcal{F}(x_0)$ is n. If this is true for every $x \epsilon M$, then (N) satisfies the <u>reachability rank condition</u>.

The following theorem illustrates the importance of the reachability rank condition.

<u>Theorem 5</u> [17]. <u>If (N) satisfies the reachability rank condition at $x_0$, then (N) is weakly locally reachable at $x_0$.</u> For $C^\infty$-systems, the converse is not quite true, but we do have

<u>Theorem 6</u> [17]. <u>If (N) is locally weakly reachable then the reachability rank condition is satisfied on an open dense subset of M (i.e., the rank condition is satisfied generically).</u>

In the event we strengthen the smoothness requirement on (N) from $C^\infty$ to analytic, we can strengthen Theorems 5 and 6 to

Theorem 7 [17].  If (N) is analytic then (N) is weakly reachable if and only if it is locally weakly reachable if and only if the reachability rank condition is satisfied.

The simplest illustration of the use of these results is to recapture the linear result of Theorem 1.  In this case

$$\mathscr{F}_0 = \{Fx + Gu : u \varepsilon \Omega\}$$

so the Lie algebra is generated by the vector fields $\{Fx, g_1, g_2, \ldots, g_m\}$, where $g_i$ denotes the ith column of G regarded as a constant vector field.  Computing brackets yields

$$[Fx, g_j] = -Fg_j \quad , \qquad\qquad [g_i, g_j] = 0 \quad ,$$

$$[Fx, [Fx, g_j]] = F^2 g_j \quad , \qquad [g_i, [Fx, g_j]] = 0 \quad , \qquad \text{etc.}$$

The Cayley-Hamilton Theorem implies that $\mathscr{F}$ is spanned by the vector fields Fx and the constant vector fields $F^i g_j$, $i = 0, 1, \ldots, n-1$, $j = 1, 2, \ldots, m$.  Thus, in this context the reachability rank condition reduces to the condition of Theorem 1, namely, (N) is locally reachable if and only if

$$\text{rank } [G|FG|F^2G|\ldots|F^{n-1}G] = n \quad .$$

However, for linear systems local reachability and reachability are equivalent, so the usual results are obtained.

The practical problem with applying the preceding results is that we have no nonlinear version of the Cayley-Hamilton Theorem insuring that the test for complete reachability can be concluded in a finite number of steps.  In principle, we

could compute bracket after bracket in the Lie algebra gener-
ated by the $\{f^i\}$ with no assurance that the next bracket might
not yield a vector field linearly independent of those already
computed.

In order to rule out the above type of behavior, we intro-
duce the following definition.

Definition 6. A set of vector fields $\left\{f^i\right\}_{i=1}^{r}$ is called
involutive if there exist constants $\gamma_{ijk}$ such that

$$[f^i, f^j](x) = \sum_{k=1}^{r} \gamma_{ijk} f^k(x) \quad .$$

The property of being involutive is a necessary condition in
order to be able to "integrate" the vector fields $f^1, \ldots, f^r$
to obtain a solution manifold. The following theorem of
Frobenius shows that this property is (with mild regularity
assumptions) also sufficient to assert the existence of
maximal solutions.

Theorem 8 [18]. Let $\left\{f^i\right\}_{i=1}^{r}$ be an involutive collection
of vector fields which are

a) analytic on an analytic manifold M. Then given any
point $x_0 \varepsilon$ M , there exists a maximal submanifold N containing
$x_0$ such that $\{f^i\}$ spans the tangent space of N at each point
of N.

b) $C^\infty$ on a $C^\infty$ manifold M with the dimension of the span
of $\{f^i\}$ constant on M. Then given any point $x_0 \varepsilon$ M, there
exists a maximal submanifold N containing $x_0$ such that $\{f^i\}$
spans the tangent space of N at each point of N.

As an illustration of Frobenius' Theorem, consider the analytic vector fields in $R^3$

$$f^1(x) = \begin{bmatrix} 0 \\ x_3 \\ -x_2 \end{bmatrix} \quad , \qquad f^2(x) = \begin{bmatrix} -x_3 \\ 0 \\ x_1 \end{bmatrix} \quad , \qquad f^3(x) = \begin{bmatrix} x_2 \\ -x_1 \\ 0 \end{bmatrix} .$$

It is easily verified that this collection is involutive and if we look at any point $x \in R^3$ then we can integrate the distribution through that point. For instance, if $x = \frac{1}{3}(\sqrt{3}, \sqrt{3}, \sqrt{3})$, then we obtain the set

$$N = \{x : \|x\| = 1\}$$

as the corresponding integral manifold. In fact, in this example, the vectors $f^1$, $f^2$, $f^3$ are tangent to the spherical shell N at each point. Additional details on this example are provided in [18].

In terms of the Frobenius Theorem, the problem of complete reachability for an involutive system of vector fields may be re-stated: does the maximal submanifold $N = M$? In order to answer this question, it is necessary to have a more explicit characterization of the submanifold N. This is provided by a theorem of Chow, which also provides the underpinning for our earlier results, Theorems 5-7. But first a bit of additional notation.

Given a vector field f on M, for each t exp tf defines a map of $M \to M$, which is the mapping produced by the flow on M defined by the differential equation $\dot{x} = f(x)$. We denote by diff (M) the group of diffeomorphisms of M and let $\{\exp \{f^i\}\}_G$

be the smallest subgroup of diff (M) which contains $\exp tf$ for all $f \in \{f^i\}$. Finally, $\{f^i\}_{LA}$ denotes the Lie algebra of vector fields generated by $\{f^i\}$ under the Jacobi bracket multiplication defined above. We are now in a position to state the following control-theoretic version of Chow's Theorem.

Theorem 9 [18]. Let $\left\{f^i(x)\right\}_{i=1}^r$ be a collection of vector fields such that $\{f^i(x)\}_{LA}$ is

a) analytic on an analytic manifold M. Then given any $x_0 \in M$, there exists a maximal submanifold $N \subset M$ containing $x_0$ such that

$$\left\{\exp\{f^i\}\right\}_G x_0 = \left\{\exp\{f^i\}_{LA}\right\}_G x_0 = N \quad ;$$

b) $C^\infty$ on a $C^\infty$ manifold M with $\dim\left(\text{span}\{f^i(x)\}_{LA}\right)$ constant on M. Then given any point $x_0 \in M$, there exists a maximal submanifold $N \subset M$ containing $x_0$ such that

$$\left\{\exp\{f^i\}\right\}_G x_0 = \left\{\exp\{f^i\}_{LA}\right\}_G x_0 = N \quad .$$

## Linear-Analytic Systems

The conclusions of Chow's Theorem enable us to effectively resolve the reachability problem for systems of the form

$$\dot{x} = \sum_{i=1}^r u_i f_i(x) \quad , \qquad x(0) = x_0 \quad .$$

However, in applications we are often confronted with systems of the form

$$\dot{x} = p(x) + \sum_{i=1}^r u_i(t) g^i(x) \quad , \qquad x(0) = x_0 \quad . \tag{3}$$

In this situation, Chow's Theorem has the serious drawback that it does not distinguish between positive and negative time. Thus, the submanifold N may include points which can only be reached by passing backward along the vector field p(x). This means that the reachable set will, in general, only be a proper subset of N.

If we let (exp tp)$(x_0)$ denote the solution to (3) at time t corresponding to all $u_i \equiv 0$, while $\mathscr{R}(t,x_0)$ denotes the reachable set at time t, then the problem of <u>local reach-ability</u> is to find necessary and sufficient conditions that (exp tp)$(x_0)$ $\varepsilon$ interior $\mathscr{R}(t,x_0)$ for all $t > 0$. Denoting (ad X, Y) = [X,Y], $(ad^{k+1}X,Y)$ = [X,$(ad^k X,Y)$], the basic known results on this problem are contained in

<u>Theorem 10</u> [19].

a) <u>A necessary and sufficient condition that</u> <u>interior $\mathscr{R}(t,x_0) \neq \emptyset$ for all $t > 0$ is that dim $\left( \{p,g^i\}_{LA} \right)\left( x_0 \right) = n$.</u>

b) <u>A sufficient condition that (exp tp)$(x_0)$ $\varepsilon$ interior $\mathscr{R}(t,x_0)$</u> <u>for all $t > 0$ is that</u>

$$\{ (ad^j p,g^i) : j = 0,1,2,\ldots;\ i = 1,2,\ldots,r\}$$

<u>contain n linearly independent elements.</u>

<u>Remark</u>: The condition (b) of Theorem 10 is also necessary in the case $n = 2$. In general, though, more stringent hypotheses are required for the "rank condition" to be necessary.

To illustrate the application of the foregoing results, consider the dynamical system

$$
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = - \begin{bmatrix} x_2 \\ \dfrac{\sin x_1}{x_3} \\ 0 \end{bmatrix} + u \begin{bmatrix} 0 \\ \dfrac{-2x_2}{x_3} \\ 1 \end{bmatrix} = p(x) + ug(x) \quad .
$$

Computing the Lie brackets, we have

$$
\begin{bmatrix} p,g \end{bmatrix} = \begin{bmatrix} \dfrac{2x_2}{x_3} \\ \dfrac{\sin x_1}{x_3^2} - \dfrac{2x_2}{x_3} \\ 0 \end{bmatrix} \quad ,
$$

so that p,g and [p,g] span $R^3$ unless $x_1 = 0$ or $\pi$ or $x_2 = 0$. That is, the system satisfies the reachability rank condition for all non-zero $x_0$.

Let us return now to the problem of local reachability. If we assume that the origin is an equilibrium point for the vector field $p(x)$, i.e., $p(0) = 0$, and if we measure the system to be in some state q at a future time $t_1$, then we can consider the local reachability problem to consist in determining the existence of a stabilizing control which would drive the trajectory of the system $x(t)$ in the "direction" -q.

To be more explicit, consider the system

$$
\dot{x} = p(x) + u(t) g(x) \quad , \tag{4}
$$

where $|u(t)| \leq 1$. Further, assume that

$$\text{dim span } \{ (ad^k p, g) : k = 0, 1, \ldots \}(0) = n$$

so that a stabilizing control law exists, at least locally (Theorem 10 (b)). The problem in the construction of such a law is that the directions that are "instantaneously" possible are $p(q) + \mu g(q)$, $-1 \le \mu \le 1$, and $-q$ need not be among these directions. Let us write $q$ as

$$q = \sum_{i=1}^{n} \alpha_i (ad^{i-1} p, g)(0) \quad .$$

Then if we can generate the directions $\pm (ad^j p, g)(0)$ via compositions of solutions of (4) with controls $|u| \le 1$, it follows that we can generate the direction $-q$.

A specific illustration of how to construct the locally stabilizing law is the following taken from [19]. Let $n = 3$ and define

$$q(\varepsilon) x = q_2^{e_3} \left( |\alpha_3| \varepsilon \right) \circ q_1^{e_2} \left( |\alpha_2| \varepsilon \right) \circ q_0^{e_1} \left( |\alpha_1| \varepsilon \right) x \quad , \tag{5}$$

where

$$e_j = \begin{cases} + & \text{if} \quad \alpha_j < 0 \quad , \\ \\ - & \text{if} \quad \alpha_j > 0 \quad , \end{cases}$$

and

$$q_0^{\pm}(s)(x) = (\exp s(p \pm g))(x) \quad ,$$

$$q_1^{\pm}(s)(x) = (\exp \sqrt{s}(p \mp g)) \circ (\exp \sqrt{s}(p \pm g))(x) \quad ,$$

$$= (\exp (2\sqrt{s}\, p \pm s[p,g] + 0(s)))(x) \quad ,$$

$$q_2^{\pm}(s)(x) = (\exp (2s^{\frac{1}{4}}p \pm s(ad^2 p,g) + 0(s)))(x) \quad .$$

These flows are chosen so that if $p(0) = 0$ and $|p(x)| \le c|x|$,

then $\left. \dfrac{dq^{\pm}}{ds}(s)(x) \right|_{s=0} = \pm(ad^j p,g)(x)$.

Thus, if $x = \sum\limits_{i=1}^{3} \alpha_i (ad^{i-1}p,g)(0)$, then

$$q(s)\, x - x = (s + s^{\frac{1}{2}} + (s/2)^{\frac{1}{4}})\, p(x) - s \sum_{i=1}^{3} \alpha_i (ad^{i-1}p,g)(x) + 0(s) \quad .$$

Hence, if $x$ is near 0 and $s$ is sufficiently small, $q(s)\, x - x =$ $-sx + 0(s)$ and the above formula shows how to choose a control over the time interval $[0, \sum\limits_{i=1}^{3} |\alpha_i| s]$ so as to move the state essentially in the direction $-x$, i.e., toward the origin. Summarizing, the steps in the process are

     i)    measure the state $x$;

     ii)   express $x = \sum\limits_{i=1}^{3} \alpha_i (ad^{i-1}p,g)(x)$ ;

     iii)   use (5) to determine an "open-loop" control $u(t,x)$ on the interval $0 \le t \le \sum\limits_{i=1}^{3} |\alpha_i| s$ ;

     iv)   remeasure the state and repeat the process.

(Note: Even though the measured state $x$ is used to compute the control, the law $u$ is still open-loop since no state over the

interval $0 \leq t \leq \sum_{i=1}^{3} |\alpha_i|$ s is measured). The formulae for the general case of the above result are given in [19] along with a report on the convergence of the algorithm sketched in steps (i) - (iii) above.

The formulae given above for generating $\pm(ad^k p, g)(x)$ are but one of many possible schemes. The question (as yet unanswered) arises as to whether a different scheme can be derived in which the terms $0(s)$ are actually <u>insignificant</u> when compared to $\pm s(ad^k p, g)$ for large k. (In the formulae given above the term $0(s)$ in $q_k^{\pm}(s)(x)$ is of the form $(s^{1+\frac{1}{2}k})w$, for some vector field w in $\{(ad^i p, g) : i = 0, 1, \ldots\}_{LA}$. Numerically, this is <u>not</u> insignificant when compared to $\pm s(ad^k p, g)$ for k large).

Before moving on to results for important special classes of nonlinear systems, it is of value to cite the works [20-22] for additional reachability results. Of special note is [20] in which global results are obtained for systems in which the Lie algebra $\{p, g^i\}_{LA}$ is not necessarily finite-dimensional.

### Bilinear Systems

By far the most detailed and explicit results for the reachability of nonlinear systems are those developed for bilinear processes. Bilinear systems are characterized by the equations

$$\dot{x} = Fx + Gu + \sum_{i=1}^{m} N_i x u_i(t) , \qquad (6)$$

where F and $N_i$ are n × n real matrices and G is an n × m real matrix.

There are a number of theoretical and practical motivations for the study of bilinear processes, which are well-detailed in [23]. For now we only note that the type of nonlinearity (multiplicative) makes the system structure in some sense "closest" to the linear case. This fact enables us to employ many of the techniques and procedures already set up for linear systems.

For studying the reachability properties of (6), we consider the case $G = 0$ (homogeneous-in-the-state systems) since the inhomogeneous case $(G \neq 0)$ is in a somewhat less settled state. However, it should be noted that by adding extra components to the state and/or to the control, and constraining them to be equal to 1, an inhomogeneous bilinear system may be formally studied as a homogeneous-in-the-state system.

Given a homogeneous-in-the-state system

$$\dot{x} = \left( F + \sum_{i=1}^{m} N_i u_i(t) \right) x \ , \qquad x(0) = x_0 \ , \qquad (7)$$

we may write the solution as $x(t) = X(t) x_0$, where $X(t) \in GL(n)$, the nonsingular $n \times n$ real matrices. Thus, the reachability properties of (7) are directly related to those of the system

$$\dot{X} = FX + \sum_{i=1}^{m} N_i X u_i(t) \ , \qquad X(0) = I \ . \qquad (8)$$

Here the system state space is taken to be $M = GL(n)$. To study reachability properties of (8), we need the notion of a matrix Lie algebra.

Definition 7. Given two n × n matrices A and B, their Lie product is defined as

$$[A,B] \doteq AB - BA \quad .$$

A Lie algebra of n × n matrices is a subspace of n × n matrices closed under the Lie product operation.

Let $\mathscr{L}$ denote the Lie algebra generated by the matrices $\{F,N_1,N_2,\ldots,N_m\}$ and let $\mathscr{R}(t,I)$ denote the reachable set for (8) at time t. Then the main reachability result for homogeneous-in-the-state bilinear systems is

Theorem 11 [24]. For the system (8), if GL(n)($\mathscr{L}$) is compact then

a)  $\bigcup_{t>0} \mathscr{R}(t,I) = GL(n)(\mathscr{L})$ ;

b)  there exists a $0 < T < \infty$ such that

$$\mathscr{R}(T,I) = \bigcup_{t\geq 0} \mathscr{R}(t,I) = GL(n)(\mathscr{L}).$$

Here

$$GL(n)(\mathscr{L}) \doteq \left\{ \Gamma \in GL(n) : \Gamma = e^{A_1} e^{A_2} \ldots e^{A_m}, A_i \in \mathscr{L} \right\}$$

$$i = 1,\ldots,m \quad , \quad m = 1,2,\ldots$$

In short, Theorem 11 says that the reachable set for (8) from the identity is GL(n)($\mathscr{L}$) and that all points that can be reached will be attained after some finite time T.

For the inhomogeneous system (6), a convenient sufficient condition for controllability is given by the following result.

Theorem 12 [25]. The inhomogeneous system (6) is controllable from the state $x_0$ if the sequence of vectors $\{s_0^1, \ldots, s_0^m, s_1^1, \ldots, s_{n-1}^1, \ldots, s_{n-1}^m\}$ contains n linearly independent elements, where

$$s_k^i = F^k g_i + (\text{ad}_F^k N_i) x_0 \quad , \qquad k = 0, 1, \ldots, n-1 \quad ,$$
$$i = 1, 2, \ldots, m \quad ,$$

$$\text{ad}_F^k N_i = [F, \text{ad}_F^{k-1} N_i] \quad , \qquad \text{ad}_F^0 N_i = N_i \quad ,$$

$g_i$ = ith column of G.

An alternate approach to the study of controllability of bilinear processes is to study the equilibrium points of (6). Let $\bar{u}$ be a constant control in the unit hypercube H. Then the equilibrium point $x^*(\bar{u})$ is the solution of the equation

$$Fx + Nx\bar{u} + G\bar{u} = 0 \quad .$$

(Note: Here we adopt the more compact notation $\sum_{i=1}^m N_i x u_i \doteq N x u$ .)

Let us assume that whenever $F + N'\bar{u}$ is singular, $G\bar{u}$ is not in its range. Then the expression

$$x^*(\bar{u}) = -(F + N'\bar{u})^{-1} G\bar{u} \tag{9}$$

is the form of all possible equilibrium points, and as $\bar{u}$ ranges over H, (9) describes the equilibrium set.

A sufficient condition for the controllability of (6) is now given by

Theorem 13 [14]. The bilinear system (6) is completely controllable using piecewise-continuous inputs if

a) there exist constant controls $u^+$ and $u^-$ in H such that $Re[\lambda_i(F + N'u^+)] > 0$ and $Re[\lambda_i(F + N'u^-)] < 0$, with $x^*(u^+)$ and $x^*(u^-)$ contained in a connected subset of the equilibrium set and

b) for each $x^*(\bar{u})$, there exists a $v \in R^m$ such that the pair $\{F + N'\bar{u}, [Nx^*(\bar{u}) + G]v\}$ is controllable.

A more thorough investigation of the above criterion, together with many auxiliary results and examples is given in the book [23].

Important properties of the reachable set for a compact control set are that it be convex and closed, regardless of the initial state. These properties are important for understanding the time-optimal control problem and for generating computational algorithms for determining optimal controls. For bilinear systems the reachable set is usually not convex (or even closed unless the control set is both compact and convex).

Since the general case is not yet settled, we consider the special case of (7) when the matrices $N_i$ have rank 1, i.e., we can write $N_i = b_i c_i'$, where $b_i$ and $c_i$ are n-dimensional vectors. The first convexity result involves the case of small t.

Theorem 14 [15]. Let $x_0$ be given and assume that $c_i'x_0 \neq 0$, $i = 1,2,\ldots,m$. Then there exists a $T > 0$ such that for each t, $0 \leq t \leq T$, the reachable set for (7) is convex for bounded controls $u_i(t)$.

In order to "globalize" this result to the case $T = \infty$, additional conditions on $F$, $b_i$ and $c_i$ are needed.

Theorem 15 [15]. <u>Suppose each component of $c_i$ is non-negative and that for all $t > 0$ the matrix</u> $F + \sum_{i=1}^{m} u_i(t) \, b_i c_i'$ <u>has non-negative off-diagonal entries. Then the reachable set at time t is convex for $t > 0$ for bounded controls $u_i(t)$.</u>

Other reachability/controllability results for nonlinear systems have been reported, but space precludes their inclusion. Specifically, we refer to [26] for global controllability results for perturbed linear systems. In a highly algebraic treatment, the case of systems governed by discrete-time polynomial dynamics is covered in detail in [27].

## 5. Observability and Constructability

The general notion of observability can be stated in the following terms: given a canonical model (N) of an input/output map f, and an input function $u \, \varepsilon \, \Omega$ applied after $t = t_0$, determine the state $x_0$ of (N) at $t = t_0$ from knowledge of the output function $y(t)$, $t_0 \leq t \leq T$. Another way of looking at the question is to ask if every possible pair of initial states $x_0, x_0'$ can be distinguished by every admissible input $u \, \varepsilon \, \Omega$.

There are several delicate issues which arise in the theory of nonlinear observability which are masked in the linear case discussed earlier. Let us consider two of the technical considerations.

i) <u>choice of inputs</u> — in the linear case, it is easy to show that if <u>any</u> input distinguishes points then <u>every</u> input does. So, it suffices to consider the case $u \equiv 0$. However, for nonlinear systems this is not the case. There may be certain

inputs which do not separate points. Thus, we must be critically aware of the observability definition employed.

ii) <u>length of observation</u> — for continuous-time linear systems, observing the output $y(t)$ over <u>any</u> interval $t_0 \leq t < t_0 + \varepsilon$, $\varepsilon$ arbitrary, suffices to separate points for a completely observable system. However, it may be necessary to observe $y(t)$ over a long, even infinite, interval in order to determine $x_0$ for a nonlinear process. Thus, it is desirable to modify the <u>global</u> concept of observability by introducing a <u>local</u> version involving only the separation of points "near" $x_0$ in either a spatial or temporal sense.

In what follows, we shall adopt definitions to deal with the foregoing difficulties, motivated by a desire to obtain a simple algebraic test for observability analogous to that given earlier for controllability.

We consider the system

$$\dot{x} = f(x,u) \quad , \qquad x(t_0) = x_0 \quad , \qquad \text{(N)}$$

$$y(t) = h(x) \quad ,$$

as given in Section 4.

<u>Definition 8</u>. A pair of points $x^0$, $x^1 \varepsilon M$ are termed <u>indistinguishable</u> if the systems $(N, x^0)$ and $(N, x^1)$ realize the same input/output map, i.e., under the same input $u \varepsilon \Omega$, the system (N) produces the same output $y(t)$ for the initial states $x^0$ and $x^1$. The system (N) is termed <u>observable</u> if for all $x \varepsilon M$, the only point indistinguishable from $x$ is $x$ itself.

Remark. Observability of (N) does not imply that _every_ input in $\Omega$ distinguishes all points of M. This is true, however, if the output y is a sum of a function of the initial state and a function of the input, as in the linear case.

Since observability is a global concept, we localize the concept with the following definitions.

Definition 9. (N) is _locally_ _observable_ at $x^0 \varepsilon$ M if for every open neighborhood U of $x^0$, the set of points indistinguishable from $x^0$ consists of $x^0$ itself. (N) is _locally_ _observable_ if it is locally observable for every $x \varepsilon$ M.

Practical considerations suggest that it may be sufficient only to distinguish points which are near to $x^0$, leaving open the possibility of $x^0$ being equivalent to states $x^1$ which are far removed. This heuristic idea motivates

Definition 10. (N) is _weakly_ _observable_ at $x^0$ if there exists an open neighborhood U of $x^0$ such that the only point in U which is indistinguishable from $x^0$ is $x^0$ itself. The system (N) is _weakly_ _observable_ if it is weakly observable at every $x \varepsilon$ M.

Again, weak observability may require that we travel far from U in order to distinguish the points of U. The following definition deals with this problem.

Definition 11. (N) is _locally_ _weakly_ _observable_ at $x^0$ if there exists an open neighborhood U of $x^0$ such that for every open neighborhood V of $x^0$ contained in U, we have that the set of points indistinguishable from $x^0$ in V is $x^0$ itself. The system (N) is _locally_ _weakly_ _observable_ if it is locally weakly observable for all $x \varepsilon$ M.

As for controllability, the following diagram of implications exists:

(N) locally observable $\Longrightarrow$ (N) observable

$\Downarrow$ $\Downarrow$

(N) locally weakly observable $\Longrightarrow$ (N) weakly observable
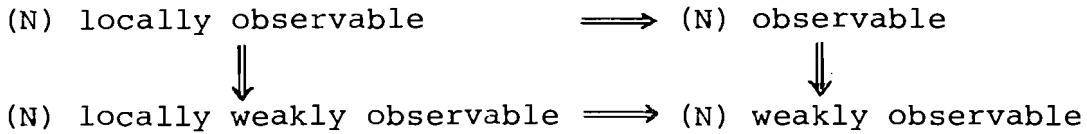
For linear systems, all four concepts coincide.

As noted in Section 1, reachability and observability are dual concepts in the precise meaning of vector space duality. In order to generalize this result to the manifold setting, additional machinery is required. In essence, we shall employ the duality between the space X(M) of vector fields on a manifold M and the space $X^*$(M) of one-forms on M. This duality, coupled with the role X(M) played in the controllability situation, strongly suggests that the space of one-forms $X^*$(M) will be the appropriate vehicle for the study of nonlinear observability.

Definition 12. Let $\phi$(x) be a $C^\infty$ function on M with q an element of X(M). Then the Lie derivative of $\phi$ (in the direction q), $L_q(\phi)$, is defined as

$$L_q(\phi)(x) = \frac{\partial \phi}{\partial x}(x) \, q(x) \quad .$$

(Note that the gradient $d\phi = \frac{\partial \phi}{\partial x}$ is an n-dimensional row vector.)

Now let $\mathcal{G}_0$ denote the subset of $C^\infty$(M) consisting of the functions $h_1(x), h_2(x), \ldots, h_p(x)$, i.e., the components of the observation vector function h(x). Further, we let $\mathcal{G}$ denote the smallest vector space generated by $\mathcal{G}_0$ and elements obtained from $\mathcal{G}_0$ by Lie differentiation in the direction of elements of $\mathcal{F}_0$ (recall: $\mathcal{F}_0$ is the set of all vector fields generated from

f(x,·) using constant controls). A typical element of $\mathscr{G}$ is a finite linear combination of elements of the form

$$L_{f^1}(\ldots(L_{f^k}(h_i))\ldots) \quad ,$$

where $f^i(x) = f(x,u^i)$ for some constant $u^i \varepsilon \Omega$. It is easily verified that $\mathscr{G}$ is closed under Lie differentiation by elements of $\mathscr{F}$ also.

Define $X^*(M)$ as the real vector space of one-forms on M, i.e., all finite $C^\infty(M)$ linear combinations of gradients of elements of $C^\infty(M)$. Further, let $d\mathscr{G}_0 = \{d\phi : \phi \varepsilon \mathscr{G}_0\}$, $d\mathscr{G} = \{d\phi : \phi \varepsilon \mathscr{G}\}$. From the well-known identity

$$L_q(d\phi) = d(L_q(\phi)) \quad ,$$

it follows that $d\mathscr{G}$ is the smallest linear space of one-forms containing $d\mathscr{G}_0$ and which is closed with respect to Lie differentiation by elements of $\mathscr{F}$. The elements of $d\mathscr{G}$ are finite linear combinations of elements of the form

$$d(L_{f^1}(\ldots(L_{f^k}(h_i))\ldots)) = L_{f^1}(\ldots(L_{f^k}(dh_i))\ldots)$$

where $f^i(x) = f(x,u^i)$ for some constant $u^i \varepsilon \Omega$. Let $d\mathscr{G}(x)$ denote the space of vectors obtained by evaluating the elements of $d\mathscr{G}$ at x.

Definition 13. (N) is said to satisfy the observability rank condition at $x^0$ if the dimension of $d\mathscr{G}(x^0)$ equals n. If dim $d\mathscr{G}(x) = n$ for all $x \varepsilon M$, then (N) is said to satisfy the observability rank condition.

The observability rank condition provides an <u>algebraic</u> test for local weak observability as the next result demonstrates.

<u>Theorem 16</u> [17]. <u>If (N) satisfies the observability rank</u> <u>condition at $x^0$ then (N) is locally weakly observable at $x^0$</u>.

The observability rank condition is "almost" a necessary condition for local weak controllability, as well, as is seen from

<u>Theorem 17</u> [17]. <u>If (N) is locally weakly observable then</u> <u>the observability rank condition is satisfied generically</u>.

We refer to [17] for the precise meaning of "generic" in Theorem 17. Intuitively, the set of locally weakly observable systems for which the observability rank condition fails is a null set in the space of all locally weakly observable systems.

For analytic systems (N), we have the stronger result

<u>Theorem 18</u> [17]. <u>If (N) is an analytic system then the</u> <u>following conditions are equivalent</u>:

    i)   <u>(N) satisfies the observability rank condition</u>;

    ii)  <u>(N) is weakly observable</u>;

    iii)  <u>(N) is locally weakly observable</u>.

<u>Example</u>. To show that the observability rank condition generalizes Theorem 2, consider the linear system

$$\dot{x} = Fx + Gu \quad ,$$

$$y = Hx \quad .$$

In this case, the space of vector fields $\mathscr{F}$ is generated by the elements

$$\{Fx, F^i g_j, \quad i = 1,2,\ldots,n-1; \quad j = 1,2,\ldots,m\} \quad .$$

If we let $h_j$ denote the jth row of H, then the relevant Lie derivatives are

$$L_{Fx}(h_j F^i)(x) = h_j F^{i+1} x \quad ,$$

$$L_{F^i g_j}(h_k F^\ell)(x) = h_k F^{i+\ell} g_j \quad ,$$

$$L_{Fx}(h_j F^\ell g_k) = 0 \quad .$$

Thus, by the Cayley-Hamilton Theorem $\mathscr{G}$ is generated by the set

$$\{h_i F^k x, h_i F^k g_j : i = 1,\ldots,p, \quad j = 1,\ldots,m, \quad k = 0,1,\ldots,n-1\}$$

and $d\mathscr{G}(x)$ is generated by

$$\Theta = \{h_i F^k : i = 1,\ldots,p, \quad k = 0,1,\ldots,n-1\} \quad .$$

Since $d\mathscr{G}(x)$ is independent of x, it is of constant dimension and the observability rank condition reduces to the requirement that the set $\Theta$ consists of n linearly independent elements.

Other important observability results for general systems are given in [28-30]. Now we consider some specific classes of nonlinear processes.

Bilinear Systems

As in the case of controllability, considerably more detailed results are available on the observability question when we impose a bilinear structure upon the system dynamics f. For instance, consider the homogeneous system

$$\dot{x} = \left( F + \sum_{i=1}^{m} u_i(t) N_i \right) x \quad , \qquad x(0) = x_0 \quad , \qquad (10)$$

$$y(t) = Hx(t) \quad .$$

We have the following result for testing whether or not indistinguishable initial states exist.

Theorem 19 [31]. The homogeneous bilinear system (10) has indistinguishable initial states if and only if there exists a state coordinate transformation T such that

$$T F T^{-1} = \begin{bmatrix} F_{11} & 0 \\ F_{21} & F_{22} \end{bmatrix} \quad ,$$

$$T N_i T^{-1} = \begin{bmatrix} N_{11}^i & 0 \\ N_{21}^i & N_{22}^i \end{bmatrix} \quad , \qquad H T^{-1} = \begin{bmatrix} H_1 & 0 \end{bmatrix} \quad .$$

An alternate characterization of the same result is given by

Theorem 20 [32]. The set of all unobservable (i.e., indistinguishable) states of the system (10) is the largest subspace $\Theta$ of $R^n$ invariant under $F, N_1, \ldots, N_m$, which contains the kernel of H. Theorem 20 suggests the following computational algorithm for calculating the subspace $\Theta$:

i)  Let $U_1$ = range (H');

ii)  Calculate the subspace $U_{i+1} = U_i + N_1' U_i + \ldots + N_m' U_i$ ;

iii)   there exists an integer $k^*$ such that $U_{k^*} = U_{k^*-1}$.
Continue step (ii) until $k^*$ is determined and
set $Z = $ range $U_{k^*}$ .

iv)   $\Theta = Z^{\perp}$, the orthogonal complement of Z.

Additional results on observability for bilinear systems may
be found in the papers already cited in the previous section.

### Factorable Systems

An interesting class of nonlinear systems is that composed
of linear systems connected in parallel with outputs multiplied.
Such "factorable" systems are surprisingly general since a
broad class of systems with separable Volterra kernels may be
expressed as finite sums of factorable systems. Thus, the fac-
torable systems might be thought of as comprising the basic
building blocks for the representation of constant parameter
nonlinear systems.

The mathematical form of a factorable system is

$$\dot{x} = Fx + gu(t) \quad , \tag{11}$$

$$y(t) = \prod_{i=1}^{K} h_i x_i(t) \quad ,$$

where we adopt the notation

$$x(t) = (x_1(t), \ldots, x_K(t))' \quad , \qquad g = (g_1, \ldots, g_K)' \quad ,$$

$$F = \begin{bmatrix} F_1 & & & 0 \\ & F_2 & & \\ & & \ddots & \\ 0 & & & F_K \end{bmatrix} \quad,$$

with $x_i$ being an $n_i$-dimensional vector, and the elements $h_i$, $g_i$, $F_i$ being of corresponding sizes. Thus, the overall state vector $x(t)$ is of dimension $n = n_1 + \ldots + n_K$.

Since the nonlinearity occurs only in the system output, the usual reachability test from the linear theory shows that the factorable system (11) is completely reachable if and only if $W_i(\lambda)$ and $W_j(\lambda)$ have no poles in common for $i \neq j$, where $W_k(\lambda)$ is the transfer matrix associated with the kth component subsystem. Thus, we turn attention to study of the observability properties of the system (11).

It turns out to be convenient to investigate observability for the system (11) by using the Kronecker product of the component subsystems comprising (11). Letting

$$x^{\otimes}(t) = x_1(t) \otimes x_2(t) \otimes \ldots \otimes x_K(t)$$

where $\otimes$ denotes the usual Kronecker product, it can be seen that $x^{\otimes}(t)$ serves as a state vector for the _linear_ system (with $u \equiv 0$). We have

$$\frac{d}{dt} x^{\otimes}(t) = F^{\otimes} x^{\otimes}(t) \quad, \tag{12}$$

$$y(t) = h^{\otimes} x^{\otimes}(t) \quad,$$

with

$$F^{\otimes} = F_1 \otimes I_{n_2} \otimes \ldots \otimes I_{n_K} + I_{n_1} \otimes F_2 \otimes I_{n_3} \otimes \ldots \otimes I_{n_K}$$

$$+ \ldots + I_{n_1} \otimes I_{n_2} \otimes \ldots \otimes I_{n_{K-1}} \otimes F_K \quad ,$$

$$h^{\otimes} = h_1 \otimes h_2 \otimes \ldots \otimes h_K \quad .$$

Knowledge of the initial state $x^{\otimes}(0)$ enables us to compute (up to certain ambiguities in sign) the state $x(0)$. So, we say that the system (11) is completely observable if its associated linear system (12) is observable in the usual sense.

A convenient characterization of the observability of (12) is possible if we define the vector $\Lambda_i$ of _distinct_ characteristic roots of the matrix $F_i$, i.e.,

$$\Lambda_i = (\lambda_{i1}, \ldots, \lambda_{i,p_i}) \quad ,$$

where $i = 1, 2, \ldots, K$, $p_i \leq n_i$. The Kronecker sum of two such vectors is given by

$$\Lambda_1 \oplus \Lambda_2 = \begin{bmatrix} \lambda_{11} + \lambda_{21} \\ \lambda_{11} + \lambda_{22} \\ \vdots \\ \lambda_{11} + \lambda_{2,p_2} \\ \vdots \\ \lambda_{1,p_1} + \lambda_{21} \\ \lambda_{1,p_1} + \lambda_{22} \\ \vdots \\ \lambda_{1,p_1} + \lambda_{2,p_2} \end{bmatrix}$$

In terms of the Kronecker sum of the $\{\Lambda_i\}$, we characterize observability of (12) by the following result.

**Theorem 21** [33]. The factorable system (11) is completely observable if and only if the vector $\Lambda_1 \oplus \Lambda_2 \oplus \ldots \oplus \Lambda_K$ has distinct entries and at most one of the subsystems has multiple characteristic values.

## Polynomial Systems

Very few results exist on the observability question for general continuous-time polynomial systems, i.e., systems of the form

$$\dot{x} = P(x,u) \quad ,$$

$$y(t) = h(x) \quad ,$$

where $P(\cdot,\cdot)$ and $h(\cdot)$ are polynomial functions of their arguments. However, in the discrete-time case a considerable body of knowledge has been reported in [34]. For brevity, let us consider a representative case, the so-called (polynomial) state-affine system

$$x(t+1) = F(u(t)) x(t) + G(u(t)) \quad , \tag{13}$$

$$y(t) = Hx(t) \quad ,$$

where $F(\cdot)$ and $G(\cdot)$ are polynomial functions of $u$ and $H$ is a constant matrix. A particular case is that of internally-bilinear systems, when $F$ and $G$ are themselves linear functions of $u$. The observability of the state-affine system (13) is settled by the following test, which is a restatement of a result taken from [34].

Theorem 22 [34]. The input sequence $w = u_1, u_2, \ldots, u_{n-1}$

distinguishes all pairs of initial states for the state-affine

system (13) if and only if the matrix

$$\Theta(w) = \begin{bmatrix} H \\ HF(u_1) \\ HF(u_2) \, F(u_1) \\ \vdots \\ HF(u_{n-1}) \, \ldots \, F(u_1) \end{bmatrix}$$

has rank n.

Thus, Theorem 21 shows that any input sequence w such that the

observability matrix $\Theta(w)$ is of full rank suffices to distinguish

initial states for the system (13).

For a more complete discussion of various observability

concepts for discrete-time polynomial systems and their inter-

relations, the work [34] should be consulted.

6.    Realization Theory

The specification of the realization problem for linear

systems is simplified by the fact that it is easy to parametrize

the input, output and state spaces via a globally defined coordi-

nate system.  This fact enables us to reduce the problem of

construction of a canonical model from input/output data to a

problem of linear algebra involving matrices.  In the nonlinear

case no such global coordinate system exists, in general, and

it is necessary to take considerable care in defining what we

mean by the problem "data."  We can no longer regard the input/

output data as being represented by an object as simple as an infinite sequence of matrices or, equivalently, a matrix transfer function. So, the first step in the construction of an effective nonlinear realization procedure is to develop a generalization of the transfer matrix suitable for describing the input/output behavior of a reasonably broad class of nonlinear processes.

If we consider the nonlinear system (N)

$$\dot{x} = f(x,u) \quad , \qquad x(0) = x_0 \quad ,$$

$$y = h(x) \quad ,$$

then it is natural to attempt to represent the output of (N) in terms of the input as a series expansion

$$y(t) = w_0(t) + \int_0^t w_1(t,s) \, u(s) \, ds$$

$$+ \int_0^t \int_0^{s_1} w_2(t,s_1,s_2) \, u(s_2) \, u(s_1) \, ds_2 ds_1 + \dots \quad .$$

Formally, the above <u>Volterra series</u> expansion is a generalization of the linear variation of constant formula

$$y(t) = He^{Ft}x_0 + \int_0^t He^{F(t-s)} Gu(s) \, ds \quad .$$

Arguing by analogy with the linear case, the realization problem for nonlinear systems may be expressed as: <u>given the sequence</u> <u>of Volterra kernels</u> $\mathscr{W} = \{w_0, w_1, w_2, \dots\}$, <u>find a canonical model</u> <u>N = (f,h) whose input/output behavior generates</u> $\mathscr{W}$.

Without further hypotheses on the analytic behavior of f, h, together with a suitable definition of "canonical model," the realization problem as stated is much too ambitious and, in general, unsolvable. $\Omega$, let us initially consider conditions under which the Volterra series exists and is unique. Further, we restrict attention to the class linear-analytic systems, i.e., $f(x,u) = f(x) + u(t) g(x)$, where $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ are analytic vector fields. The basic result for Volterra series expansions is

Theorem 23 [35]. If f, g and h are analytic vector fields and if $\dot{x} = f(x)$ has a solution on [0,T] with $x(0) = x_0$, then the input/output behavior of (N) has a unique Volterra series representation on [0,T].

In the case of a bilinear system where $f(x) = Fx$, $g(x) = Gx$, $h(x) = x$, $u(\cdot) =$ scalar control, the Volterra kernels can be explicitly computed as

$$w_n(t,s_1,\ldots,s_n) = e^{Ft}e^{-Fs_1}Ge^{Fs_1}e^{-Fs_2}Ge^{Fs_3} \ldots e^{-Fs_n}Ge^{Fs_n} x_0 .$$

It can be shown [36] that for bilinear systems the Volterra series converges globally for all locally bounded u.

The global convergence of the Volterra series for bilinear processes suggests an approach to the construction of a Volterra expansion in the general case. First, expand all functions in their Taylor series, forming a sequence of bilinear approximations of increasing accuracy. We then compute the Volterra series for each bilinear approximation. However, the simple system

$$\dot{x} = u^2 x \quad , \qquad x(0) = 1 \quad , \qquad h(x) = x \quad ,$$

shows that, in general, no Volterra expansion exists which is valid for all u such that $\|u\|$ is sufficiently small. Further details on the above bilinear approximation technique can be found in [18].

By taking the Laplace transform of the Volterra kernels $\{w_i\}$, it is possible to develop a nonlinear analogue of the standard matrix transfer function of the linear theory. Such an approach as carried out in [37], for example, provides an alternate "frequency-domain" approach to the realization problem. We shall forego the details of such a procedure here due to space considerations, and focus our attention solely upon nonlinear systems whose input/output data is given in terms of the infinite sequence of Volterra kernels $\{w_i\}$.

Now let us turn to the definition of a canonical model for a nonlinear process. As noted earlier, in the linear case we say a model is canonical if it is both reachable (controllable) and observable (constructible). Such a model is also minimal in the sense that the state space has smallest possible dimension (as a vector space) over all such realizations. In order to preserve this minimality property, we make the following

Definition 14. A system N is called locally weakly minimal if it is locally weakly controllable and locally weakly observable.

The relevance of Definition 14 to the realization problem is seen from the following result.

Theorem 24 [17]. Let N, $\hat{N}$ be two nonlinear systems with input sets $\Omega = \hat{\Omega}$, and state manifolds M and $\hat{M}$ of dimensions m, $\hat{m}$, respectively. Suppose $(N, x_0)$ and $(\hat{N}, \hat{x}_0)$ realize the same input/output map. Then if $\hat{N}$ is locally weakly minimal, $\hat{m} \le m$.

Thus, we see that two locally weakly minimal realizations of the same input/output map must be of the same state dimension which is minimal over all possible realizations.

Remark. Two locally weakly minimal realizations need not be diffeomorphic, in contrast to the linear case. This is seen from the two systems

$$ N: \quad \dot{x} = u \quad , \quad y_1 = \cos x \quad , \quad y_2 = \sin x \quad , $$

$$ \hat{N}: \quad \dot{\theta} = u \quad , \quad y_1 = \cos \theta \quad , \quad y_2 = \sin \theta \quad , $$

with $\Omega = \hat{\Omega} = R$, $M = R$, $\hat{M} = S^1$, the unit circle, $y \in R^2$, $x_0 = 0$, $\theta_0 = 0$. Here N and $\hat{N}$ realize the same input/output map. Furthermore, both systems are locally weakly controllable and observable.

The above result leaves open the question if two canonical realizations are isomorphic, i.e., given two nonlinear systems N and $\hat{N}$, with state manifolds M and $\hat{M}$,

$$ \dot{x} = f(x,u) \quad , \tag{N} $$

$$ y = h(x) \quad , $$

$$ \dot{z} = \hat{f}(z,u) \quad , \tag{$\hat{N}$} $$

$$ y = \hat{h}(z) \quad , $$

when does there exist a diffeomorphism $\phi : M \to \hat{M}$ such that $x = \phi(z)$, $z = \phi^{-1}(x)$ or

$$\frac{\partial \phi}{\partial z} f(\phi(z), u) = \hat{f}(\cdot, u) \quad ,$$

$$h(\phi(z)) = \hat{h}(\cdot) \quad .$$

The answer to this question is provided by the following restatement of a result of Sussman.

**Theorem 25** [38]. **Let there be given a mapping** $G_{x_0, u}$ **which to each input** $u(t)$, $0 \leq t \leq T$, **assigns a curve** $y(t)$ **and assume that there exists a finite-dimensional analytic complete system**

$$\dot{x} = f(x, u) \quad , \qquad\qquad x(0) = x_0 \quad ,$$

$$y = h(x) \quad , \qquad\qquad x \in M \quad ,$$

**which realizes the map** $G_{x_0, u}$. **Then** $G_{x_0, u}$ **can also be realized by a system which is weakly controllable and observable. Furthermore, any two such realizations are isomorphic.**

Remark:

In all the results above, as well as those to follow, the conditions of analyticity and completeness of the defining vector fields is crucial. The reason is clear: analyticity forces a certain type of "rigidity" upon the system, i.e, the global behavior of the system is determined by its behavior in an arbitrarily small open set. Completeness is also a natural condition since without this property the system is not totally specified, as it is then necessary to speak about the type of

behavior exhibited in the neighborhood of the vector field
singularity. Fortunately, analyticity and completeness are
properties possessed by any class of systems defined by sets
of algebraic equations, h⁻ ing a reasonable amount of homo-
geneity. For instance, linear systems, bilinear systems and
polynomial systems are all included in this class, together
with any other type of system which is both finite-dimensional
and "algebraic."

Now let us turn to some realization results for specific
classes of nonlinear systems. For ease of notation, we consider
only single-input, single-output systems referring to the refer-
ences for the more general case.

### Bilinear Systems

Given a sequence of Volterra kernels $\{w_i\}_{i=0}^{\infty}$, the first
question is to determine conditions under which the sequence
may be realized by a bilinear system. For this we need the
concept of a factorizable sequence of kernels.

Definition 15. A sequence of kernels $\{w_i\}_{i=2}^{\infty}$ is said to
be factorizable if there exist three matrix functions $F(\cdot)$,
$G(\cdot)$, $H(t,\cdot)$ of sizes $n \times n$, $n \times 1$, $1 \times m$, resp. such that

$$w_i(t,s_1,\ldots,s_i) = H(t,s_1) F(s_2 - s_1) \ldots F(s_{i-1} - s_{i-2}) G(s_i - s_{i-1}) \ ,$$

$$s_1 \leq s_2 \leq \cdots \leq s_i \ .$$

The set $\{F,G,H\}$ is called the factorization of $\{w_i\}$ and the
number n is its dimension. A factorization $\{F_0,G_0,H_0\}$ of min-
imal dimension is called a minimal factorization.

We can now characterize those Volterra kernels which can be realized by a bilinear system.

<u>Theorem 26</u> [36]. <u>The sequence of Volterra kernels $\{w_i\}_{i=0}^{\infty}$ is realizable by a bilinear system if and only if $w_1$ has a proper rational Laplace transform and $\{w_i\}_{i=2}^{\infty}$ is factorizable by functions F, G, H with proper rational Laplace transforms.</u>

Let us assume that a given sequence of kernels $\{w_i\}$ is bilinearly realizable. We then face the question of the construction of a minimal realization and its properties. The main result in this regard is

<u>Theorem 27</u> [36]. <u>For a sequence of bilinearly realizable kernels $\{w_i\}$, the minimal realizations are such that</u>

i) <u>the state space dimension $n_0$ is given by the dimension of the linear system whose impulse response matrix is</u>

$$W(s) = \left[ \begin{array}{c|c} w_1(t,s) & H_0(t,s) \\ \hline G_0(s) & F_0(s) \end{array} \right] \quad ;$$

ii) <u>any two minimal realizations</u>

$$\dot{x} = Ax + Bu + Nxu \quad ,$$

$$y = Cx \quad ,$$

$$\dot{z} = \hat{A}z + \hat{B}u + \hat{N}zu \quad ,$$

$$y = \hat{C}z \quad ,$$

are related by a linear transformation of their state spaces, i.e., there exists an $n_0 \times n_0$ matrix T such that

$$\hat{A} = TAT^{-1} \quad , \qquad \hat{B} = TB \qquad \hat{N} = TNT^{-1} \quad , \qquad \hat{C} = CT^{-1} \quad .$$

Theorem 27 provides the basic information needed in order to actually construct the matrices A, B, C, N of a minimal realization. Since W(s) is the impulse response of a linear system of dimension $n_0$, there must exist three matrices P, Q, R of sizes $n_0 \times n_0$, $n_0 \times (n+1)$, $(n+1) \times n_0$ such that

$$W(s) = Re^{Ps}Q \quad .$$

By partitioning Q and R as

$$R = \begin{pmatrix} R_1 \\ \\ R_2 \end{pmatrix} \quad , \qquad Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \quad ,$$

where $R_1$ is $1 \times n_0$ and $Q_1$ is $n_0 \times 1$, we obtain

$$w_1(t,s) = R_1 e^{Ps}Q_1 \quad , \qquad H_0(t,s) = R_1 e^{Ps}Q_2 \quad ,$$

$$G_0(s) = R_2 e^{Ps}Q_1 \quad , \qquad F_0(s) = R_2 e^{Ps}Q_2 \quad .$$

We now define the matrices of our minimal realization as

$$A = P \quad , \qquad B = Q_1 \quad , \qquad C = R_1 \quad , \qquad N = Q_2 R_2 \quad .$$

Thus, the surprising conclusion is that the realization procedure for bilinear systems can be carried out using essentially

the same techniques as those employed in the linear case once the minimal factorization $\{F_0, G_0, H_0\}$ has been found.

Other approaches to the construction of bilinear realizations are discussed in [39], while results for the discrete-time case are given in [40]. The case of multilinear systems is similar to the bilinear situation and is discussed in detail in [41].

### Linear-Analytic Systems

The general question of when a given Volterra series $\{w_i\}_{i=0}^{\infty}$ admits realization by a finite-dimensional linear-analytic system $\{f, g, h\}$ of the form

$$\dot{x} = f(x) + ug(x) \quad ,$$

$$y = h(x) \quad ,$$

has no easily computable answer, although some difficult to test conditions are given in [42]. On the other hand, if the Volterra series is finite then the results are quite easy to check and reasonably complete. For their statement, we make

Definition 16. A Volterra kernel $w(t, s_1, \ldots, s_r)$ is called separable if it can be expressed as a finite sum

$$w(t, s_1, \ldots, s_r) = \sum_{i=1}^{m} \gamma_i^1(t) \gamma_i^2(s_1) \ldots \gamma_i^r(s_r) \quad .$$

It is called differentiably separable if each $\gamma_i$ is differentiable and is stationary if

$$w(t,s_1,\ldots,s_r) = w(0,s_1-t,s_2-t,\ldots,s_r-t) \quad .$$

The main theorem characterizing the realization of finite
Volterra series by a linear-analytic system is

Theorem 28 [42]. A finite Volterra series is realizable
by a (stationary) linear-analytic system if and only if each
term in the series is individually realizable by a (stationary)
linear-analytic system. Furthermore, this will be the case if
and only if the kernels are (stationary and differentiably)
separable.

The above result leaves open the question of actual com-
putation of the vector fields {f,g,h} defining the linear-
analytic realization of a finite Volterra series. However,
this problem is formally bypassed by the following result.

Theorem 29 [42]. A finite Volterra series has a (stationary)
linear-analytic realization if and only if it has a (stationary)
bilinear realization.

From Theorem 29 it is tempting to conclude that there is
no necessity to study linear-analytic systems when given a finite
Volterra series, since we can always realize the data with a
bilinear model. Unfortunately, the situation is not quite so
simple since the dimension of the canonical bilinear realization
will usually be somewhat greater than that of the corresponding
linear-analytic model. To illustrate this point, consider the
finite Volterra series

$$w_0(t) = 0 \quad , \qquad w_1(t,s_1) = \exp(t-s_1) \quad , \qquad w_2(t,s_1,s_2) = 0 \quad ,$$

$$w_3(t,s_1,s_2,s_3) = \tfrac{1}{3}\exp[3(t-s_1)]\exp[2(s_1-s_2)]\exp[(s_2-s_3)] \quad ,$$

$$w_i = 0 \quad , \qquad i \geq 4 \quad .$$

This series is realized by the three-dimensional bilinear model

$$\dot{x} = Fx + Gu + Nxu \quad , \quad \text{where}$$

$$y(t) = x(t) \quad ,$$

$$F = \begin{bmatrix} 1 & 0 & \tfrac{1}{6} \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad , \qquad N = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \quad , \qquad G = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

However, the same set of kernels is also realized by the one-dimensional linear-analytic system

$$\dot{x} = \sin x + u(t) \quad , \qquad x(0) = 0 \quad ,$$

$$y(t) = x(t) \quad .$$

## Polynomial Systems

If the system input/output map is of polynomial type, i.e., each term in the Volterra series is a polynomial function of its arguments, then an elegant realization theory for such maps has been developed by Sontag [27] in the discrete-time case. Since presentation of the details would entail too large an excursion into algebraic geometry, we loosely summarize the main results referring to the references for a more complete account.

For simplicity, we restrict our account to <u>bounded</u> polynomial input/output maps f, which means that there exists an integer α such that the degree of each term in the Volterra series for f is uniformly 'ɔunded by α. The main realization result for bounded polynomial input/output maps is

Theorem 30 [27]. <u>If a bounded input/output map is at all</u> <u>realizable by a polynomial system, then it is realizable by an</u> <u>observable state-affine system of the form</u>

$$x(t+1) = F(u(t)) x(t) + G(u(t)) \quad , \qquad x(0) = 0 \quad ,$$

$$y(t) = Hx(t) \quad ,$$

<u>where F(·) and G(·) are polynomial matrices, H is a linear map</u> <u>and the system state space is</u> $R^n$.

An observable state-affine realization is termed <u>span-</u> <u>canonical</u> if the subspace of reachable states is all of $R^n$. Then it can be shown that a span-canonical realization of a given bounded finitely realizable f always exists and any two such realizations are related by a state coordinate change. Furthermore, a realization is span-canonical if and only if its dimension n is minimal among all state-affine realizations of the same input/output map.

Somewhat less complete results are also reported in [27] for unbounded polynomial input/output maps. The relationship between the foregoing discrete-time results and the continuous-time case is still far from clear, due mainly to the nonreversibility of difference (as opposed to differential) equations and to the different algebraic properties of difference and

differential operations. To bridge this gap may turn out to be a nontrivial task, as is seen by the recent result [43] that a "finite" continuous-time map has its canonical state space unconstrained, which is far from true in the discrete-time setting.

Some additional work on polynomial systems taking a functional-analytic, rather than algebraic, approach is reported in [44].

### "Almost"-Linear Systems

By imposing special types of nonlinearities upon a standard linear system, it is possible to employ techniques similar to the usual linear methods for realization of input/output maps. In this regard we note the "factorable" Volterra systems considered earlier, having the internal form

$$\dot{x} = Fx + gu(t) \quad ,$$

$$y(t) = \prod_{i=1}^{K} c_i x_i (t) \quad .$$

Here the nonlinearities enter only through the system output. Utilizing tensor products, it can be shown [33] that the input/output behavior of such a process can be described by a so-called Volterra transfer function $H(s_1, \ldots, s_K)$. Since a factorable Volterra system consists of $K$ linear subsystems connected in parallel, with the outputs multiplied, the realization problem reduces to determining the transfer functions $H_1(s), \ldots, H_K(s)$ of each subsystem from $H(s_1, \ldots, s_K)$.

If the $H_i^{'}(s)$ are known, then standard linear theory provides the overall system realization. Techniques for solving this problem are reported in [33].

In another direction, we could consider cascade combinations of linear subsystems and static power nonlinearities as in [45]. For inputs of the form

$$u(t) = \sum_{i=1}^{P} a_i e^{\lambda_i t} \quad ,$$

the output of such a system is

$$y(t) = \sum_{i_1=1}^{P} \sum_{i_2=1}^{P} \cdots \sum_{i_m=1}^{P} \left[ \prod_{j=1}^{m} a_i \right] \exp\left( \sum_{k=1}^{m} \lambda_{i_k} t \right) \quad ,$$

where $m > 0$ is an integer defining the degree of the static non-linearity, i.e., the block diagram of the system is

$$u(t) \rightarrow \boxed{H_0(s)} \rightarrow \boxed{(\cdot)^{P_1}} \rightarrow \boxed{H_1(s)} \rightarrow \cdots \rightarrow \boxed{H_q(s)} \rightarrow \boxed{(\cdot)^{P_q}} \rightarrow \boxed{H_q(s)} \rightarrow y(t) \quad ,$$

where $\prod_{j=1}^{q} P_j = m$ and $H_j(s)$ is a strictly proper rational function of degree $\leq n$, $j = 0, 1, \ldots, q$. In the work [45] an algorithm is given for solution of the minimal realization problem for such a system.

## 7. Conclusions and Future Research

The foregoing results leave little doubt that substantial progress has been made in nonlinear system theory over the past decade. As noted in the introduction, we have focused only upon problems of reachability, observability and realization, omitting the more well-known areas of stability and optimal control. Advances in these areas have also been impressive as can be seen from the works [46-47]. Thus, the inescapable conclusion is that nonlinear system theory is alive and well and it is to be expected that progress on outstanding issues will be rapid in the years to come.

By way of closing remarks, let us now engage in a bit of crystal ball-gazing and sketch some problem areas which seem to be most important for future research in nonlinear systems.

1) Computational Methods — the effective employment of any of the results given here relies upon efficient computational algorithms. For those procedures which mimic the linear case (e.g., bilinear realization), good methods already exist for computing the necessary quantities. However, much remains to be done to develop comparable methods for, say, computing the reachable set for a nonlinear process or determining the Volterra series of a given input/output map from measured data;

2) Stochastic Effects — a cornerstone of linear system theory is the Kalman filter and its associated apparatus for determining the "best" estimate of system parameters in the presence of noise. This is a special case of the more general stochastic realization problem, in which the input/output data

itself is corrupted by noise and "best" estimates of the system
model must be made. Again in the linear case results are avail-
able [48]. However, almost nothing has been accomplished along
these lines for nonlinear ɼ_ocesses. It seems likely, though,
that with the increased understanding now available good progress
can be made. We should note the works [49-50] as promising
initial forays in this area;

3) Non-Analytic Systems — almost all interesting results
for nonlinear systems are for processes whose defining vector
fields are analytic. As pointed out earlier, there is good
reason for this since the local behavior of analytic systems
entirely determines the global behavior. However, there are
interesting and important processes which do not fall into
this category (e.g., systems with threshold effects, processes
with phase transitions, and so on). A concerted attempt at
relaxation of the analyticity assumptions can be expected to
yield substantial dividends in furthering our ability to tackle
a variety of problems in the social and biological sciences;

4) Infinite-Dimensional Processes — in general, systems
whose underlying dynamics are governed by partial differential
equations or processes involving time-lag effects cannot be
modeled by a finite set of ordinary differential or difference
equations. Even in the linear case such processes lead to
thorny analytical questions which are, as yet, far from being
well under control. So, it is perhaps wildly optimistic to
think that substantial advances can be made in this direction
for nonlinear processes. Nonetheless, we have seen that many

of the results and techniques of the linear theory can be
extended to classes of nonlinear systems with modest additional
effort.  So, it seems reasonable to attempt an investigation
of those nonlinear problems which are the counterparts of the
corresponding infinite-dimensional linear processes.

# References

[1] Mayne, D., and R. Brockett, eds., <u>Geometric Methods in System Theory</u>, Reidel, Dordrecht, 1973.

[2] Martin, C., and R. Hermann, eds., <u>The 1976 Ames Research Center (NASA) Conference on Geometric Control Theory</u>, Math Sci Press, Brookline, Mass., 1977.

[3] Byrnes, C., and N. Hurt, On the Moduli of Linear Dynamical Systems, <u>Advances in Mathematics</u>, <u>4</u> (1979), 83-122.

[4] Hermann, R., <u>Linear System Theory and Introductory Algebraic Geometry</u>, Interdisciplinary Mathematics, Vol. 8, Math Sci Press, Brookline, Mass., 1974.

[5] Fuhrmann, P., Algebraic System Theory: An Analyst's Point of View, <u>J. Franklin Institute</u>, <u>301</u> (1976), 521-540.

[6] Kalman, R., P. Falb, and M. Arbib, <u>Topics in Mathematical System Theory</u>, McGraw-Hill, New York, 1969.

[7] Casti, J., <u>Dynamical Systems and their Applications: Linear Theory</u>, Academic Press, New York, 1977.

[8] Brockett, R., <u>Finite-Dimensional Linear Systems</u>, John Wiley, New York, 1970.

[9] Kalman, R.E., On Partial Realizations of a Linear Input/ Output Map, in N. de Claris and R. Kalman, eds., <u>Systems and Networks</u>, Holt, New York, 1968.

[10] Kalman, R.E., On Partial Realizations, Transfer Functions and Canonical Forms, <u>Acta Polytechnica Scandinavica</u>, <u>31</u> (1979), 9-32.

[11] Bryson, A., and Y.C. Ho, Applied Optimal Control, Blaisdell, Waltham, Mass., 1969.

[12] Aggarwal, J., and M. Vidyasagar, eds., Nonlinear Systems: Stability Analysis, Dowden, Hutchinson and Ross, Stroudsburg, Penn., 1977.

[13] Peixoto, M., ed., Dynamical Systems, Academic Press, New York, 1973.

[14] Lee, E., and L. Markus, Foundations of Optimal Control, John Wiley, New York, 1967.

[15] Brockett, R., On the Reachable Set for Bilinear Systems, in R. Mohler and A. Ruberti, eds., Proc. 1974 Conference on Bilinear Systems, Springer, New York, 1975.

[16] Kalman, R.E., Pattern Recognition Problems of Multilinear Machines, in Proc. IFAC Symposium on Technical and Biological Problems of Control, Yerevan, Armenia, USSR, 1968.

[17] Hermann, R., and A. Krener, Nonlinear Controllability and Observability, IEEE Transactions on Automatic Control, AC-22 (1977), 728-740.

[18] Brockett, R., Nonlinear Systems and Differential Geometry, Proc. IEEE, 64 (1976), 61-72.

[19] Hermes, H., On Local and Global Controllability, SIAM Journal of Control, 12 (1974), 252-261.

[20] Hirschorn, R., Global Controllability of Nonlinear Systems, SIAM Journal of Control, 14(1976), 700-711.

[21]   Haynes, G.W., and H. Hermes, Nonlinear Controllability

        via Lie Theory, <u>SIAM Journal of Control</u>, <u>8</u> (1970),

        450-460.

[22]   Sussman, H., and V. Jurdjevic, Controllability of Nonlinear

        Systems, <u>Journal of Differential Equations</u>, <u>12</u> (1972),

        95-116.

[23]   Mohler, R., <u>Bilinear Control Processes</u>, Academic Press,

        New York, 1973.

[24]   Sussman, H., and V. Jurdjevic, Control Systems on Lie

        Groups, <u>Journal of Differential Equations</u>, <u>12</u> (1972).

[25]   Imbert, N., M. Clique, and A.-J. Fossard, Un Critère de

        Gouvernabilité des Systèmes Bilinéaires, <u>RAIRO</u>,

        J-3 (1979), 55-64.

[26]   Lukes, D.L., On the Global Controllability of Nonlinear

        Systems, in L. Weiss, ed., <u>Ordinary Differential</u>

        <u>Equations</u>, Academic Press, New York, 1971.

[27]   Sontag, E., <u>Polynomial Response Maps</u>, Springer Lecture

        Notes in Control, Vol. 13, Berlin, 1979.

[28]   Griffith, E., and K.S.P. Kumar, On the Observability of

        Nonlinear Systems - I, <u>Journal of Mathematical Analysis</u>

        <u>and Applications</u>, <u>35</u> (1971), 135-147.

[29]   Kou, S.R., D. Elliott, and T.J. Tarn, Observability of

        Nonlinear Systems, <u>Information and Control</u>, <u>22</u> (1973),

        89-99.

[30]   Kostynkovskii, Y., Observability of Nonlinear Controlled

        Systems, <u>Automation and Remote Control</u>, <u>9</u> (1968),

        1384-1396.

[31] Brockett, R., On the Algebraic Structure of Bilinear
    Systems, in R. Mohler and A. Ruberti, eds., <u>Theory</u>
    <u>and Application of Variable Structure Systems</u>,
    Academic Press, New York, 1972.

[32] Isidori, A., and A. Ruberti, Realization Theory of
    Bilinear Systems, in [1].

[33] Harper, T., and W. Rugh, Structural Features of Factorable
    Volterra Systems, <u>IEEE Transactions on Automatic</u>
    <u>Control</u>, AC-21 (1976), 822-832.

[34] Sontag, E., On the Observability of Polynomial Systems, I:
    Finite-Time Problems, <u>SIAM Journal of Control</u>, <u>17</u>
    (1979), 139-151.

[35] Lesiak, C., and A. Krener, The Existence and Uniqueness of
    Volterra Series for Nonlinear Systems, <u>IEEE Trans-</u>
    <u>actions on Automatic Control</u>, AC-23 (1978), 1090-1095.

[36] d'Alessandro, P., A. Isidori, and A. Ruberti, Realization
    and Structure Theory of Bilinear Systems, <u>SIAM</u>
    <u>Journal of Control</u>, <u>12</u> (1974), 517-535.

[37] Mitzel, G., S. Clancy, and W. Rugh, On Transfer Function
    Representation for Homogeneous Nonlinear Systems,
    <u>IEEE Transactions on Automatic Control</u>, AC-24 (1979),
    242-249.

[38] Sussman, H., Minimal Realizations of Nonlinear Systems,
    in [1].

[39] Isidori, A., Direct Construction of Minimal Bilinear
    Realizations from Nonlinear Input/Output Maps, <u>IEEE</u>
    <u>Transactions on Automatic Control</u>, AC-18 (1973)

[40] Fornasini, E., and G. Marchesini, Algebraic Realization Theory of Bilinear Discrete-Time Input/Output Maps, J. Franklin Institute, 301 (1976), 143-159.

[41] Marchesini, G., and G. Picci, Some Results on the Abstract Realization Theory of Multilinear Systems, in R. Mohler and A. Ruberti, eds., Theory and Application of Variable Structure Systems, Academic Press, New York, 1972.

[42] Brockett, R., Volterra Series and Geometric Control Theory, Automatica, 12 (1976), 167-176.

[43] Crouch, P., Finite Volterra Series, doctoral dissertation, Harvard University, 1977.

[44] Porter, W.A., An Overview of Polynomic System Theory, Proc. IEEE, 64 (1976), 18-23.

[45] Smith, W., and W. Rugh, On the Structure of a Class of Nonlinear Systems, IEEE Transactions on Automatic Control, AC-19 (1974), 701-706.

[46] Gurel, O., and O. Rössler, eds., Bifurcation Theory and Applications in Scientific Disciplines, NY Academy of Sciences, Vol. 316, 1979.

[47] Brunovsky, P., On the Structure of Optimal Feedback Systems, Proc. International Congress of Mathematicians, Helsinki, 1978.

[48] Picci, G., Stochastic Realization of Gaussian Processes, Proc. IEEE, 64 (1976), 112-122.

[49] Willsky, A., and S. Marcus, Analysis of Bilinear Noise Models in Circuits and Devices, J. Franklin Institute, 301 (1976), 103-122.

[50] Marcus, S., Modeling and Analysis of Linear Systems with
Multiplicative Poisson White Noise, in [2].