AN ABSTRACT OF THE THESIS OF

Chun Sik Hwang for the degree of Doctor of Philosophy in

Electrical and Computer Engineering presented on October 17, 1985
(Major Department)

Title: OBSERVABILITY AND INFORMATION STRUCTURE OF NONLINEAR SYSTEMS

Redacted for Privacy

Approved: ______________________________

An observability problem for both deterministic and stochastic systems is studied here.

Deterministic observability is a determination of whether every state of the system is connected to the observation mechanism and how it is connected, if connected. On the other hand, stochastic observability discusses the "tightness" of the connection in terms of the chosen statistical sense.

For the deterministic system observability two conditions, connectedness and univalence, are obtained from modification of the global implicit-function theorem. Depending on how the conditions are satisfied observability is classified in three categories; observability in the strict sense, observability in the wide sense and the unobservable case.

Two underwater tracking examples, the bearing-only-target (BOT) problem described in the mixed-coordinate system, and an array SONAR

problem described in terms of a small number of sensors and various measurement policies are analyzed.

For the stochastic system observability, an information theoretic approach is introduced. The Shannon concepts of information are considered instead of Fisher information. Computed here is the mutual information between the state and the observation. Since this quantity is expressed as an entropy difference between a priori and a posteriori processes, two densities are required for computation. Due to the difficulty in solving the density equation, the second moment approximation of the densities is considered here. Then, the mutual information is used as a criterion to determine the "degree of observability."

Information sensitivity with respect to various coordinate systems, including rectangular, modified polar and mixed coordinates are analyzed for the BOT system. In an array SONAR, a combination of relative delay and Doppler measurements for up to three sensors are compared.

OBSERVABILITY AND INFORMATION STRUCTURE OF NONLINEAR SYSTEMS

by

Chun Sik Hwang

A thesis

submitted to

Oregon State University

in partial fulfillment of

the requirement for the

degree of

Doctor of Philosophy

Completed

October 17, 1985

Commencement June, 1986

APPROVED:

Redacted for Privacy

Professor of Electrical and Computer Engineering in charge of major

Redacted for Privacy

Head of Department of Electrical and Computer Engineering

Redacted for Privacy

Dean of Graduate School

Date Thesis is presented October 17, 1985

Typed by Modern Office Management Systems

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to Dr. Ronald R. Mohler, his major professor, for his good-natured guidance in every aspect of his research.

Sincere thanks are also due to the other members of his doctoral committee: Dr. Richard J. Bucolo, Dr. Wojciech J. Kolodziej, Dr. Rudolf S. Engelbrecht, and Dr. Robert M. Burton for their invaluable suggestions and assistance.

Financial support for this research was provided both by the ONR and the Korean Army, and it is gratefully acknowledged.

Finally, the author expresses his deep thanks to his wife, Hoil, and two children, Man Soo, and Se Jeong for their faith and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

OBSERVABILITY AND INFORMATION STRUCTURE OF NONLINEAR SYSTEMS

CHAPTER 1: INTRODUCTION

A state space description is one way widely used to describe a physical dynamic system in a mathematical model. Here every individual state represents some property of the actual system characteristics. So, to understand the nature of the system from outside the dynamic model, one is required to observe or measure necessary states. But, sometimes, it is not possible to access and measure all of the necessary states from the outside. Even in case of such possibility, it may be too expensive economically to measure specific states. In this case, one thinks about another indirect way instead of direct measuring at high cost or unmeasurable states, i.e., if one can somehow reconstruct every necessary state by utilization of less expensive or measurable states only, then one might be satisfied. Observability is a basic system study relevant to this subject. One is interested, here, in determination of whether measured data is enough to reconstruct all of the states. Importance of system observability stems from another aspect. I.e., if the system is not observable for some reason, then certain states which are estimated from this insufficient information may be inaccurate and thus any further action, for example, feedback control which is evaluated based on inaccurate states may exhibit undesirable performance.

If noise is involved in the description of system and/or measurement dynamics then the observability concept is changed from the above deterministic case. Here, one is more interested in "how much" the system is observable in terms of a chosen probabilistic sense, i.e., degree of observability rather than a "yes" or "no" type answer. Of course, there are many different ways to measure the degree of observability. Apparently, one way is using information theory. Here, evaluated is the quantity of common information, so called, mutual information between the state x_t and the observation y_t, and this quantity is used as a criterion to determine the degree of observability, i.e., a calculation is made of the amount of information about state x_t which is contained in the observation y_t.

In Chapter Two, deterministic observability is studied. After defining the problem, observability criteria for linear systems and former results for nonlinear systems are summarized. Since, nonlinear observability is a geometric functional structure problem, a functional analytic approach is used. A modified version of the global implicit function theorem is obtained from the result of Palais [1]. To apply the modified version of this theorem in the nonlinear observability problem, appropriate algebraic modification of the observation equation is required. Thus two conditions, connectedness and univalence, are derived. Depending on how the conditions are satisfied, observability is classified in three categories; observability in the strict sense, observability in the wide sense and the unobservable case. Two important applicational examples are analyzed using the result. I.e., BOT tracking which is described in the mixed-coordinate system, and an array SONAR with a small number of sensors and with various measurement policies are analyzed.

In Chapter Three, stochastic-system observability is studied using an information-theoretical approach. The term "information" is interpreted in the Shannon sense rather than the Fisher sense here. So, information is not an abstract quantity but a substantial quantity having appropriate units. With the basic definitions of information and entropy concepts, mutual information is introduced and expressed in terms of entropy difference, i.e., difference between unconditional and conditional entropies. Since the evaluation of the mutual information of stochastic processes requires more conditions than simple random variables that is introduced using measure theory. Under the proper conditions, entropy is expressed in terms of estimation covariances. Therefore, the mutual information can be obtained from two covariances - unconditional and conditional covariances. Both can be obtained from an adopted filter algorithm. But the non-Gaussian case generally requires knowledge of the probability distribution or higher order moments. Here the second moment approximations of the densities are considered.

A brief discussion on the relationship between deterministic and stochastic observability follows. A result on the relationship between the Fisher information and Shannon's mutual information is discussed.

Chapter Four shows simulation results of various practical problems in view of observability and information structure. Followed by a simple linear-system example is BOT tracking and array SONAR problems which are analyzed in Chapter two.

Information structures of observable and unobservable cases for all examples are compared with various parameter changes. Estimation error analysis in terms of the contents of information is shown.

Chapter Five summarizes the results.

Notation

The following notations will be used throughout:

R^n	Euclidean n-dimensional space
$\|\|\cdot\|\|$	Euclidean norm
trA	Trace of a matrix A
A*	Conjugate transpose of matrix or vector A (A^T will be used when A is real)
$A^{(n)}(t)$	n-th time derivative of A(t)
$\frac{\partial}{\partial x}$	Gradient vector of nonanticipative functionals
$\frac{\partial^2}{\partial \mathbf{x}\, \partial \mathbf{x}^T}$	Jacobian matrix of nonanticipative functionals
<<	Absolute continuity or negligibly small
~	Equivalence or approximated quantity
(Ω, F, μ)	Complete measure space
(Ω, F, P)	Complete probability space
F_t	Sub - σ - algebra of F
$\|\cdot\|$	Absolute value
$\dot{x}(t)$	Denotes $\frac{dx}{dt}$
x(t)	Deterministic time variable of vector x.
$\chi(t)$	Scalar quantity of x(t)
x_t	Stochastic time variable of vector x for a particular elementary event $\omega \in \Omega$

Notation (cont.)

x_t	Scalar quantity of x_t
$\{x_t\}$	Stochastic vector process
$E[x_t]$	Expectation of x_t
$E[x_t \mid y_t]$	Conditional expectation with respect to a given measurement y_t
$E[x_t \mid F_t^Y]$	Conditional expection with respect to a given sub-σ - algebra generated by $\{y_s, 0 \le s \le t\}$
C^1	Space of continuous functions
$[a,b]$	Closed interval
(a,b)	Open interval
$a \in A$	a is an element of A
P	Covariance matrix
p	Probability distribution (probability density function when not confused with distribution)
p	Probability density function
**	End of proof
$f_{\mid U}$	f is restricted by U

CHAPTER 2: OBSERVABILITY OF DETERMINISTIC NONLINEAR SYSTEMS

2-1 The observability problem and former results.

Consider a mathematical description of physical dynamic system which is expressed in the first-order vector differential equation

$$\frac{dx(t)}{dt} = f(x(t),u(t),t), \tag{2-1}$$

where $x(t)$ is an n-dimensional state vector, $u(t)$ is an r-dimensional control input, and t is the time variable. Assume the dynamic property of the system is known, i.e., an n-vector valued function $f(.)$ and $u(t)$ is known for $t \geq t_o$. Further assume that $f(.)$ satisfies the existence and uniquness conditions of the $x(t)$, i.e.,

1. $f(.)$ is continuous in t and once continuously differentiable in x and u for fixed t, $t \varepsilon [0, \infty)$.
2. $f(.)$ satisfies uniform Lipschitz condition on x.

$$||f(x^1(t),.)-f(x^2(t),.)|| \leq M \; || x^1(t) - x^2(t)||, \tag{2-2}$$

where $||.||$ is the Euclidean norm, M is a bounded real positive constant. Under the above conditions one wants to know the time trajectory of $x(t)$ from (2-1). For this purpose one constructs an integral operator $g(.)$ such that

$$x(t)=g(x(t_o),u(t),t). \tag{2-3}$$

But knowing the operator $g(.)$ does not mean that, actually, one can get the solution trajectory $x(t)$ of (2-1) because the initial state $x(t_o)$ in (2-3) is not known. So, if one can somehow establish $x(t_o)$, then the problem will be solved. To establish the initial state $x(t_o)$, in practice, one might construct another equation known as a "measurement" or "observation" equation since there is no way to know $x(t_o)$ from the system model equation (2-1) in itself. Using appropriate measuring or observing devices, necessary state variables or other variables are observed for some period of time, say $[t_o, t_1]$. Then using the observed data, $x(t_o)$ might be determined indirectly. This obervation mechanism might be modelled mathematically as

$$y(t) = h(x(t),t), \tag{2-4}$$

where $h(.)$ is an m-dimensional vector function and $y \in R^m$. Here m is not necessarily the same as n. Usually from the physical availability and economic point of view, m is less than n.

If (2-4) is uniquely solvable for $x(t)$, then every state $x_i(t)$, $i=1,2,\ldots,n$ can be computed with only currently measured $y(t)$, i.e., the information measured is in a sense complete. But if observed information is incomplete, i.e., (2-4) is not uniquely solvable for $x(t)$, then there arises the problem of evaluating the state $x(t)$ by some indirect method using state equation (2-1) as well as observation equation (2-4).

The observability problem has been well investigated and the result is clear for the linear system where the test of nonsingularity of the observability matrix or equivalently rank test is enough.

But for the general nonlinear system these techniques are not applicable, unfortunately, since even in case of nonsingular or full rank condition of the observability matrix, one cannot solve uniquely $x(t)$ from (2-1) and (2-4). Thus $x(t_o)$ can not be determined uniquely. Before investigating this problem further, a summary of the former results are made.

2.1.1 Former results on system observability

1. Linear system.

Consider the time-varying linear system

$$\dot{x}(t) = A(t)x(t)+B(t)u(t), \tag{2-5}$$

$$y(t) = C(t)x(t)+D(t)u(t), \tag{2-6}$$

where matrices $A(t)$, $B(t)$, $C(t)$, $D(t)$ are known $n \times n$, $n \times r$, $m \times n$, $m \times r$, respectively and entries are continuous in t over $(-\infty, \infty)$. Observability of the system (2-5), (2-6) is dealt with in the most standard textbooks [2], [3].

First define the observability of the linear system (2-5), (2-6) as follows:

Definition [3]

The system (2-5), (2-6) is completely observable at t_o if for any $x(t_o)$, there exists a finite $t_1 > t_o$ such that the knowledge of $u(t)$ and

$y(t)$, $t \varepsilon [t_o, t_1]$ is sufficient to determine $x(t_o)$.

From solution of (2-5), $y(t)$ of (2-6) becomes

$$y(t) = C(t)\phi(t,t_o)x(t_o)+C(t)\int_{t_o}^{t}\phi(t,s)B(s)u(s)ds + D(t)u(t), \quad (2\text{-}7)$$

where $\phi(.)$ is the transition matrix of the homogeneous part of (2-5). From (2-7) observability criterion is derived as [2];

Criterion 1

The system (2-5), (2-6) is observable at t_o if and only if the columns of the m x n matrix function $C(t)\phi(t,t_o)$ are linearly independent on $[t_o, t_1]$.

By multiplying $\phi^*(t,t_o)C^*(t)$, integrating from t_o to t and retaining the zero input response of (2-7), Criterion 2 is obtained.

Criterion 2

The system (2-5), (2-6) is observable at t_o if and only if the Grammian matrix N(.)

$$N(t_o,t)= \int_{t_o}^{t}\phi^*(s,t_o)C^*(s)C(s)\phi(s,t_o)ds \quad (2\text{-}8)$$

is nonsingular.

Another criterion which is more convenient to apply can be derived from Criterion 1, i.e.,

$$F(t) = C(t)\phi(t,t_o), \tag{2-9}$$

are linearly independent on $[t_o,t_1]$ if the matrix

$$V_1(t) = [F^*(t)|F^{(1)*}(t)|\ldots|F^{(n-1)*}(t)], \tag{2-10}$$

has rank n. Thus we have the third criterion.

Criterion 3

System is observable at t_o if and only if there exists a $t \varepsilon [t_o,t_1]$ such that observability matrix

$$V^*(t) = \begin{bmatrix} Q_o(t) \\ Q_1(t) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Q_{n-1}(t) \end{bmatrix}, \tag{2-11}$$

has rank n, where

$$Q_{k+1}(t) = Q_k(t)A(t) + \frac{d}{dt}Q_k(t),\ k=0,1,\ldots,n-1, \tag{2-12}$$

$$Q_o(t) = C(t) .$$

For the time-invariant linear case the following observability conditions are equivalent. The time-invariant linear system is,also, observable at t_o in $[o, \infty)$ if one of the following conditions is satisfied,

1) The columns of ce^{At} are linearly independent on $[0,\infty)$

2) The columns of $C(SI-A)^{-1}$ are linearly independent. S is Laplace transform parameter.

3) $N(t_o,t) = \int_{t_o}^{t} e^{A^*(s-t_o)}C^*Ce^{A(s-t_o)}ds,$

is nonsingular for any $t_o>o$ and $t>t_o$.

4) The mn x n observability matrix

$$V^* = \begin{bmatrix} C \\ CA \\ CA^2 \\ \cdot \\ \cdot \\ \cdot \\ CA^{n-1} \end{bmatrix}, \qquad (2\text{-}13)$$

has rank n.

2. Nonlinear system.

As known, the observability property of the general nonlinear system is not a global property,i.e.,an observable nonlinear system in one time interval or one portion of state space may be unobservable in a different interval. In a geometric sense, a functional relation between measurement space and state space might not be in one-to-one correspondence such that the inverse function between the two spaces is not uniquely defined globally even though it is so defined locally.

Various authors have studied the nonlinear observability problem in many ways. Extension of the linear system observability criteria to the nonlinear case is attempted in [4], [5]. The observability rank condition using Lie algebra [6], [7], [8] and Taylor series expansion [9] are reviewed. As the observability problem is, sometimes, called "an inverse problem," the inverse function theorem in analysis is used widely. In this approach the Jacobian matrix of the function which is related to the observation equation plays a central role. [10] - [17] can be viewed in this category.

1) Linearization method

The nonlinear system and observation equations

$$\dot{x}(t) = f(x(t),u(t),t), \tag{2-14}$$

$$y(t) = h(x(t),t), \tag{2-15}$$

are linearized around some reference point, for example, the origin or the equilibrium point or a proper operation point to study the neighborhood property around them. Here, a linearized version of (2-14), (2-15) is obtained as

$$\delta\dot{x}(t) = F\,\delta x(t) + G\,\delta u(t), \tag{2-16}$$

where

$$F = \frac{\partial f}{\partial x}\bigg|_{x=x^*},$$

$$G = \frac{\partial f}{\partial u}\bigg|_{u=u^*},$$

$$H = \frac{\partial h}{\partial x}\bigg|_{x=x^*},$$

with x^* a certain reference point. Lee and Markus[4] chooses x^* to be the origin under the null condition

$$f(0,0,0) = 0, \tag{2-18}$$

$$h(0,0) = 0, \tag{2-19}$$

and applied the rank test to the system (2-16), (2-17). Hwang and Seinfeld [5] extended the work of [4] to the arbitrary entire domain of the initial condition.

2) Observability rank condition

A geometric approach using Lie Algebra for the continuous [6] or discrete [7] nonlinear system is studied. Define

$$L_{f}i(h(x)) = \frac{\partial h}{\partial x} f(x,u^{i},t), \; i = 1, 2, \ldots, r,$$

where $f^{i}(x) = f(x(t),u_{i}(t),t)$ and L is closed under Lie algebra

$$L_{[f^1,f^2]}(h) = L_{f}1(L_{f}2(h)) - L_{f}2(L_{f}1(h)) .$$

Let $g(x)$ be the set with elements consisting of a finite linear combination of functions of the form

$$L_f^{1}(\ldots\ldots(L_f^{k}(h))\ldots\ldots),\ k=1,\ 2,\ \ldots,\ m$$

The Lie differential dg(x) is, then a finite linear combination

$$dg(x) = \{d(L_f^{1}(\ldots\ldots(L_f^{k}(h))\ldots))\},$$
$$= \{L_f^{1}(\ldots\ldots(L_f^{k}(dh))\ldots)\}. \qquad (2\text{-}21)$$

The observability rank condition is satisfied if dg(x) in (2-21) has rank n.

3) Taylor Series expansion [9]

The Taylor series expansion of (2-15) about an initial condition $x(t_o) = x_o$ at t_o is

$$y(t) = y(t_o)+y'(t_o)\,\Delta t + y''(t_o)\,\frac{\Delta t^2}{2} + \ldots\ldots,$$
$$= \sum_{i=0}^{\infty} y^{(i)}(t_o)\frac{\Delta t^i}{i!}\ . \qquad (2\text{-}22)$$

Define the collection of all the coefficients of (2-22) to be Y such that

$$Y = [y^{(i)}(t_o),\ i = 1,2,\ldots\infty\ ]^T,$$
$$= H(x_o). \qquad (2\text{-}23)$$

Then one-to-one relation of the function (2-23) is checked. In actual application $y^{(i)}(t_o)$, $i = 1,2,\ldots$ is checked if it is an even function in x_o.

4) Jacobian matrix approach

Observation equation $y(t)$ is differentiated with appropriate substitution according to the system equation (2-14) successively. Then the Jacobian matrix $J(.)$ evaluated at x_o is analyzed as follows;

i) Rank test of determinant $J(.)$ [10], [11]

or, equivalently nonzero of det J is tested [17].

ii) Ratio condition [13], [14], [15]

Ratio condition is satisfied if the absolute value of the leading principle minor of $J(.)$ is greater than $\varepsilon > 0$, i.e.,

$$|\Delta_1| = |\det J_1| > \varepsilon ,$$

$$\left|\frac{\Delta_2}{\Delta_1}\right| = \left|\frac{\det J_2}{\det J_1}\right| > \varepsilon ,$$

$$\left|\frac{\Delta_n}{\Delta_{n-1}}\right| = \left|\frac{\det J_n}{\det J_{n-1}}\right| > \varepsilon , \tag{2-24}$$

where J_i is obtained by taking only the first i rows and columns of J. Singh [14] checked the ratio condition for the matrix, AJ, where, A, is an arbitrary, $n \times mk$ matrix for the k-th derivation of $y(t)$ such that $mk \geq n$.

iii) Positive semidefinite of AJ [13], [14], [16].

Again A is an arbitrary $n \times mk$ matrix to make AJ to be $n \times n$

matrix. Then the system is said to be observable if one can find matrix A such that AJ is positive semidefinite.

iv) Minor matrix analysis of J [12].

Minor matrix of J matrix J_1, J_2, ..., J_{n-1} is constructed. Then for each J_i, an unobservable set D_i is obtained as

$$D_i = \{ x \mid J_i \neq 0,\ J_{i+1} = 0 \},\ i = 1, 2, \ldots, n-1. \qquad (2\text{-}25)$$

In spite of many results, it is found that some are insufficient [9] - [11], [13], [14], or too complicated to apply in practice [12], or applicable for only special class of nonlinear system such as in [18] or for linearized systems.

Introduced in the subsequent section is a new method which is simple to apply in practical problems and provides not only the test of observability of the system, but also, identifies the unobservable states when the system is unobservable. This approach is based on Palais' global implicit-function theorem [1] and its late versions [19], [20].

Modification of both the non-zero Jacobian condition and finite covering condition are required to be applied to the system observability. A modified version of the global implicit-function theorem is used in section three to demonstrate its simplicity and effectiveness by providing various examples including tracking of a maneuvering target where only bearing information is extracted from the measurement and array SONAR tracking problem with a small number of sensors.

2.2 A modified form of global implicit-function theorem

The most common inverse-function theorem guarantees only the existence of a local inverse in terms of the nonzero determinant of the Jacobian of the function f(.). The implicit-function theorem is an extension of this theorem to include argumented variables in it. The global versions of these theorems are the global inverse – function theorem and the global implicit-function theorem, respectively. Both theorems, in a global sense, require nonzero det J(.) and finite-covering conditions. It is shown here that both conditions can be modified further to be sufficient conditions for f to be invertible uniquely. I.e., without losing a global homeomorphic property of f, one can relax the nonzero Jacobian condition from the n dimensions to the n-1 dimensions for the special structure of f. However, the finite-covering condition needs to be added to the one-covering condition. The modified version of the global implicit-function theorem then will be used to determine the observability of the given nonlinear system. See Appendix A for the inverse and implicit function theorems and some related definitions.

Global versions of the local inverse and implicit function theorems are studied by several authors [25], [26], [27]. Here these theorems are restated without proof which can be found from cited references.

Theorem 2-1 Global inverse function theorem

Let f be an n real function of n real variables. The necessary and sufficient conditions that the function $f: R^n \to R^n$ defined by

$$f(x)=y, \qquad x \in R^n, y \in R^n$$

be a C^1 diffeomorphism of R^n onto itself are

i) each $f_i(x)$ is of class C^1,

ii) $\det Jf(x) \neq 0$,

iii) $\lim \| f \| = \infty$, as $\| x \| \to \infty$.

Theorem 2-2 Global implicit function theorem

Let f be a n real function of n + r real variables ($n \geq 1$, $r \geq 1$). Consider the function $f: R^n x\, R^r \to R^n$ such that

$$f(x,v)=y,$$

where $x \in R^n, v \in R^r$, $y \in R^n$ and f is C^1 in x and v. Then there exists a unique C^1 function such that

$$g: R^n x R^r \to R^n,$$

if

i) $\det Jf(.) \neq 0$ for all x and v, where $J = \partial f / \partial x$.

ii) $\lim \| f(x,v) \| = \infty$, as $\| x \| \to \infty$.

Condition iii) in the Theorem 2-1 or condition ii) in the Theorem 2-2 is called a "finite-covering" condition (see below).

Next it is shown that both the nonzero-Jacobian and the finite - covering conditions of both theorems are not enough for f to be one-to-one correspondence. Appropriate modification is required to provide sufficient conditions. Before a discussion is presented the following terms are defined.

Definitions [26], [31]

A cover for a set A is a collection v of sets such that $A \subset \bigcup_{V \varepsilon v} V$. Let X and Y each be connected spaces. If f maps X onto Y with the property that for each $y \varepsilon Y$ has an open neighborhood V such that each component of $u \varepsilon U$, $U = f^{-1}(V)$ is mapped homeomorphically onto V by f, then f is called a covering map. In this case if the cardinal number is n, then f is an n-covering map. If n is finite, then it is a finite-covering map, and if n=1, then it is a one-covering map .

Note that the finite covering condition excludes the possibility that f oscillates infinitely as $|| x || \to \infty$. With the above definitions, next two lemmas show that the homeomorphism of f (at least in a local sense) provides sufficiency for f to be a finite-covering function. But, the converse is not true (See Example 2-1).

Lemma 2-1 [27]

Let $f:X \mid \to Y$, $X \in R^n$, $Y \in R^n$ be a local homeomorphism. A necessary and sufficient condition that f be a finite covering is that

$$\lim_{\|x\| \to \infty} \| f(x) \| = \infty .$$

Lemma 2-2 [26]

Let $f:X \mid \to Y$, $X \in R^n$, $Y \in R^n$. If f is a homeomorphic function of R^n onto itself, then

$$\lim_{\|X\| \to \infty} \| f(X) \| = \infty .$$

Example 2-1

Consider the two-dimensional function f which is given by

$$f(x) = \begin{bmatrix} x_1^2+x_2^2 \\ \\ 2x_1^2+4x_2^2 \end{bmatrix} .$$

Then

$$\lim_{\|x\| \to \infty} \|f(x)\| = \lim_{x_1^2+x_2^2 \to \infty} \{(x_1^2+x_2^2)^2 + (2x_1^2+4x_2^2)^2\} = \infty .$$

Clearly the finite-covering condition is satisfied, but actual solution of the two equations yields

$$f_1(x) = x_1^2 + x_2^2 = y_1 ,$$

$$f_2(x) = 2x_1^2 + 4x_2^2 = y_2 ,$$

with non-unique solutions

$$x_1 = \pm \sqrt{\frac{4y_1-y_2}{2}} ,$$

$$x_2 = \pm \sqrt{\frac{y_2-2y_1}{2}} .$$

Thus f is only locally homeomorphic, i.e., f is not one-to-one globally. Both x_1 and x_2 are covered by the two "sheets" of cover. However, the existence of the two independent solutions is guaranteed by a nonzero determinant of the Jacobian,

$$\det Jf(x) = x_1 x_2 \neq 0,$$

i.e., with $x_1 \neq 0$ and $x_2 \neq 0$.

From the above two lemmas and example, it is clear that the finite-covering condition only provides a "weak" sufficient condition for f to be a homeomorphic function, globally.

Even though the global functions have played a fundamental role in many research works in nonlinear system studies, both the nonzero Jacobian and the finite covering conditions are not enough to provide sufficient conditions for f to be one-to-one correspondence. To discuss this more specifically next further definitions are made.

Definition

An individual finction $f_i(x)$, $i=1, 2, \ldots, n$ of f is called an absolutely independent function if it consists of only one coordinate of x, say x_j. x_j is called an absolutely independent variable.

A nonzero Jacobian condition provides functional independency and thus at most guarantees the existence of local inverses. But it does not say how many inverses exist, including the possibility of an infinte number which may appear when f involves trigonometric functions.

On the other hand,a finite covering condition furnishes a little narrower restriction to f than the nonzero Jacobian condition by excluding an infinite covering possibility, but still allows multiple coverings as well as functional dependence. So, we must modify both conditions as follows. In case f has absolutely independent functions, f(.) can still hold functional independence even if $\det Jf(.) = 0$ as far as $\det Jf^{-}(.) \neq 0$, where f^{-} denotes the remaining portion of f while deleting one absolutely independent function from f. The next example shows that f can be functionally independent, and thus can have a global inverse in spite of $\det Jf(.) = 0$ as far as $\det Jf^{-}(.) \neq 0$.

Example 2-2

Let $f:R^3 \rightarrow R^3$ is defined by

$$f(x) = \begin{bmatrix} x_1^3 \\ x_2^3 \\ x_1 + x_2 + x_3 \end{bmatrix} = y \quad .$$

The function has a global inverse on R^3 as

$$x_1 = (y_1)^{1/3} \quad ,$$

$$x_2 = (y_2)^{1/3} \quad ,$$

$$x_3 = y_3-(y_1)^{1/3}-(y_2)^{1/3} \quad .$$

Hence f is a homeomorphic - onto function unless

$$\det Jf(x) = 9x_1^2x_2^2 = 0 \ ,$$

by $x_1 = 0$ and $x_2 = 0$.

Det $Jf(x) = 0$ is allowed either by $x_1 = 0$ or $x_2 = 0$ without loosing functional independence. Note that both x_1 and x_2 are absolutely independent variables.

Thus the nonzero-Jacobian condition can be weakened to (n-1) dimensions instead of n dimensions in the special form of f. Meanwhile a finite-covering condition must be modified to a one-covering condition instead of finite-covering condition. But neither one is not enough for f to be a globally homeomorphic function since a nonzero-Jacobian condition alone lacks globallity of the inverse and the one-covering condition alone lacks independency of f. Consequently,we have the following adaptation of the previous theorem.

<u>Theorem 2-3</u>

Let $f: x \to y$, $x \in R^n$, $y \in R^n$ be an onto C^1 function. f is globally homeomorphic x onto y if

i) $\det Jf(x) \neq 0$ for all x

($\det Jf^{-}(x) \neq 0$ if f contains absolutely independent functions)

ii) f(x) is a one-covering function for all x.

<u>Proof</u>

We need to prove that the two conditions mean a global homeomorphism of f. First, consider for the case when f has no absolutely independent functions. Then by the inverse function theorem f is a local homeomorphism from x to y. So, by addition of restriction U on f, $f_{|U}(x)$ is one-to-one from onto y. Next if f has some absolutely independent function, then $\det Jf^{-}(x) \neq 0$ provides a local homeomorphism from x^{-} to y. The function f_i which is excluded from f is already independent from f^{-}; thus f_i is at least locally homeomorphism from condition ii). So, f is locally homeomorphic and again restriction U exists such that f be one-to-one from U to y. Hence if we can show that U=x, then proof will be completed. Suppose U is a proper subset of x. Since U is open in x, U is an open proper subset of x. Let χ be a boundary point of U, and V be an open connected neighborhood of $f(\chi)$. Since f is a one-covering map on x, $f^{-1}(V)$ is not empty and consists of one component. Let N_χ denote this component. Surely N_χ contains χ. Let $N_\chi^* = U \cap f^{-1}(V)$. Since f is continuous f^{-1} is open. Hence both N_χ and N_χ^* are open and connected.

Also note that f maps both N_χ and N^*_χ onto V. Since N_χ is open and contains x, the set $N_\chi \cap U$ is also not empty. It follows that $N_\chi \cap N^*_\chi$ is not empty, otherwise there will be at least one point x_1 in $N_\chi \cap U$ a point x_2 in N^*_χ such that $f(x_1) = f(x_2) \varepsilon V$, and $f_{|U}$ will not be one-to-one on U which constitutes contradiction. Hence, $N_\chi = N^*_\chi$, i.e., x is in N^*_χ and, therefore, is in U . This implies U can't be an open proper subset of x. That is U is closed in x. So, U is both open and closed in x and nonempty. Therefore U = x. **

Remarks

1. Globally homeomorphic from x to y is indentical to global one-to-one correspondence and continuity [30].
2. Every homeomorphic onto function is a covering map, and every covering map is locally homeomorphic.
3. Even a nonzero-Jacobian condition can be relaxed to n-1 dimensions. Here n dimensions will be assumed in the general discussion since $\det Jf \neq 0$ always includes $\det J\bar{f} \neq 0$.

Lemma 2-3

If every entry of the Jacobian J of f does not make any sign change along the real line of x, then f is globally a one-covering map.

Proof

Entry $J_{ij} = \dfrac{\partial f_i}{\partial x_j}$, i,j = 1, 2, ..., n is variation of function f_i with respect to j-th direction of x. If f_i does not make any sign change due to x_j, then f_i is monotone in j-th direction, i.e., f_i is one-covering function with respect to x_j. If every function does not have any sign change in any direction, then f is a one-covering function globally. **

In order to be a multiple-covering function in any direction, the slope of a corresponding entry must be changed due to that direction. Then the number of possible covers are one plus the number of sign changes. The nonzero-Jacobian condition may be combined to constitute one method to determine one-to-one correspondence of f. See Theorem 2-4 below.

<u>Lemma 2-4</u>

If the Jacobian J of f(x) is either positive or negative definite for all x, then f(x) is a global one-covering map.

<u>Proof</u>

Proof for the part of the positive definite case is given in [19]. Negative definite case can be proven similarily. **

In Lemma 2-4, the nonzero-Jacobian condition is already implied hence not required here. A modified version of the global inverse function theorem allows us to adopt the global implicit function theorem as follows;

<u>Theorem 2-4</u>

Consider $f: x \times u \rightarrow y$, $x \in R^n, u \in R^r$, $y \in R^n$ such that

$f(x,u) = y$.

Suppose f is C^1 function in x and u. If f satisfies the following two conditions,

i) $\det Jf(.) \neq 0$ for all x.

ii) $f(x,u)$ is a one-covering map on all x, then there exists a unique continuous function g such that

$$x = g(y,u). \tag{2-27}$$

Proof

Define a vector $\hat{x}$ and vector-valued function $\hat{f}$ as

$$\hat{x} = \begin{bmatrix} x \\ u \end{bmatrix}$$

$$\hat{f} = \begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \end{bmatrix} = \begin{bmatrix} f(x,u) \\ u \end{bmatrix},$$

$$= \begin{bmatrix} y \\ u \end{bmatrix} = \hat{y}, \tag{2-28}$$

which maps R^{n+r} onto itself. Obsviously $\hat{f}$ is continuously differentiable with respect to $\hat{x}$ and its Jacobian matrix is

$$J\hat{f}(\hat{x}) = \left[\frac{\partial \hat{f}}{\partial \hat{x}}\right],$$

$$= \begin{bmatrix} \dfrac{\partial \hat{f}_1}{\partial x} & \dfrac{\partial \hat{f}_1}{\partial u} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial u} \end{bmatrix},$$

$$= \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial u} \\ 0 & I_r \end{bmatrix}, \qquad (2\text{-}29)$$

where I_r is an identity matrix with dimension r. Since $\dfrac{\partial f}{\partial x} \neq 0$, det $\left(\dfrac{\partial \hat{f}}{\partial \hat{x}}\right) \neq 0$ from (2-29). And since $\hat{f}_1 = f(x,u)$ is a one-covering map on x, and $\hat{f}_2 = u$ is also a one-covering map on u, $\hat{f}(\hat{x})$ is a one-covering map on $\hat{x} = \begin{bmatrix} x \\ u \end{bmatrix}$. Therefore by the Theorem 2-3, there exists a globally continuous function $g = \hat{f}^{-1}$ such that

$$g(\hat{y}) = \hat{x}, \qquad (2\text{-}30)$$

i.e.,

$$\hat{x} = \begin{bmatrix} x \\ u \end{bmatrix} = g(\hat{y}),$$

$$= \begin{bmatrix} g(y,u) \\ g_1(y,u) \end{bmatrix}, \qquad (2\text{-}31)$$

for all $\hat{y} \varepsilon R^{n+r}$.

Take the first n equations from (2-31).

$$x=g(y,u), \tag{2-32}$$

which is also a globally continuous function mapping from R^{n+r} into R^n. **

As shown a nonzero-Jacobian determinant guarantees the existence of a local homeomorphic inverse, i.e., provides the "connectedness" of every component of x to Y, the measurement space. But the connection may not be necessarily unique. For this reason nonzero-Jacobian condition will be called "connectedness condition" in the observability problem which will be discussed in the next section.

A one-covering condition, on the other hand, provides the uniqueness of the connection globally. So, the one-covering contiion will be called the "univalence condition" in the observability problem. Heuristiclly, Theorem 2-4 says that the mapping (2-26) is a one-to-one correspondence globally if every x_i, i = 1, 2, ..., n can be expressed uniquely in terms of only Y and u for all x.

With this background about the nonlinear functions, observability of nonlinear systems is studied next.

2-3. Observability of Nonlinear systems

State and observation equations are given, again, as

$$\dot{x}(t) = f(x(t),u(t),t), \tag{2-33}$$

$$y(t) = h(x(t),t). \tag{2-34}$$

As assumed earlier f(.) satisfies necessary conditions to guarantee the existence and uniqueness of the solution x(t). Further it is assumed that h(.) is differentiable up to (n-1)-th order with respect to t. Then, define system observability as follows.

Definition

System (2-33), (2-34) is observable at t_o if knowledge of the input u(t) and the output y(t), $t \varepsilon [t_0, t_1]$ is sufficient to determine $x(t_o)$ uniquely for finite t_1. If every state $x(t) \varepsilon R^n$ is observable on the time interval $[t_o, t_1]$, then the system is completely observable.

Note here that due to the assumption of the existence and uniqueness of the solution in (2-33), x(t) can be uniquely determined from proper construction of the integral operator g(.) as in (2-3)

$$x(t) = g(x(t_o),\ u(t),t), \tag{2-3}$$

once $x(t_o)$ is known.

So, the definition of the $x(t_o)$-observability above implies, also, x(t)-observability for the considered time interval $t \varepsilon [t_o, t_1]$.

Next, to derive more definitions on the system, differentiates (2-34) with respect to t and makes appropriate substitution (2-33) (with suppression t in the variables)

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

$$y' = \frac{\partial h}{\partial t} + \frac{\partial h}{\partial x}\frac{\partial x}{\partial t} = h_t + h_x f \ ,$$

$$= h_1(x,u,t) \ .$$

$$y'' = \frac{\partial h_1}{\partial t} + \frac{\partial h_1}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial h_1}{\partial u}\frac{\partial u}{\partial t} = h_{1t} + h_{1x} f + h_{1u} u' \quad ,$$

$$= h_2(x,u,u',t) \ .$$

$$\vdots$$

$$y^{(n-1)} = h_{(n-2)t} + h_{(n-2)x} f + h_{(n-2)u} u' + , \ldots , + h_{(n-2)u^{(n-3)}} u^{(n-2)}$$

$$= h_{n-1}(x,u,u',\ldots,u^{(n-2)},t) \quad , \tag{2-35}$$

where $y^{(i)}$ denotes i-th time derivatives of $y(t)$.

Define an mn-dimensional vector Y, <u>measurement vector</u> of the system (2-33), (2-34) as the left hand side of (2-35), i.e.,

$$Y = [y, y', y'', \ldots\ldots, y^{(n-1)}]^T \ , \tag{2-36}$$

and an mn-dimensional function H(.), <u>measurement function</u> of (2-35) as

$$H(.) = [h, h_1, h_2, \ldots\ldots, h_{n-1}]^T \ . \tag{2-37}$$

Then one obtains an mn-functional relation in vector form

$$Y = H(x,v,t) \quad , \tag{2-38}$$

where $v(t)$ is a function $u^{(i)}$, $i=1,2,\ldots,n-2$.

From equation (2-38) next can be proved.

Theorem 2-5

If every state $x(t_o)$ is uniquely determined from (2-38), then the system (2-33), (2-34) is observable at t_o.

Proof

The proof will be completed if one can show that the unique determination of every state $x(t_o)$ from (2-38) is equivalent to that every state is uniquely determined from the measurement $y(t), t\varepsilon [t_o,t_1]$.

Let us expand the function y(t) in a Taylor series for any $t\varepsilon [t_o,t_1]$ at t_o

$$y(t) = y(t_o)+y'(t_o)(t-t_o)+0.5y''(t_o)(t-t_o)^2+,\ldots,+ \frac{1}{(n-1)!}y^{(n-1)}(t_o)(t-t_o)^{n-1} + r(t) . \qquad (2\text{-}39)$$

Since the Taylor-series expansion of an arbitrary function is unique, each coefficient $y^{(i)}(t_o)$, $i = 1, 2, \ldots, n-1$ is also unique. So, once y(t) is determined, then $y^{(i)}(t_o)$ is determined uniquely. However, each coefficient of (2-39) is an exact element of the measurement vector Y in (2-38). Therefore, if $x(t_o)$ is uniquely solveable in terms of Y, V and t in (2-38), then the system is observable at t_o by the definition. **

Thus, the observability problem of the system is equivalent to find the condition under which (2-38) has a unique inverse about state

$x(t)$. Or geometrically, the system is observable if the mapping (2-38) is one-to-one from the state space $x \varepsilon R^n$ into or onto the measurement space $Y \varepsilon R^{mn}$ for all $t \varepsilon [t_o, t_1]$. (See Figure 1.)

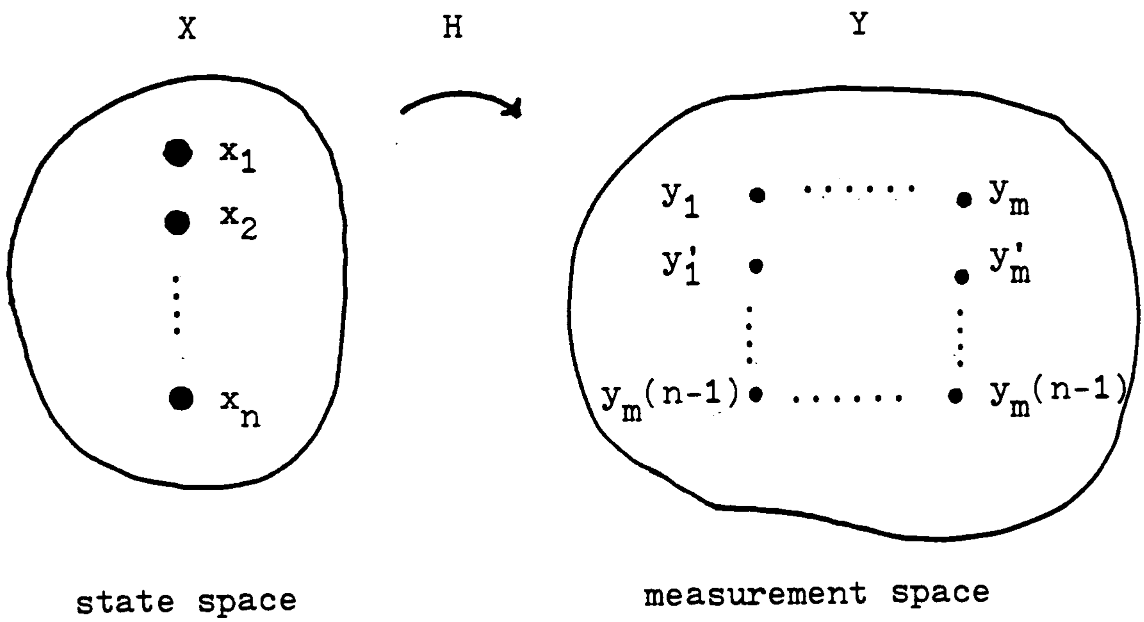


Figure 1, Geometric interpretation of system observability

So, from the functional analysis results of the previous section and Theorem 2-5, the system is observable if the following two conditions are satisfied.

1. <u>Connectedness</u>

Every state x_i, $i = 1, 2, \ldots, n$ must be connected to any elements of measurement space Y, i.e., (2-38) constitutes n independent function with respect to x in time interval $t \varepsilon [t_o, t_1]$.

2. Univalence

Further, every state x_i, $i = 1, 2, \ldots, n$ must be connected uniquely to the measurement space Y.

As mentioned earlier, the first condition is related to the functional independency and thus nonzero Jacobian condition of (2-38) and the second condition is related to the one-covering condition. Before applying Theorem 2-4 it is required to rearrange (2-38) to reduce computational complexity as follows. This procedure helps to maximize the functional independence before applying the non-zero Jacobian condition by deleting functionally dependent elements from the mn functional H.

$$y = h(x,t), \tag{2-40}$$

$$y' = h_1(x,u,t). \tag{2-41}$$

By appropriate replacment of $h_1(.)$ by $h(.)$ one can obtain

$$y' = h_{1a}(y,x,u,t). \tag{2-42}$$

Repeating this procedure up to (n-1)th order gives

$$y'' = h_{2a}(y,y',x,u,u',t),$$

$$\vdots$$

$$y^{(n-1)} = h_{n-1,a}(y,y',\ldots,y^{(n-2)},x,u,u',\ldots,u^{(n-2)},t) . \tag{2-43}$$

Denote Y^- the set consisting of

$$Y^{-} = \{y, y', y'', \ldots, y^{(n-2)}\} , \tag{2-44}$$

and

$$V = (u, u', u'', \ldots, u^{(n-2)}) . \tag{2-45}$$

Then the vector notation of (2-42), (2-43) becomes

$$Y = H_a(Y^{-}, x, V, t) . \tag{2-46}$$

Successive replacement of lower order derivatives to the higher order derivatives as in (2-43) means minimizing functional dependence between the individual functional elements h, $h_1, \ldots, h_{n-1}$ since the procedure is exactly the same as the successive elimination of unknown variables in solving (2-38) for x. Thus maximum independence between functional elements is obtained. Next let

$$p = (Y^{-}, V, t),$$

then (2-46) becomes

$$Y = Ha(x, p) . \tag{2-47}$$

with (2-47) and Theorem 2-4 determination of the system observability can be made using the following result.

<u>Main Result</u>

System (2-33), 2-34) is observable (in the strict sense) if (2-47) satisfies the following two conditions for all $t \in [t_o, t_1]$.

i) Connectedness condition

$$\det JH_a^n(.) \neq 0 , \qquad (2\text{-}48)$$

where $JH_a^n = \frac{\partial}{\partial x} H_a^n$, and H_a^n is any subset of H_a consisting of n functions.

ii) Univalence condition

For the chosen $H_a^n(.)$, every state x_i, i = 1, 2, ..., n can be uniquely expressed in terms of only Y and p.

The assertation is obvious from the Theorem 2-4 and 2-5. Actual proof is similar to the proof of the Theorem 2-4.

Depending on the satisfaction of the conditions i) and/or ii) of the result, define and categorize system observability as follows:

1. Observable in the strict sense.

Both of the two conditions are satisfied for at least one combination of H_a^n out of mn function H_a.

2. Observable in the wide sense.

Only the connectedness condition is satisfied for any one or more states, i.e, multiple covering appears in any component of x for any time t.

3. Unobservable

One ore more components of x cannot be expressed in terms of Y and P. In this case these states are unconnected to Y and thus the system is unobservable.

The above observability determination is demonstrated by the following examples.

<u>Example 2-3</u>

A falling body in the constant gravity field with position variable x_1 and velocity x_2 can be expressed as

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -g, \qquad g \text{ is constant.}$$

If one measures position x_1, then

$$y = x_1, \text{ and}$$
$$y' = \dot{x}_1 = x_2.$$

So, both states are uniquely determined from $Y = (y, \; y')^T$, and hence the system is observable. On the other hand if velocity x_2 is measured, then

$$y = x_2,$$
$$y' = \dot{x}_2 = -g.$$

Only x_2 is connected uniquely to Y. x_1 is disconnected and unobservable; hence the system is unobervable. Classic rank test can be used to verify this.

Example 2-4

$$\dot{x}_1 = x_1 + u ,$$
$$\dot{x}_2 = 2x_1 - x_2 + 3x_3 + 2u ,$$
$$\dot{x}_3 = x_3 ,$$

$$y = 2x_2 + x_3 ,$$

then

$$y' = 4x_1 - 2x_2 + 7x_3 + 2u ,$$
$$y'' = 2x_2 + x_3 = y .$$

Only x_2, x_3 can be obtained uniquely if x_1 is given, i.e., x_1 is unobservable. Decoupling procedures show that x_1 is unobservable.

Example 2-5

A gyrocompass precessional motion is descrived as [17]

$$\dot{x}_1 = ax_2 + bx_3 , \quad a>0, \quad b=a(1-\rho), \; 0<\rho<1 ,$$
$$\dot{x}_2 = -cx_1 - dx_1^3 ,$$
$$\dot{x}_3 = -Fx_2 - Fx_3 , \qquad F>0 ,$$

$y = x_1$, then (2-49)

$$y' = ax_2 + bx_3 , \tag{2-50}$$

$$y'' = -acy - ady^3 - bF(x_2 + x_3) . \tag{2-51}$$

$$\det J = bF(b-a) \neq 0 . \tag{2-52}$$

From (2-49) ~ (2-51)

$$x_1 = y,$$

$$x_2 = \frac{-(acy + ady^3 + Fy' + y'')}{F(b-a)},$$

$$x_3 = \frac{bFy' + a(acy + ady^3 + y'')}{bF(b-a)},$$

Clearly, all the states are observable from the last three equations. So, the system is observable.

Example 2-6 [9],[13]

$$\dot{x}_1 = x_2 x_3,$$
$$\dot{x}_2 = -x_1 x_3,$$
$$\dot{x}_3 = 0.$$

$$y = x_1, \quad \text{then} \tag{2-53}$$

$$y' = x_2 x_3, \tag{2-54}$$

$$y'' = -x_1 x_3^2 = -y x_3^2. \tag{2-55}$$

So, $\det J = 2x_1 x_3^2 \neq 0$ implies that the initial state of the form $\{x_{10} \neq 0, x_{30} \neq 0\}$ satisfies the connectedness condition. But from (2-53) to (2-55),

$$x_1 = y,$$

$$x_2 = \pm y' / \sqrt{\frac{y''}{-y}},$$

$$x_3 = \pm \sqrt{\frac{y''}{-y}}.$$

x_2 and x_3 have multiple expressions or two covers. So, the univalence condition is not satisfied. The system is only observable in the wide sense if $\{x_{10} \neq 0,\ x_{30} \neq 0\}$.

Example 2-7 [12]

$$\dot{x}_1 = x_2,$$
$$\dot{x}_2 = -2x_1 - 3x_2 - x_1^3 x_3,$$
$$\dot{x}_3 = -x_3 x_4,$$
$$\dot{x}_4 = 0.$$

$$y = x_1. \tag{2-56}$$

So,

$$y' = x_2, \tag{2-57}$$

$$y'' = -2x_1 - 3x_2 - x_1^3 x_3 = -2y - 3y' - y^3 x_3, \tag{2-58}$$

$$y''' = -2y' - 3y'' - 3y^2 y' x_3 + y^3 x_3 x_4 \tag{2-59}$$

$\det J = -y^6 x_3 = -x_1^6 x_3 \neq 0$ implies the connectedness is satisfied when $\{x_{10} \neq 0,\ x_{30} \neq 0\}$. Here, note that (2-56), (2-57) are absolutely independent functions. So, det J=0 is allowed as far as $\det J^- \neq 0$, where J^- is the Jacobian after deleting any one of the two absolutely independent functions. In this case only

$$x_{20} = y'(t_o) = 0.$$

is allowed since $x_{10} = 0$ makes $\det J^- = 0$..

From (2-56) ~ (2-59)

$$x_1 = y,$$

$$x_2 = y',$$

$$x_3 = \frac{-(2y+3y'+y'')}{y^3},$$

$$x_4 = \frac{-(2y'+3y''+y''')}{2y+3y'+y'} + \frac{3y'}{y}.$$

Obviously, the univalence condition is satisfied. So, the system is observable if $\{x_{10} \neq 0,\ x_{30} \neq 0\}$ is preserved.

Two practically more important examples are shown in the next section which will be used also for stochastic system observability.

2-4 BOT and array SONAR tracking examples

System observability determination of two important examples in underwater tracking are demonstrated here. The first example is a bearing-only-target tracking problem where only bearing information of the target is extracted from the measurement device and used to determine the observability of the other state variables as well as whole system observability.

Consider an object or target (T) and observer or ownship (O) configuration as in Figure 2. When T and/or O move with velocity components v_{Tx}, v_{Ty}, v_{Ox}, v_{Oy}, relative coordinates $x(t)$ and $y(t)$ can be generated as

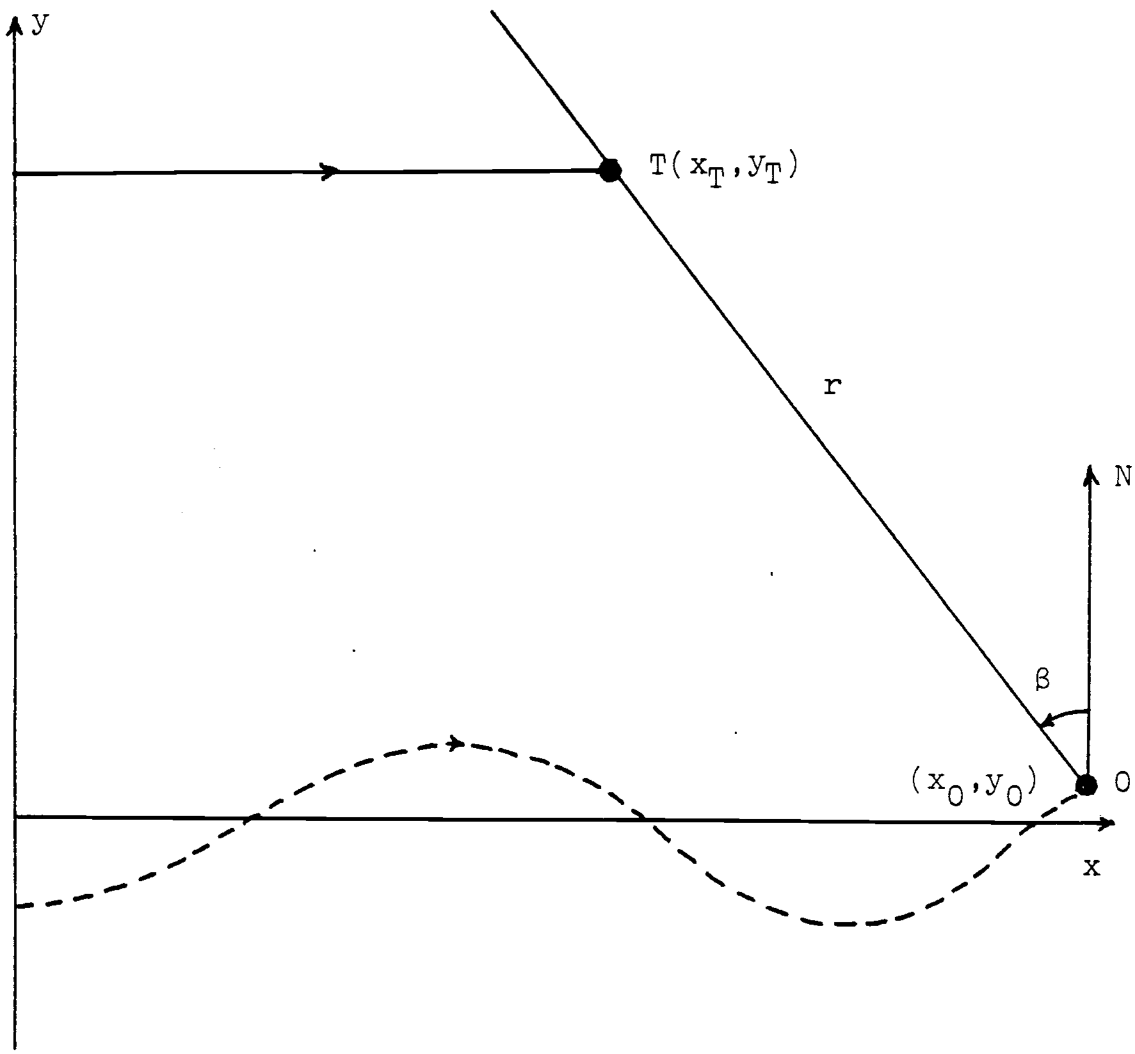


Figure 2, BOT configuration

$$x(t) = x_T(t)-x_O(t), \tag{2-60}$$

$$y(t) = y_T(t)-y_O(t). \tag{2-61}$$

Define the state variables in mixed coordinates which consists of mixed components of polar and rectangular coordinate as

$$x_1(t)= \beta(t), \tag{2-62}$$

$$x_2(t)= r(t), \tag{2-63}$$

$$x_3(t)= v_{Tx}(t)-v_{Ox}(t) = v_x(t), \tag{2-64}$$

$$x_4(t)= v_{Ty}(t)-v_{Oy}(t) = v_y(t), \tag{2-65}$$

where $\beta(t)$ is bearing of T from O with respect to some reference, North N here, and $r(t)$ is range. Then from the relations

$$x(t) = r(t)\sin\beta(t), \tag{2-66}$$

$$y(t) = r(t)\cos\beta(t), \tag{2-67}$$

and their derivatives with proper algebra, the state equation in this coordinate system becomes

$$\dot{x}(t) = \begin{pmatrix} \dfrac{x_3\cos x_1 - x_4\sin x_1}{x_2} \\ x_3\sin x_1 + x_4\cos x_1 \\ a_x \\ a_y \end{pmatrix}, \tag{2-68}$$

where $a_x(t)$, $a_y(t)$ are accelerations in their directions. Due to bearing measurement the observation equation is

$$y(t) = [1 \quad 0 \quad 0 \quad 0]x(t) . \tag{2-69}$$

To make the system simpler, it is assumed that $a_x(t) = 0$, $a_y(t) = a(t) \neq 0$ in (2-68), i.e., maneuvering exists only in x-direction. Then successive replacements yield

$$y = x_1, \tag{2-70}$$

$$y' = \frac{x_3 \cos y - x_4 \sin y}{x_2} , \tag{2-71}$$

$$y'' = \frac{-(a.\sin y + 2y' \cos y.x_4 + 2y' \sin y.x_3)}{x_2} , \tag{2-72}$$

$$y''' = \frac{3ay' \cos y + [3y'' \sin y + 2(y)^2 \cos y]x_3 + [3y'' \cos y - 2(y')^2 \sin y]x_4 + a' \sin y}{x_2} . \tag{2-73}$$

So, from (2-70)-(2-73)

$$x_1 = y, \tag{2-74}$$

$$x_2 = \frac{-2y'x_4 - a\cos y.\sin y}{y'' \cos y + 2(y')2\sin y} , \tag{2-75}$$

$$x_3 = \frac{[y'' \sin y - 2(y')^2 \cos y]x_4 - ay' \sin y}{y'' \cos y + 2(y')2\sin y} , \tag{2-76}$$

$$x_4 = \frac{a[4(y')^3\cos y\sin y+6y'y''\cos^2 y-3y'y''-y'''\cos y\sin y]+a'\sin y[y''\cos y+2y')^2\sin y]}{2y'y'''-3(y'')2+4(y')4}$$

,(2-77)

From (2-77) it is clear that x_4 is connected to the measurment vector Y and it is unique when a(t) and/or a'(t) are nonzero, i.e., maneuvering exists. This implies from (2-75) and (2-76) that x_2 and x_3 are also uniquely connected to Y. So, the system satifies the connectedness condition if T and/or O maneuver. But when a(t) = 0, a'(t) = 0, i.e., when non-maneuvering, (2-77) says that x_4 is not connected to Y and is unobservable. This causes again from (2-75) and (2-76) that x_2, x_3 are disconnected from Y, and thus these states are unobservable from Y. Only x_1 is observable ,in this case,which is itself a measurement variable. After lengthy computation, the determinant of the Jacobian becomes

$$\det J = \frac{-2a'y'\sin y+3a[2(y')^2\cos y+y''\sin y]-[12y'y''\sin y(1+\cos^2 y)+}{x_2^4}$$

$$\underline{8\cos^3 y(y')^3]x_3+4y'\cos y\sin y[2(y')^2\cos y+3y''\sin y]x_4} \quad . \qquad (2\text{-}78)$$

From (2-78) the system is unobservable with det J = 0 for the following cases:

1. Infinite range, $x_2 = \infty$,
2. Non-maneuvering, $x_3 = x_4 = 0$ with a(t) = a'(t) = 0 (Including parallel stationary movement and tail chasing.),

3. Zero heading rate and acceleration, $\beta'(t) = \beta''(t) = 0$,

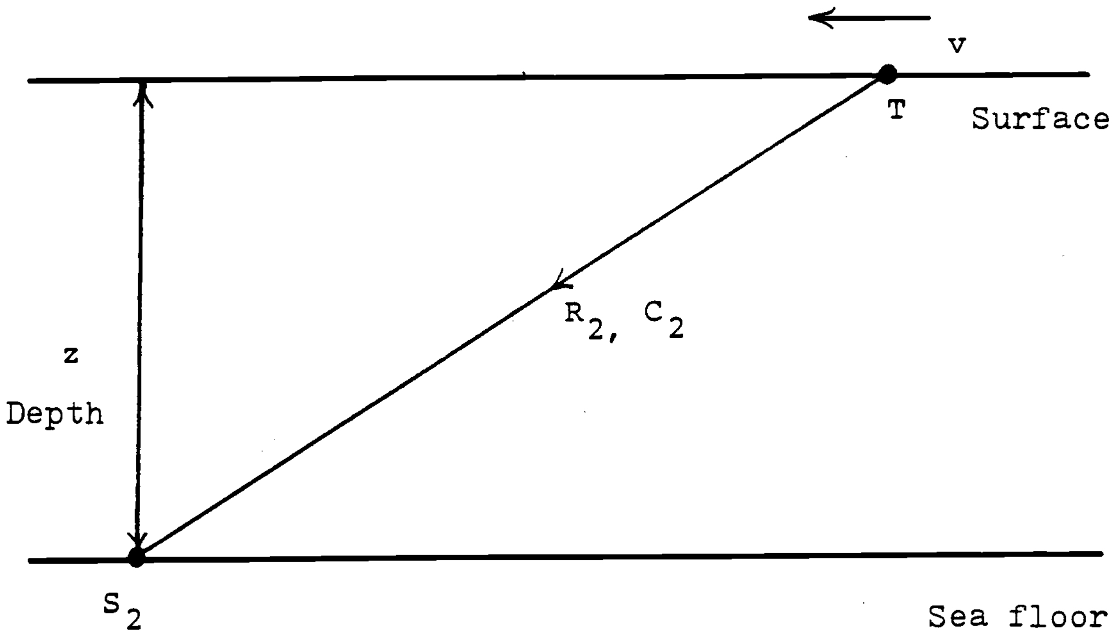
4. Constant range with special heading such that

$$\tan\beta = \frac{6\,a(\beta')^2}{2a'\beta' - 3a\beta''} \quad . \tag{2-79}$$

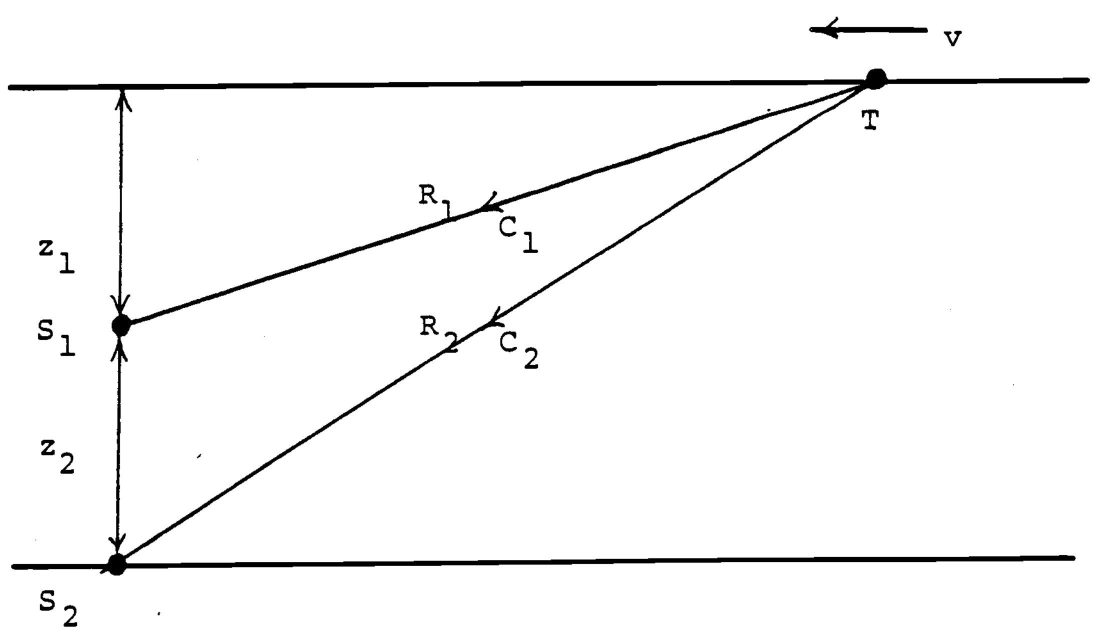
As well as certain others, the system is unobservable due to the lack of rank when any one or more conditions of above are satisfied. Consequently, from (2-74)-(2-78), it is shown again that for BOT tracking, the system is observable only when maneeuvering exists.

The second applicational example is the underwater SONAR tracking problem where the number of sensors, deployment and measurement schemes are changed. For good system observability, the number of sensors and their configuration are very important. Further, with the same number of sensors and the same deployment structures, measurement policy is even more important for many cases. One can measure absolute wave-propagation time-delay between the target and sensor or time delay difference between the two sensors, Doppler or Doppler difference or any combination thereof. Each of these measurement policies requires different observability analysis. Deployment can be considered as either horizontal (towed linear array) or vertical to the surface (vertically planted array). Figure 3 shows sensor and target configuration for up to three sensors which are deployed vertically. Only directly propagated wave is considered here. In the one-sensor case, only absolute time delay or absolute Doppler shift between T and S_2 can be measured. It implies that synchronization of

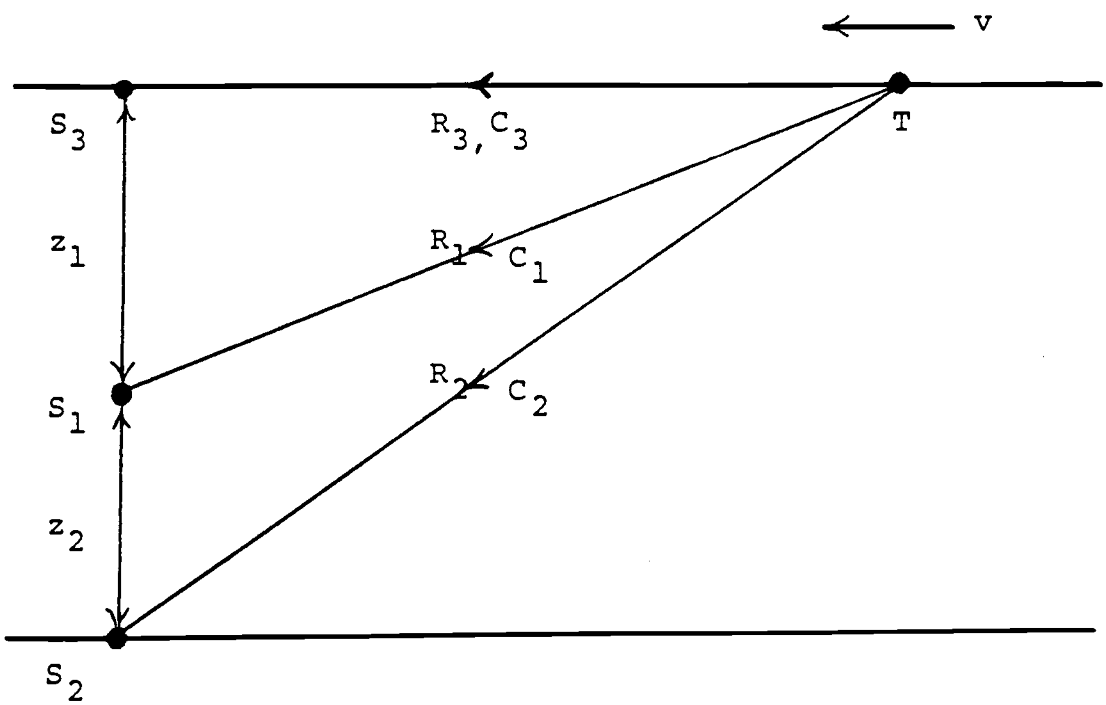
Figure 3 , Sensor configuration



1). One-sensor



2). Two-sensor



3). Three-sensor

T and S_2 is required for the passive case or can be used for only active SONAR case.

In two-sensor measurement, either absolute quantities or comparative differencies of intersensor delay and/or Doppler can be measured.

Here it is assumed that three measurement policies occur .

1. One relative delay; 2S1D
2. One relative Doppler; 2S1P
3. One relative delay and Doppler; 2S1D1P

In the three-sensor deployment, several possible measurements are considered as follows:

1. Two relative delay; 3S2D
2. Three relative delay; 3S3D
3. Two relative delay and one Doppler; 3S2D1P

Of course, more than three sensors can be considered. But it is known that [68] for optimal range and bearing estimation in sense of a minimum uncertainty ellipse, the best array configuration of M sensors is three groups of M/3 sensors each with equal spacing between groups. In this case, all sensors in a "pod" are assumed to be in the same location, i.e., there is no delay between sensors in the same group. Equally spaced M sensors showed much inferior performance than the three clusters of M/3 sensors except $M \to \infty$. So, the number of sensors considered here are limited up to three.

In a two-dimensional coordinate system, at least four states are required to describe the motion of the point target - two for position and two for velocity in each direction, respectively. Since sound speed varies quite significantly with depth, salinity, temperature - specially in coastal inlets [64], [69], [70] it affects the time delay and the Doppler shift. So, it is considered as a state variable, also.

I.e., define the state variables as follows:

x_1 is target position in x-direction,

x_2 is target velocity in x-direction,

x_3 is target position in y-direction,

x_4 is target velocity in y-direction,

x_5 is C_1 (accoustic wave speed in R_1),

x_6 is C_2 (accoustic wave speed in R_2).

With the above state the system equations can be written as (under the assumption of constant wave speed in depth)

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} x(t). \tag{2-80}$$

The basic measured quantities are time delay difference t_{ij} between sensors i and j, and Doppler frequency shift difference f_{ij} from carrier frequency f_c = 3500 Hz, which seems widely used in practical SONAR systems. So, for example, if two delay and one Doppler shift is measured with three sensors (3S2D1P), the observation equation becomes

$$y(t) = \begin{bmatrix} \tau_{12}(t) \\ f_{12}(t) \\ \tau_{32}(t) \end{bmatrix},$$

$$= \begin{bmatrix} \dfrac{R_2(t)}{C_2(t)} - \dfrac{R_1(t)}{C_1(t)} \\[2ex] f_c\{\dfrac{\dot{R}_2(t)}{C_2(t)} - \dfrac{\dot{R}_1(t)}{C_1(t)}\} \\[2ex] \dfrac{R_2(t)}{C_2(t)} - \dfrac{R_3(t)}{C_3} \end{bmatrix},$$

$$= \begin{bmatrix} \dfrac{(x_1^2+x_3^2)^{1/2}}{x_6} - \dfrac{(x_1^2+(x_3-z_2)^2)^{1/2}}{x_5} \\[2ex] \dfrac{f_c(x_1x_2+x_3x_4)}{x_6(x_1^2+x_3^2)^{1/2}} - \dfrac{f_c(x_1x_2-(x_3-z_2)x_4)}{x_5(x_1^2+(x_3-z_2)^2)^{1/2}} \\[2ex] \dfrac{(x_1^2+x_3^2)^{1/2}}{x_6} - \dfrac{x_1}{c_3} \end{bmatrix},$$

$$= h(x(t), f_c, C_3), \tag{2-81}$$

where surface sound speed C_3 is assumed to be a known value.

The other cases of measurement equations have a similar form except measuring different quantities. Therefore, in all cases, the system equations are simple linear equations if nonlinear drag, etc., are neglected. But the observation equations are nonlinear.

To oberve deterministic observability for this system, categorize the measurement scheme into three groups for convenience as

1. An absolute dèlay; 1S1D
2. Pure relative delay; 2S1D, 3S2D, 3S3D
3. Relative Doppler; 2S1P, 2S1D1P, 3S2D1P

The first case for an absolute time propagation delay of the acoustic wave with one-sensor deployment gives the observation equation as

$$y(t) = \frac{R_2(t)}{C_2(t)} \quad . \tag{2-82}$$

Considering system equation (2-80) and the relation (with omission of time variable t)

$$R_2 = (x_1^2 + x_3^2)^{1/2} \ ,$$

$$\dot{R}_2 = \frac{x_1 x_2 + x_3 x_4}{R_2} \quad . \tag{2-83}$$

Then, by algebraic manipulation

$$y = \frac{R_2}{x_6} = \frac{(x_1^2 + x_3^2)^{1/2}}{x_6} \quad , \tag{2-84}$$

$$y' = \frac{x_1x_2 + x_3x_4}{x_6R_2} \quad , \tag{2-85}$$

$$y'' = \frac{x_2^2 + x_4^2}{x_6R_2} - \frac{(y')^2}{y} \quad . \tag{2-86}$$

Let

$$A = \frac{(y')^2}{y} \quad ,$$

$$B = 2(y')^2 - yy'' \quad ,$$

then,

$$y''' = \frac{-y'(x_2^2 + x_4^2)}{R_2^2} - A' \quad , \tag{2-87}$$

$$y^{(4)} = \frac{(x_2^2 + x_4^4)B}{R_2^2 y} - A'' \quad , \tag{2-88}$$

$$y^{(5)} = \frac{(x_2^2 + x_4^2)}{R_2^2 y_2} (yB' - 3By') - A''' \quad . \tag{2-89}$$

From (2-84)-(2-89), it is clear even before solving them for x that x_5 does not appear in any equation, explicitly. So, x_5 is not connected to the measurement vector Y,

$$Y = \{y, y', \ldots., y^{(5)}\} .$$

Obviously x_5 is unobservable, and makes the system unobservable deterministically. Actual solution of these equations shows that other variables have multiple solutions, i.e., they are connected to Y multiply, thus they are observable at least in a wide sense.

In the second case when pure relative delay is measured as in 2S1D, for example, then

$$y = \tau_{12} ,$$

$$= \frac{R_2}{C_2} - \frac{R_1}{C_1} ,$$

$$= \frac{(x_1^2 + x_3^2)^{1/2}}{x_6} - \frac{[x_1^2 + (x_3 - Z_2)^2]^{1/2}}{x_5} , \tag{2-90}$$

$$y' = \frac{x_1 x_2 + x_3 x_4}{x_6 R_2} - \frac{x_1 x_2 + (x_3 - Z_2) x_4}{x_5 R_1} . \tag{2-91}$$

Continuation up to (n-1)th order derivatives shows that the results are almost identical to the first case except x_5 appears in the expressions. It implies immediately that all the states are observable at least in a wide sense. When adding more measurements by addition of more sensors like 3S2D or 3S3D, the system becomes more observable due to increasing the possibility of uniqueness of the solution in terms of state x.

In the last case when the measurement equations include Doppler shift as in 2S1P, 2S1D1P or 3S2D1P shows very interesting results. For example when observing one Doppler shift in a two-sensor deployment (2S1P)

$$
\begin{aligned}
y &= f_{12} \, , \\
&= f_c \left(\frac{\dot{R}_2}{C_2} - \frac{\dot{R}_1}{C_1} \right) , \\
&= f_c \left(\frac{x_1 x_2 + x_3 x_4}{x_6 R_2} - \frac{x_1 x_2 + (x_3 - Z_2) x_4}{x_5 R_1} \right) , \\
&= f_c y_D' \, , \qquad (2\text{-}92)
\end{aligned}
$$

where y_D' is the time differentiation of the delay (2-91). Continuation gives

$$
\begin{aligned}
y' &= f_c y_D'' \, , \\
y'' &= f_c y_D''' \, , \\
&\vdots \\
y^{(5)} &= f_c y_D^{(6)} \, . \qquad (2\text{-}93)
\end{aligned}
$$

Doppler measurement is just scaling up of one step higher delay differentiation with scaling factor f_c. However, as discussed earlier the 2S1D system itself is already observable (at least in a wide sense). So, this system is also observable in the same context. The same argument can be applied for the 2S1D1P or 3S2D1P measurement cases, also. Thus the Doppler measurement system is observable deterministically as far as a delay measurement system is observable. Of course, a scaling factor influences the magnitude of the information obtained from the measurement. The effect of this will be discussed in Chapter Four where information structures of the various measurement schemes are analyzed.

CHAPTER 3: INFORMATION-THEORETIC OBSERVABILITY OF STOCHASTIC SYSTEMS

3-1. Introduction to information theory

Involvement of the noises in the stochastic system description makes it very difficult to extend the deterministic system observability condition to apply in the stochastic system case. A "yes" or "no" type answer to the observability question has little meaning in this case. Attempts on this problem must be interpreted in a probabilistic sense.

Contrary to the former results [34]-[39] where Fisher information is mainly used to study the stochastic observability, here Shannon information is utilized instead. Specifically, mutual information is computed and used as a criterion to determine the degree of observability of any states or whole system.

Information theory has two general orientations: one developed by Wiener and another by Shannon. Although both Wiener and Shannon shared a common probabilistic basis, there is some distinction between them. The significance of Wiener's work is that, if a signal is corrupted by some noises, then it is attempted to recover the signal from the corrupted one. It is for this purpose that Wiener orignated optimum filtering theory. However, Shannon's work goes to the next step. He showed that the signal can be transferred optimally provided it is properly formed. That is, the signal to be transferred can be processed before and after sending to counter the disturbance and to be recovered properly at the destination. For this purpose, Shannon developed the theories of information measure, channel capacity, coding processors, and so on.

To define the information measure, consider the simple information channel Figure 4 and assume that x_i is an input event and y_j is a corresponding output event, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

Now define a measure of the amount of information provided by the output (or measurement) y_j about the input x_i. It is not difficult to expect that the transmission of x_i through the noisy channel causes a change in the probablility of x_i from an a priori $p(x_i)$ to an a posteriori $p(x_i|y_j)$. In measuring this change, take the logarithmic ratio of the two probabilities. It turns out to be appropriate for the definition of information measure which is suggested first by Hartley [40]. I.e., the amount of information provided by y_j about x_i can be defined as [40], [41].

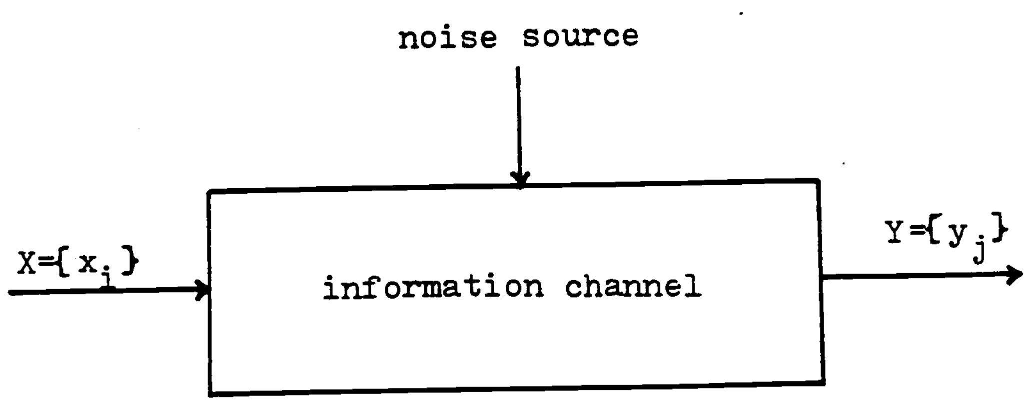


Figure 4, Input-output block diagram for information channel

$$I(x_i, y_j) = \log_2 \frac{p(x_i|y_j)}{p(x_i)} \text{ , bits ,}$$

$$= \log_{10} \frac{p(x_i|y_j)}{p(x_i)} \text{ , hartleys ,}$$

$$= \ln \frac{p(x_i|y_j)}{p(x_i)} \text{ , nats .} \quad (3\text{-}1)$$

(3-1) is defined by Shannon and used as a measure of mutual information between event x_i and y_j. If $p(x_i|\ y_j) = 1$

$$I(x_i, y_j) = I(x_i),$$

$$= \ln\ (1/p(x_i)) = -\ln p(x_i). \quad (3\text{-}2)$$

(3-2) is called self information. If (3-2) is true for all i, then the channel is noiseless. Averaged amount of information which is represented by H(x)

$$H(x) = \sum_{i=1}^{n} p(x_i) I(x_i) \text{ ,}$$

$$= - \sum_{i=1}^{n} p(x_i) \ln p(x_i), \quad (3\text{-}3)$$

has been, traditionally, called "information entropy," or just "entropy" of x. In statistical thermodynamics H is a measure of "disorder" or "uncertainty." Boltzmann showed [42] that in an isolated thermodynamic system H could never decrease, i.e.,the system tends to its maximum disorder. To decrease the entropy, one must add information to it either by transferring entropy out of the system boundary or by making observation (measurement). Here we are interested in the latter method.I.e., to decrease the uncertainty of the general stochastic system, measurement will be made and observe the decreased amount of uncertainty, and thus will use this quantity as a test criterion of the observability of the system.For an n random vector x with continuous probability density p(x) with natural logarithm base, H(x) becomes

$$H(x) = -\int_x p(x)\ln\frac{1}{p(x)}\,dx\ ,$$

$$= -\int_x p(x)\ln p(x)dx\ ,$$

$$= -E[\ln p(x)]\ , \tag{3-4}$$

where E is expectation operator.

Another quantity of information content which is commonly used is the Fisher information. For the same x and density p(x), Fisher information is defined as [43]-[47] and [66].

$$J(x) = -\int_x p(x)\frac{\partial^2 \ln p(x)}{\partial x\partial x^T}\,dx,$$

$$= \int_x p(x)\left(\frac{\partial \ln p(x)}{\partial x}\right)\left(\frac{\partial \ln p(x)}{\partial x}\right)^T dx,$$

$$= \int_x \frac{1}{p(x)}\left(\frac{\partial p(x)}{\partial x}\right)\left(\frac{\partial p(x)}{\partial x}\right)^T dx. \tag{3-5}$$

Algebraic identity

$$\frac{\ln p(a)}{a} = \frac{1}{p(a)} \cdot \frac{\partial p(a)}{\partial a} ,$$

was used in the last equality of (3-5). More compactly (3-5) becomes

$$J(x) = -E[\frac{\partial^2 \ln p(x)}{\partial x \cdot \partial x^T}] ,$$

$$= E[(\frac{\partial \ln p(x)}{\partial x})(\frac{\partial \ln p(x)}{\partial x})^T] \qquad (3\text{-}6)$$

From the two definitions (3-4) and (3-5) above, it is clear that the Fisher information J is a nxn matrix quantity and that the Shannon information H is a scalar valued quanitity. The general relation between these two information concepts will be discussed briefly later. However, immediate comparison of (3-4), (3-5) shows that a simple relation can be derived if a specific density p(x) is given for any random variable x. For example, a scalar random variable χ with Gaussian density having zero mean and variance σ^2 has a Fisher information

$$J(\chi) = -E[\frac{\partial^2 \ln p(\chi)}{\partial \chi^2}] = \frac{1}{\sigma^2} \quad . \qquad (3\text{-}7)$$

Meanwhile its entropy is

$$H(\chi) = -E[\ln p(\chi)] = 1/2\ln(2\pi\sigma^2). \qquad (3\text{-}8)$$

So, from (3-7), (3-8) one can get the relation

$$\frac{dH(x)}{d(\sigma^2)} = 1/2J(x) \quad . \tag{3-9}$$

Generalization of this relation can be found in [43] and [44].

Appendix B shows that the maximum entropy density funcitcn varies depending on the constraints which are added to the density $p(x)$. The Gaussian density has maximum entropy under the given mean and variance condition when x ranges from $-\infty$ to $+\infty$.

It is known that [48, and from private communication with R.W. Hamming, Naval Postgraduate School, March 1985] entropy of commonly used random variables $H(x)$ and its variance σ_x^2 have one-to-one relation

$$H(x) = 1/2 \ln(A\sigma_x^2), \tag{3-10}$$

if the density and expectation of x exist. So , for example, the inverse-Gaussian or Cauchy density does not have the relation (3-10) due to nonexistence of mean and variance expressions. Constant A is determined once density is known. $A = 2\pi e$ for Gaussian case, for example, from (3-8).

Table 1 shows this relationship for some commonly used densities [48].

Table 1. Entropy-variance relationship

Distribution	Pdf $p(\chi)$	Const. A
Gaussian	$\frac{1}{\sqrt{2\pi a}} \exp(-\frac{\chi^2}{2a^2})$	$2\pi e$ (~17.079456)
Uniform	$1/a;\ -a/2 \leq \chi \leq a/2$	12.
Triangular	$a + a^2\chi;\ -1/a \leq a \leq 0$	$6e$ (~16.30968)
Exponential	$ae^{-a\chi},\ \chi > 0$	e^2 (~7.389046)
Double Exponential	$1/2\ ae^{-a\lvert\chi\rvert}$	$2e^2$(~14.778092)
Rayleigh	$\frac{\chi}{a^2} \exp(-\frac{\chi^2}{2a^2})$	$\frac{e^{\mu+2}}{4-\pi}$ (~15.331182)
Poission	$\frac{a^{n+1}\chi^n e^{-a\chi}}{n!}$; $\chi > 0$; $n > 1$	$\frac{(n!)^2}{n+1} \exp[2+2n(\mu+1-\sum_{i=1}^{n} 1/i)]$ (15.98307 for n = 10)

$$\text{Euler Const.} = \lim_{n\to\infty} (1 + 1/2 + 1/3 + \ldots + 1/n - \ln(n))$$

$$=0.577215664$$

3-2. The concept of mutual information.

Calculation of the amount of the information about one random function contained in another random function, so called mutual information, has many important applications. In communication this concept is used to detect or decode a transmitted signal from a noise contaminated received signal [41], [49], [50]. The extended application of the mutual information to a more general system to identify unknown parameters is tried by Weidemann and Stear [51]. Later with the help of measure theory, its utilization is widened into the area of filtering of general stochastic systems [45], [46], [52]-[54]. Here an attempt is made further to apply the same concept in the observability problem. The main feature of this approach lies in the transition of the definition of the term "information" from Fisher to Shannon, i.e., the meaning of information here is understood in the sense of Shannon.

Define two random vectors x and y as

$$x = (x_1, x_2, \ldots, x_n),$$
$$y = (y_1, y_2, \ldots, y_m),$$

and assume a joint density p(x,y), and marginal densities p(x) and p(y) are defined as usual. Then the entropy of x, H(x) is defined as by (3-4). Entropy of y, H(y) is defined similarly

$$H(y) = -E[\ln p(y)].$$

In the same context conditional entropy $H(x|y)$ can be defined as in [41], [51]-[54], i.e., for a given conditional density $p(x|y)$ and chosen specific value $y = \mathcal{y}$ then

$$H(x|\mathcal{y}) = -\int_X p(x|\mathcal{y}) \ln p(x|\mathcal{y}) dx. \tag{3-11}$$

From the average over all possible y

$$H(x|y) = -\int_Y p(y)H(x|y)dy,$$

$$= -\int_{X,Y} p(y)p(x|y)\ln p(x|y)dxdy,$$

$$= -\int_{X,Y} p(x,y)\ln p(x|y)dxdy,$$

$$= -E[\ln p(x|y)]. \tag{3-12}$$

Next, define joint entropy H(x,y) in a similar way as

$$H(x,y) = -\int_{X,Y} p(x,y)\ln p(x,y)dxdy,$$

$$= -E[\ln p(x,y)]. \tag{3-13}$$

With the above definitions, mutual information between x and y is derived.

Upon the definition of (3-1), the average mutual information of x for specific $y = y$ is termed as <u>conditional mutual information [41] $I(x,y)$</u> which is expressed as

$$I(x,y) = \int_X p(x|y)I(x,y)dx,$$

$$= \int_X p(x|y)\ln\frac{p(x|y)}{p(x)}dx. \tag{3-14}$$

$I(x,y)$ is the measure of information gain which is provided by the measurement $y = y$. So, averaging of (3-14) for all possible values of y yields the formal definition of the <u>mutal information</u> I(x,y) [41], [45], [51]-[54] as

$$I(x,y) = \int_{x,y} p(x,y) \ln \frac{p(x|y)}{p(x)} dxdy,$$

$$= \int_{x,y} p(x,y) \ln \frac{p(y|x)}{p(y)} dxdy. \qquad (3\text{-}15)$$

Using the entropy definitions (3-4), (3-12), (3-13) I(x,y) becomes

$$I(x,y) = H(x) - H(x|y),$$

$$= H(y) - H(y|x),$$

$$= H(x) + H(y) - H(x,y). \qquad (3.16)$$

(3-16) can be diagrammed as in Figure 5.

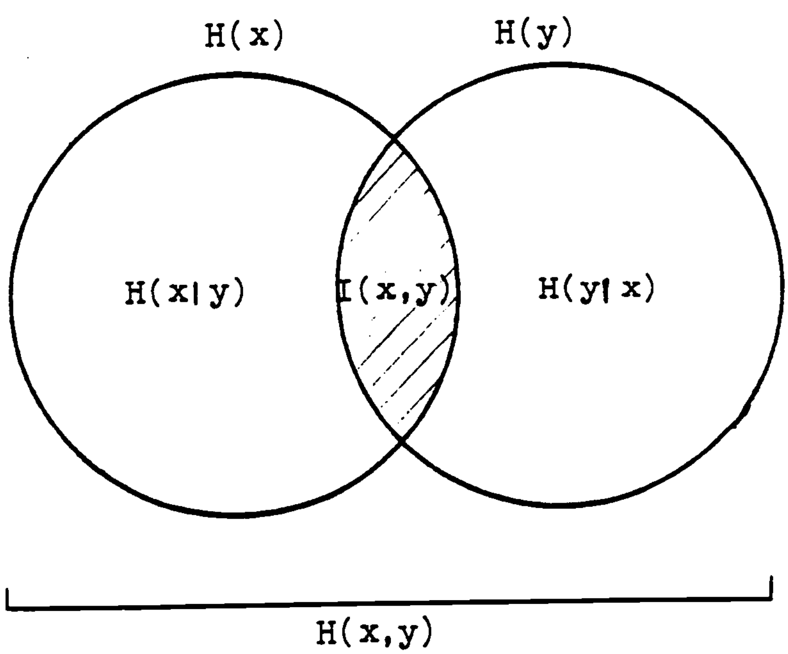


Figure 5, Entropy and mutual information

I.e., mutual information is the common portion of the information $H(x)$ and $H(y)$. So, it is clear from (3-15) that if x and y are independent, i.e,,

$$p(x|y) = p(x),$$

then, $I(x,y)$ is always zero due to $\ln(1) = 0$ and no common portion in Figure 5.

1. <u>Properties of I(x,y)</u>

Mutual information has the following important properties;

1) $I(x,y) = I(y,x) \geq 0$

This inequality is called the "Shannon inequality." Mutual information is always greater than zero except the case where x, y are stochastically independent.

2) $I(x,y) \geq I(x, L(y))$

Some information is lost by the transformation L, where L(y) is any mapping which depends on the domain of y. Equality holds if and only if the mapping is one-to-one and onto. Loss of information depends on the relation

$$H(y) = H(x) + E[\ln|J|],$$

where $y = f(x)$, J = Jacobian of $f(x)$

3) $I(x,y) \geq I(z, y)$, (3-17)

where $z = f(x,N)$, N is a random function or variable. Information loss is incurred, also, due to the random term in the transformation.

4) The information about x increases monotonically as more observation is taken, i.e.,

$$I(x_1, \ldots, x_k; y_1, \ldots, y_M) \leq I(x_1, \ldots, x_k; y_1, \ldots, y_M, y_{M+1} \ldots), \tag{3-18}$$

For our own purpose here, the first equality of (3-16) and the property 4) above play the most important role. (3-16) is used to compute mutual information between x and y by considering H(x) as an uncertainty of the system state x before an observation is made and H(x|y) as the uncertainty of x after an obsservation is made. Thus I(x,y) is interpreted here as the uncertainty decrease or, equivalently, information increase due to the observation. Since this uncertainty difference is entirely caused by the observation y, the mutual information I(x,y) can be used as the measure of the observability of the system. The increased amount of information due to the observation, then can be evaluated using the inequality (3-18). I.e., the difference

$$I(x_1, \ldots, x_k; y_1, \ldots, y_M, y_{M+1}) - I(x_1, \ldots, x_k; y_1, \ldots, y_M)$$

is the information change or information rate which is caused by the (M+1)-th observation data. In communication theory the maximum mutual information over the p(x) is defined as channel capacity C,

$$C = \max_{p(x)}(I(x,y)). \tag{3-19}$$

<u>Example 3-1</u>

Consider a simple scalar system where observation y is the sum of the random variable x and observation noise n

$$y = x + n. \tag{3-20}$$

Let x be a zero mean Guassian random variable with density

$$p(x) = \frac{1}{\sqrt{2\pi S}} \exp(-\frac{x^2}{2S}). \tag{3-21}$$

S is the power in the signal . Suppose another random variable n is independent of y and is Gaussian with zero mean, variance σ_n^2. Then

$$\begin{aligned} p(y|x) &= p_n(y-x) \\ &= \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\{-\frac{(y-x)^2}{2\sigma_n^2}\} . \end{aligned} \tag{3-22}$$

So, from Table 1, the conditional entropy is

$$H(y|x) = 1/2\ln(2\pi e\sigma_n^2). \tag{3-23}$$

Since the output is a sum of the two Gaussian signals it is also Guassian with variance $S + \sigma_n^2$, i.e.,

$$p(y) = \frac{1}{\sqrt{2\pi(S+\sigma_n^2)}} \exp\{-\frac{y^2}{2(S+\sigma_n^2)}\}. \tag{3-24}$$

So,

$$H(y) = 1/2\ln[2\pi e(S+\sigma_n^2)] \quad . \tag{3-25}$$

Thus, from (3-22), (3-25) and the definition (3-16)

$$I(x,y) = H(y) - H(y|x) \; ,$$

$$= 1/2\ln(1 + \frac{S}{\sigma_n^2}) = 1/2\ln(1 + \frac{S}{N}), \tag{3-26}$$

where N is the noise power. Note in (3-26) that as noise power becomes small, mutual information increases due to $H(y|x)$ decreasing. So, the output y approximates the input x more exactly. Oppositely, if $N \to \infty$, i.e., the input is totally "masked" by the noise, then $I(x,y)$ approaches zero. Then x and y look like independent signals. No information about x is tranferred to y. All of the information is lost during the transmission. It is clear that $I(x,y)$ increases with increasing signal to noise ratio (SNR). Since, the correlation coefficient r, in this case is

$$r^2 = \frac{\sigma_x^2}{\sigma_Y^2} = \frac{S}{S + N}$$

$I(x,y)$ can be obtained in terms of r from (3-26),

$$I(x,y) = 1/2\ln(1 + \frac{S}{N}) \quad ,$$

$$= -1/2\ln(1 - \frac{S}{S+N}) \quad ,$$

$$= -1/2\ln(1 - r^2). \qquad (3\text{-}27)$$

$I(x,y)$ is a function of only r and ranges from zero to infinite value as $|r|$ ranges from zero to one.

3-3. Mutual information of stochastic systems.

Figure 6 shows the schematic configuration of the typical stochastic system. Comparison of Figure 4 and 6 shows that the measurement mechanism h(.) can be identified as an information channel where transferring of information occurs.

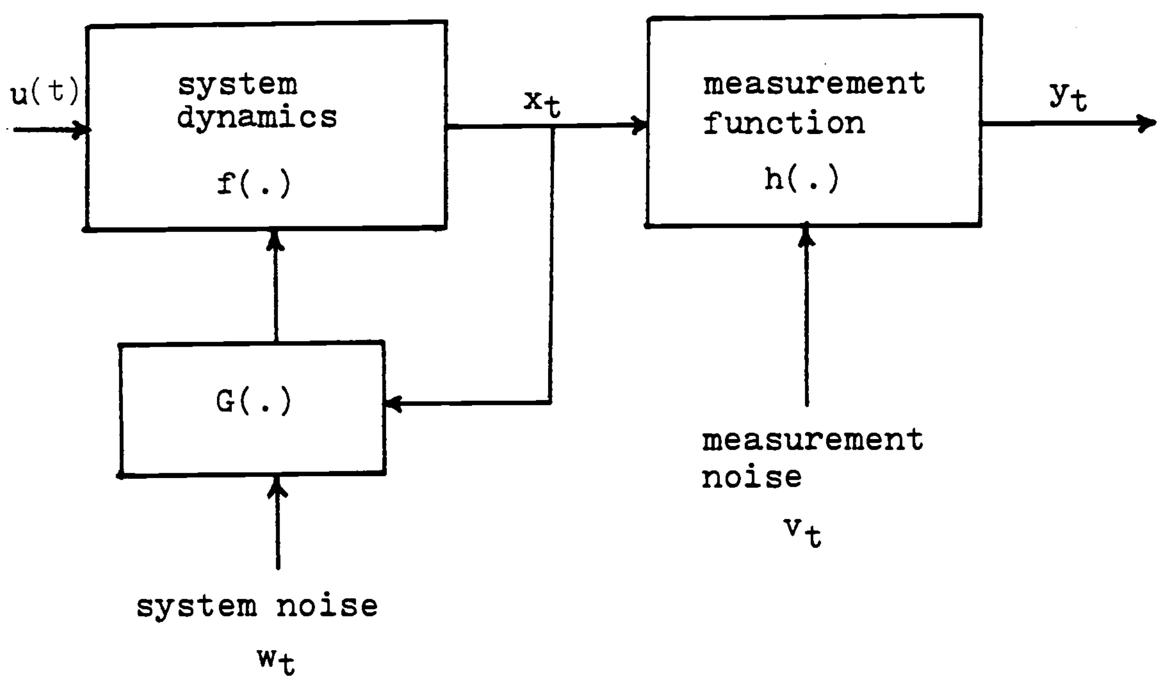


Figure 6, Typical stochastic systems

However, generalization of Shannon's result (3-15) or (3-16) to the continuous random process needs more assumptions on the measure theoretic point of view. This is discussed next.

First, consider that the observation of the process x_t which is expressed in terms of the Ito stochastic differential equation (with the suppression of deterministic control u(t))

$$dx_t = f(x_t, t)dt + G(x_t, t)dw_t, \; x_{t0} = x_o \tag{3-28}$$

is made through another stochastic equation

$$dy_t = h(x_t, t)dt + dv_t, \tag{3-29}$$

where $x_t \varepsilon R^n$, $y_t \varepsilon R^m$; f(.) and h(.) are n, m dimensional vector valued functions, respectively. w_t and v_t are independent Wiener processes with covariances Q(t), R(t) independent of x_{t0}. G is an appropriate dimensional matrix. Assume (3-28), (3-29) satisfy the existence and uniqueness conditions of the solution in the mean-square sense [34], [36]. Let (Ω, F, μ) be a measure space. Let Y = C[0,T] and F^Y be the family of Borel sets of Y and F_t^Y be non-decreasing sub-σ-algebras of F^Y generated by $\{y_s, \; 0 \le s \le t\}$. The measure induced by y_t on the space $(Y, \; F_t^Y)$ is denoted by μ_Y and the Wiener measure induced by v_t on $(Y, \; F_t^V)$ is denoted by μ_V. Let X be the vector space and F^X be the family of Borel sets of X. F_t^X is also a nondecreasing sub-σ-algebras of F^X. Then the measure induced joint measure μ_{XY} of the joint process $(x_t, \; y_t)$ is defined on the space $(X \times Y, \; F_t^X \times F_t^Y)$. Further assume that

$$\int_0^t h(x_s,s)^T h(x_s,s)ds < \infty \,, \quad \text{a.s.} \tag{3-30}$$

Then Gel'fand and Yaglon [55], Liptser and Shiryayev [56], Duncan [45] proved that the absolute continuity

$$\mu_Y << \mu_V \,, \tag{3-31}$$

$$\mu_{XY} << \mu_X \times \mu_V \tag{3-32}$$

holds. Further it is known that [46], [56] equivalence relation of the measures

$$\mu_Y \sim \mu_V$$

$$\mu_{XY} \sim \mu_X \times \mu_Y \sim \mu_X \times \mu_V$$

holds, also. If once absolute continuity condition holds, then by the Randon-Nikodym theorem [28], [31], [57] there exists a finite real valued unique F-measurable funciton ϕ on Ω such that for every $A \varepsilon F$, e.g., in (3-31)

$$\mu_Y(A) = \int_A \phi_1(\omega) d\mu_V(\omega) \,, \tag{3-33}$$

or in a differential form

$$\phi_1(\omega) = \frac{d\mu_Y}{d\mu_V}(\omega) \quad . \tag{3-34}$$

With the same reason for the (3-32)

$$\phi_2(\omega) = \frac{d\,\mu_{X,Y}}{d\mu_X \times d\mu_V}(\omega) \quad . \tag{3-35}$$

The function, known as a likelihood ratio, plays a key role in the derivation of mutual information. From the Cameron-Martin translation theorem [45], [46], [58] for the system (3-28) and (3-29), likelihood ratio becomes

$$\frac{d\mu_Y}{d\mu_V}(y) = \exp\{\int_o^t \bar{h}(x_s,s)^T R^{-1} dy_s - 1/2 \int_o^t \bar{h}(x_s,s)^T R^{-1} (\bar{h}_s,s)ds\}, \tag{3-36}$$

$$\frac{d\mu_{X,Y}}{d\mu_X \times d\mu_V}(x,y) = \exp\{\int_0^t h(x_s,s)R^{-1} dy_s - 1/2\ h(x_s,s)^T R^{-1} \quad h(x_s,s)ds\}, \tag{3-37}$$

where $\bar{h}(x_s,s) = E[h(x_s,s)|F_s^Y]$. If all the measures considered are probability quantitites P_X, P_Y, P_V and P_{xy}, respectively. Then the Radon-Nikodym derivatives ϕ_1 and ϕ_2 become density ratios

$$\phi_1 = \frac{dP_Y}{dP_V}, \quad \phi_2 = \frac{dP_{xy}}{dP_X dP_V}.$$

So, by letting ϕ be

$$\begin{aligned}\phi &= \phi_2 \cdot \phi_1^{-1}, \\ &= \frac{dP_{xy}}{dP_X dP_V} \cdot \frac{dP_V}{dP_Y}, \\ &= \frac{dP_{xy}}{dP_X dP_Y}.\end{aligned} \tag{3-38}$$

Then, from the definition of mutual information (in Shannon sense)

$$I(x_{t,}y_t) = \int \phi(x_t,y_t) \ln \phi (x_t,y_t) dP_x dP_y \quad . \tag{3-39}$$

Since, $P_{xy}(x_t,y_t) = P_{x|y}(x_t|y_t)P_y(y_t)$

$$\phi (x_t,y_t) = \frac{dP_{x|y}(x_t|y_t)dP_y(y_t)}{dP_x(x_t)dP_y(y_t)} \quad ,$$

$$= \frac{dP_{x|y}(x_t|y_t)}{dP_x(x_t)} \quad . \tag{3-40}$$

So, inserting (3-40) into (3-39) yeilds

$$I(x_t,y_t) = \frac{dP_{x|y}(x_t|y_t)}{dP_x(x_t)} \ln \frac{dP_{x|y}(x_t|y_t)}{dP_x(x_t)} dP_x(x_t)dP_y(y_t) \quad . \tag{3-41}$$

If probability density is used instead of distribution with the notations

$$p_x(x_t) = \frac{dP_x(x_t \leq x)}{d\,x}, \; p_y(y_t) = \frac{dP_y(y_t \leq y)}{d\,y}, \; p_{x|y}(x_t|y_t) = \frac{dP_{x|y}(x_t \leq x|y)}{d\,x} \quad ,$$

(3-41) becomes

$$I(x_t,y_t) = \int p_{x|y}(x_t|y_t) \ln \frac{p_{x|y}(x_t|y_t)}{p_x(x_t)} p_y(y_t) dx_t dy_t \quad ,$$

$$= \int p_{xy}(x_{t,}y_t) \ln \frac{p_{x|y}(x_t|y_t)}{p_x(x_t)} dx_t dy_t \quad ,$$

$$= E \ln \frac{p_{x|y}(x_t|y_t)}{p_x(x_t)} ,$$

$$= H(x_t) - H(x_t|y_t) . \tag{3-42}$$

Therefore, to compute mutual information for the system (3-28), (3-29) one is, again, required to know either two densities - unconnditional and conditional - or two entropies. Next is a brief discussion on the solution of these density equations and approximation methods of these densities using appropriate moments.

1. $p(x_t)$ and two-moment approximation

Consider the sytem equation (3-28) again

$$dx_t = f(x_t,t)dt + g(x_t,t)dw_t, \quad x_{to} = x_o . \tag{3-43}$$

Due to the unknown initial state x_o and the additive noise w_t, the process $\{x_t\}$ can only be described by the statistical treatment. As is known [36], [57] the probability density evolution of $p(x_t)$ obeys the Kolmogorov forward equation

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^{n} \frac{\partial (pf_i)}{\partial x_i} + 1/2 \sum_{i,j}^{n} \frac{\partial^2 (pGQG^T)}{\partial x_i \partial x_j} , \tag{3-44}$$

where all the arguments in the expression are omitted for brevity. But unfortunately the above partial differential equation can be solved only for a few sprecial simple case. So, in many practical problems one relies on an alternative approximation approach such as state estimation; e.g.,one obtains proper approximated moments of the density instead of the density itself. Particularly the first two moments are important in entropy computational purpose even though they do not completely characterize density $p(x_t)$. It is known that [36] the first two moments mean $\hat{x}_t$ and covariance P_t propagate according to

$$\dot{\hat{x}}_t = f(\hat{x}_t, P_t, t) \tag{3-45}$$

$$\begin{aligned} \dot{P}_t = {} & E[f(x_t,t)x_t^T] - E[f(x_t,t)]\hat{x}_t^T + E[x_t f^T(x_t,t)] \\ & - \hat{x}_t E[f^T(x_t,t)] + E[G(x_t,t)Q(t)G^T(x_t,t)], \end{aligned} \tag{3-46}$$

where $\hat{x}_t = E[x_t|x_s, \quad s\leq t]$. By neglecting third and higher-order moments in the evaluation of (3-45) and (3-46), one obtains the following approximated version for $\hat{x}_t$ and P_t.

$$\dot{\hat{x}}_t = f(\hat{x}_t,t) + \frac{P_t}{2} f_{xx}(\hat{x}_t, t), \tag{3-47}$$

$$\begin{aligned} \dot{P}_t = {} & f_x(\hat{x}_t,t)P_t + P_t f_x^T(\hat{x}_t, t) + G(\hat{x}_t,t)Q(t)G^T(\hat{x}_t,t) \\ & + P_t G_x(\hat{x}_t,t)Q(t)G_x^T(\hat{x}_t,t) + P_t G(\hat{x}_t,t)Q(t)G_{xx}^T(\hat{x}_t,t), \end{aligned} \tag{3-48}$$

where $f_x(.)$ and $G_x(.)$ are first partial derivatives and $f_{xx}(.)$, $G_{xx}(.)$ are second partial derivatives at x_t. Further if the second partials of (3-47), (3-48) are negligible compared to the first partials and

$G(.)$ is not a function of x_t, then

$$\dot{\hat{x}}_t = f(\hat{x}_t, t), \tag{3-49}$$

$$\dot{P}_t = f_x(\hat{x}_t, t)P_t + P_t f_x^T(\hat{x}_t, t) + G(t)Q(t)G^T(t) \tag{3-50}$$

which is a commonly used approximation. Of course, there are many other algorithms which can be practically useful .

2. $p(x_t|Y_t)$ and extended linear filter

Conditional density $p(x_t|Y_t)$ of the system (3-28), (3-29) satisfies the nonlinear stochastic partial differential equation, commonly known as the Kushner equation [34], [36]

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial (pf_i)}{\partial x_i} + 1/2 \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i \partial x_i}(pGQG^T) + \{h(x_t, t) - Eh(x_t, t)\}^T R^{-1}(t)$$

$$\{dy_t - Eh(x_t, t)dt\}p. \tag{3-51}$$

Due to the additional measurement-related third term in (3-51) it may be more complicated to solve than (3-44). To obtain the conditional moments of the pdf $p(x_t|Y_t)$ of (3-51) let

$$\hat{\psi}(x_t) = E[\psi(x_t)|F_t^Y],$$

then any conditional moment satisfies the stochastic differential equation

$$d\hat{\psi}(x_t) = \{E[\psi_x^T f] + 1/2\ \mathrm{tr}[E(GQG^T\psi_{xx})]\}dt + \{E[\psi h] - \hat{\psi}\hat{h}\}^T R^{-1}(dy_t - \hat{h}dt), \tag{3-52}$$

where $\hat{h} = E[h(x_t,t)|F_t^Y]$ and ψ_x, ψ_{xx} are the first and second partial derivatives of ψ relative to x_t, respectively. By letting $\psi(x_t) = x_t$ and $\psi(x_t) = x_t x_t^T$ obtains the mean and covariance as

$$d\hat{x}_t = f(\hat{x}_t,t)dt+\{E[x_t h^T(x_t,t)]$$

$$-\hat{x}_t E[h^T(x_t,t)]\}R^{-1}(t)\{dy_t-E[h(x_t,t)]\}, \tag{3-53}$$

$$dP_t = \{E[(x_t-\hat{x}_t)f^T]+E\{f(x_t-\hat{x}_t)^T\}+E[GQG^T]-E[(x_t-\hat{x}_t)h^T]R^{-1}(t)$$

$$E[h(x_t-\hat{x}_t)^T]\}dt+E[(x_t-\hat{x}_t)(x_t-\hat{x}_t)^T(h-Eh)^T]R^{-1}(t)\{dy_t-Ehdt\}, \tag{3-54}$$

Since, P_t is a function of the higher-order moments it can not be a finite-dimensional filter in general. So, various approximations and assumptions are made to ensure that (3-53), (3-54) to be finite dimensional and practically-implementable filter algorithms. If, again, G(.) is a function of only t, and the first-order expansion cf f(.) and h(.) is made, then (3-53), (3-54) reduce to the well known extended Kalman filter

$$d\hat{x}_t = f(\hat{x}_t,t)dt + P_t h_x^T R^{-1}(t)[dy_t-h(\hat{x}_t,t)dt], \tag{3-55}$$

$$\dot{P}_t = f_x(\hat{x}_t,t)P_t+P_t f_x^T(\hat{x}_t,t)+G(t)Q(t)G^T(t)-P_t h_x^T R^{-1}(t)h_x P_t, \tag{3-56}$$

where $f_x = \left.\dfrac{\partial f}{\partial x_t}\right|_{x_t = \hat{x}_t}$,

$h_x = \left.\dfrac{\partial h}{\partial x_t}\right|_{x_t = \hat{x}_t}$.

The Kalman-Bucy filter is obtained, of course, if the system and measurement equations are linear. Depending on the order of the expansion of f(.) and h(.), second or even higher-order filters can be derived.

Notice here that the utilization of any approximated moment expressions of the density instead of the density itself incurs the conceptional change of the mutual information from $I(x_t, y_t)$ to $I(\hat{x}_t, y_t)$, where $\hat{x}_t = E[x_t | F_t^Y]$. In the next section, the second-order moment approximation of the density functions $p(x_t)$ and $p(x_t | y_t)$ will be discussed in the computation of the mutual information $I(\hat{x}_t, y_t)$. Before this, the relationship between the Shannon and Fisher information will be summarized for the stochastic system instead of the random variable case. The following are the vector version of the results of Liptser and Shiryayev [56].

3. Relationship between Shannon and Fisher Information.

Consider the general nonlinear stochastic system as in (3-28), (3-29). Nonlinear functional dependence of f(.) and G(.) in terms of x_t makes the derivation of any relationship between the two information concepts very difficult. This difficulty can be avoided if a specific form of nonlinear system is assumed, for example, linear dynamics-nonlinear observation system. In this case, system is given as

$$dx_t = f(t)x_t dt + g(t)dw_t, \tag{3-57}$$

$$dy_t = h(x_t, y_t, t)dt + dv_t. \tag{3-58}$$

Note that $h(.)$ can, also be a function of the observation y_t itself under the bounded strong solution condition for t and the nonanticipativeness for y_t. Assume further that $h(.)$ satisfies

$$\int_{t_o}^{T} [h^T(x_t,y_t,t)h(x_t,y_t,t)]dt<\infty , \tag{3-59}$$

for each t, $t_o \le t \le T$, and two densities $p(x_t)$ and $p(x_t|y_t)$ are twice continuously differentiable with respect to x_t.

Then Fisher and Shannon information has the following relation ;

$$I(x_t,y_t)=I_o(x_t,y_t)-1/2\int_{t_o}^{t} tr\{g(s)g^T(s)[EJ(x_s,y_s)-J(x_s)]\}ds, \tag{3-60}$$

where $I_o(x_t,y_t)$ is Shannon information quantity due to the observation equation (3-58) only, i.e., the case where statistical uncertainty of the process $\{x_t\}$ is not considered, and $J(x_t,y_t)$, $J(x_t)$ are Fisher information quantities corresponding to the densities $p(x_t|y_t)$ and $p(x_t)$, respectively. $I_o(x_t,y_t)$ can be expressed according to [45], [46], [56].

$$I_o(x_t,y_t)=1/2tr.\int_{t_o}^{t} E\{[h(x_s,y_s,s)-\bar{h}(x_s,y_s,s]^T[h(x_s,y_s,s)-\bar{h}(x_s,y_s,s)]\}ds, \tag{3-61}$$

where

$$\bar{h}(x_t,y_t,t) = E[h(x_t,y_t,t)|F_t^Y].$$

The proof of the relation (3-60) can be found in the cited reference.

3.4 Observability using mutual information

As mentioned before the mutual information $I(x_t, y_t)$ is the information contents (in Shannon's sense) about the state x_t which is contained in the observation y_t, i.e., the common information of the two processes x_t and y_t. So, once it is computed then it represents the "tightness" of the connection of the state x_t to the observation y_t. Hence, it might be sued as a criterion to determine the degree of the observability of the given stochastic system. The term "observability" here is, of course, used in a different meaning from the deterministic case and even different with the traditionally used stochastic case where the Fisher information is commonly used.

As the Fisher information matrix and the observability matrix is practically used together in a traditional observability determination, Shannon's mutual information and the term "observability" will be used together henceforth.

But due to the difficulty in solving the exact density equations, Kolmogorov forward equation and Kushner equations, approximated moment expressions are utilized, alternatively.

Before this, former results on stochastic system observability are summarized next.

1. Former results on stochastic system observability.

Consider, again, a general form of stochastic system (3-28), (3-29)

$$dx_t = f(x_t,t)dt + G(x_t,t)dw_t, \tag{3-62}$$

$$dy_t = h(x_t,t)dt + dv_t. \tag{3-63}$$

The traditional approach in the determination of the observability of the system (in a Fisher sense) is as follows: using the likelihood function Λ with $\Lambda = p(x_t|y_t)$ for the noiseless system $(Q(t) = 0)$ in (3-28), (3-29), it's logarithm quantity $\ln(\Lambda)$ is maximized according to the definition of the Fisher information

$$J = -E[\frac{\partial^2 \ln(\Lambda)}{\partial x_t \partial x_t^T}] . \tag{3-64}$$

Then, the Fisher information matrix $J(t,t_o)$ for the first-order approximation of the system about the estimation $\hat{x}_t$ is obtained as [34]-[36], [38], [46], [47]

$$J(t,t_o) = \Phi^T(t_o,t)P_o^{-1}\Phi(t_o,t) + \int_{t_o}^{t}\Phi^T(s,t)H^TR^{-1}H\Phi(s,t)ds,$$

$$= J_i(t,t_o) + J_o(t,t_o), \tag{3-65}$$

where $\Phi(.)$ is transition matrix for the linearized portion of $f(.)$ at $\hat{x}_t$, i.e.,

$$\frac{\partial \Phi}{\partial t}(t,s) = F(t)\Phi(t,s), \quad \Phi(t,t) = I, \tag{3-66}$$

$$F(t) = \left.\frac{\partial f}{\partial x_t}\right|_{x_t = \hat{x}_t},$$

$$H(t) = \left.\frac{\partial h}{\partial x_t}\right|_{x_t = \hat{x}_t},$$

and J_i is due to the initial information P_o^{-1} and J_o is due to the observed information, respectively. Or after some algebraic manipulation recursive version of (3-65) is obtained as

$$\frac{dJ(.)}{dt} = -F^T(t)J(.)-J(.)F(t)+H^T(t)R^{-1}(t)H(t). \tag{3-67}$$

Traditionally $J_o(.)$ is called an observability matrix (some authors [34], [36] call it an information matrix.). Then positive definiteness or nonsingularity of $J_o(.)$ is used as a criterion of the determination of the observability for the system. Or, for some positive constants α, β , s , and unit matrix I, the relation

$$0 < \alpha I \leq J_o(t,t-s) \leq \beta I, \tag{3-68}$$

is checked for all $t \geq t_o+s$ [36].

However, the Fisher information matrix J is related to the estimation error covariance matrix P_t by [47]

$$P_t \geq (I + \frac{\partial b}{\partial x_t})J^{-1}(I + \frac{\partial b}{\partial x_t})^T. \tag{3-69}$$

with b(t) being a bias of x_t with respect to x_t. If x_t is an unbiased estimator, then

$$P_t \geq J^{-1}, \tag{3-70}$$

Further, if $\hat{x}_t$ is optimal, then the equality in (3-70) holds. I.e., the covariance of the individual state estimation error is lower bounded by the diagonal elements of J^{-1} which is, so called, Cramer-Rao lower bound.

As well as the positive definiteness of the observability matrix, eigenvalues of this matrix are, sometimes, utilized to test the system observability [37]. Appearance of any zero eigenvalue(s) means singular J_o and causes system unobservability. High stiffness between the eigenvalues means weakly observable. Condition number q of J_o

$$q = \frac{e}{s},$$

where e and s are maximum and minimum eigenvalues, respectively, is used as an indicator of the system observability.

Somewhat different approach is studied by Sunahara [59]. A stochastic system (3-62), (3-63) is said to be observable if there exists an estimator such that the associated error converges to a sufficiently small value on the time interval $[t_o, t_1]$ in some stochastic sense, i.e., for the preassigned error constants δ and ε, $0 < \varepsilon < 1$ if

$$P(||x_{t_1} - \hat{x}_{t_1}||^2 \geq \delta) \leq \varepsilon, \qquad (3\text{-}71$$

is satisfied, then the system is said to be stochastically observable. Here $\hat{x}_t$ is obtained by the pre-assigned filter form

$$d\hat{x}_t = f(\hat{x}_t,t)dt+P_tH^T(t)[dy_t-h(\hat{x}_t,t)dt], \tag{3-72}$$

for the appropriate dimensional matrices P_t and $H(t)$.

Even though the Fisher-information approach is most widely used in the observability determination of the given stochastic system, several disadvantages can be indicated when compared with the Shannon's mutual-information approach.

1) Even though the theoretical definition of the Fisher information (3-64) can accommodate system noise w_t, the practically used form (3-65) does not accommodate w_t as far as the likelihood function which is chosen as the conditional density $p(x_t|y_t)$. Neglect of the system noise may cause incorrect results when w_t is significant compared to the other noises [39]. A convenient form to handle both system and observation noises is not yet available. However, the mutual information conveniently considers both noises simultaneously since it always requires both densities $p(x_t|y_t)$ and $p(x_t)$ together from its definition.
2) If the system is unobservable or marginally observable, then singularity or almost singularity of the observability matrix makes it very difficult to compute this matrix, practically. But this problem does not occur in the mutual information computation as can be seen in the next subsection.

3) Extending linear results to the general nonlinear case requires many approximations. In the nonlinear case ,a general form of transition matrix does not exist. $I(x_t,y_t)$ requires many approximations to be practically implementable, but here one can use many well-developed nonlinear filters which are already publicaly available.

4) Even with the above problems in the Fisher information approach, simplicity in the calculation and recursive nature make it popular in the linear or linearized, negligible system noise applications.

2. Observability computation using mutual information.

From the discussion of the previous section, computation of observability in terms of mutual information may be found conveniently by an approximated filter algorithm in many cases. From (3-42), $I(x_t,y_t)$ computation requires two entropies - marginal entropy $H(x_t)$ and conditional entropy $H(x_t|y_t)$. Both entropies can be computed from the relations

$$H(x_t) = n/2 \ln A + 1/2 \ln(\det \Gamma_t^T), \tag{3-73}$$

$$H(x_t|y_t) = n/2 \ln A + 1/2 \ln(\det P_t^T), \quad P_{to}^T = P_{to}^T, \tag{3-74}$$

where Γ_t^T is the covariance for the marginal density $p(x_t)$ and P_t^T is the covariance for the conditional density $p(x_t|y_t)$. Note superscript T is not a transpose here. No approximation is assumed in both Γ_t^T and P_t^T. Therefore, from insertion of (3-73), (3-74) into (3-42)

$$I(x_t,y_t) = n/2 \ \ln A + 1/2 \ \ln(\det \Gamma_t^T)$$
$$- n/2 \ \ln A - 1/2 \ \ln(\det P_t^T),$$
$$= 1/2 \ \ln \left(\frac{\det \Gamma_t^T}{\det P_t^T}\right), \tag{3-75}$$

where $I(x_{to},y_{to}) = 0$ due to $\Gamma_{to}^T = P_{to}^T$, i.e., initial information is normalized always to zero.

Since exact covariances Γ_t^T and P_t^T are, in general, functions of the higher-order moments, computation of these matrices are also difficult. If any of the second-order approximation algorithms is used with the resultant covariances Γ_t and P_t, then

$$H(\hat{x}_t) = n/2 \ \ln A + 1/2 \ \ln(\det \Gamma_t), \tag{3-76}$$
$$H(\hat{x}_t|y_t) = n/2 \ \ln A + 1/2 \ \ln(\det P_t), \tag{3-77}$$

where $\hat{x}_t$ is the estimation of x_t obtained from the chosen second order approximation. In this case, mutual information $I(\hat{x}_t,y_t)$ becomes

$$I(\hat{x}_t,y_t) = H(\hat{x}_t) - H(\hat{x}_t|y_t),$$
$$= \frac{1}{2} \ln \left(\frac{\det \Gamma_t}{\det P_t}\right), \quad I(\hat{x}_o,y_o) = 0. \tag{3-78}$$

Equation (3-78) is the final result which will be used as a criterion of the observability of the stochastic system. According to the third property of the mutual information (3-17)

$$I(x_t,y_t) \geq I(\hat{x}_t,y_t). \tag{3-79}$$

Inequality in (3-79) is due to the information loss which is incurred during the approximation procedures. For the observability of the individual state, say i-th state, the following will be used

$$I_i(\hat{x}_i, y_t) = 1/2 \ln \left(\frac{\Gamma_{ii}}{P_{ii}}\right), \quad i = 1,2, \ldots, n, \tag{3-80}$$

where $\hat{x}_i$ is the i-th element of $\hat{x}_t$ and Γ_{ii}, P_{ii} are diagonal elements of Γ_t and P_t, respectively. Of course, Γ_t and P_t are computed as a part of state estimation. Thus, both are defined only when they are positive definite. The degree of observability at time t is easily computed by reading Γ_t, P_t and simple computation according to (3-78).

From (3-78) it is clear that for $I(\hat{x}_t, y_t)$ to be maximum, P_t must be minimum. If the minimum covariance P_t^* of the estimation error is obtained by the unbiased optimum estimator, then the maximum Fisher information is obtained, also [47], i.e., Cramer-Rao lower bound is obtained in this case. So,

$$P_t \geq P_t^* = J^{-1} . \tag{3-81}$$

To observe observability variation due to Γ_t and P_t changes, consider the simple linear system

$$dx_t = F(t)x_t dt + G(t)dw_t, \tag{3-82}$$

$$dy_t = H(t)x_t dt + dv_t, \tag{3-83}$$

where w_t and v_t have strength $Q(t)$ and $R(t)$, respectively. Covariances P_t and Γ_t, then, satisfy

$$\dot{\Gamma}_t = F(t)\,\Gamma_t + \Gamma_t F^T(t) + G(t)Q(t)G^T(t), \quad \Gamma_{to} = \Gamma_o, \tag{3-84}$$

$$\dot{P}_t = F(t)P_t + P_tF^T(t) + G(t)Q(t)G^T(t) - P_tH^T(t)R^{-1}(t)H(t)P_t,$$
$$P_{to} = P_o = \Gamma_o \quad . \tag{3-85}$$

Mutual information change in this system arises in two ways. One is through initial information Γ_{to}, P_{to}, and another is through measurement mechanism H(t).

Even assuming the same initial information such that $\Gamma_{to} = P_{to}$, the magnitude of Γ_{to} or P_{to} plays an important role at the final time. For example, a large initial covariance make system observability grow fast at the initial stage since $\dot{P}_t$ in (3-85) tends to decrease rapidly to its steady state if the filter works properly. The main reason for this is due to the last term of (3-85). However, Γ_t does not change rapidly since there is no such term in (3-84). Some guidelines of choosing proper initial covariance in simulation can be found in [60]. But choosing of specific value of P_{to} is based on the designer's "degree of confidence" of $\hat{x}_{to}$ relative to unknown true value x_{to}, in most cases. If too optimistic (choosing too small P_{to} by overconfidence), then information growth may be very slow even in the case where the system is deterministically observable. So, tuning of the filter is compromising between two extremes by trial and error until obtaining desirable performance.

The effect of measured information on observability is seen also through the last term $P_tH^T(t)R^{-1}(t)H(t)P_t$ in (3-85). Especially measurement structure matrix H(t) and noise strength R(t) are important here. So, if this term is negligible due to some reason, for example, $R(t) \to \infty$ and/or $H(t) \to 0$, then the changing rate $\dot{\Gamma}_t$ and $\dot{P}_t$ in (3-84), (3-85) will be almost the same. Thus, mutual information or observability will not grow any more in this case.

A short discussion of the relation between the deterministic observability condition and the mutual information for the linear system case is made next.

3. Linear systems: deterministic and stochastic observability.

Mutual information, or formally, stochastic observability of a system is approximated as the log ratio of the two covariances Γ_t and P_t. So, the relationship between deterministic and stochastic observability is characterized by the relation between these matrices and the satisfaction of the deterministic observability condition. To avoid complexity consider a stochastic linear (time-invariant) system

$$dx_t = Fx_t dt + g dw_t, \tag{3-86}$$

$$dy_t = Hx_t dt + dv_t , \tag{3-87}$$

where w_t, v_t have covariances Q, R respectively. For this system a theorem is cited from [56].

Theorem 3-1

Let the system (3-86), (3-87) satisfy the deterministic observability condition, i.e., observability matrix

$$V = \begin{bmatrix} H \\ HF \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ HF^{n-1} \end{bmatrix} , \qquad (3\text{-}88)$$

has rank n. Then the covariance matrix P_t of the system is uniformly bounded and converges to its limit P_∞ , where

$$\begin{aligned} P_\infty &= \lim_{t\to\infty} P_t, \\ &= \lim_{t\to\infty} E[(x_t-\hat{x}_t)(x_t-\hat{x}_t)^T], \end{aligned} \qquad (3\text{-}89)$$

is the solution of

$$FP_\infty + P_\infty F^T + GQG^T - P_\infty H^T R^{-1} HP_\infty = 0. \qquad (3\text{-}90)$$

Uniform boundedness is proved by changing the system dynamic equation (3-86) into its auxiliary control problem.

Remarks

1) For uniform boundedness and convergence of P_t to P_∞ at least an unstable state, if exists, must be observable deterministically [61]. If the system is stable then the observability rank condition (3-88) can be dropped for boundedness of P_t.

2) If matrix pair (F,G) in (3-86) constitutes a controllable system, i.e., controllability matrix M (when w_t is considered as an input control).

$$M = (G,\ FG,\ \ldots.,\ F^{n-1}G),$$

has rank n, then P_t is positive definite [36], [56].

From the previous theorem, the following result can be proved. In the theorem, covariance matrix manipulation identities are cited from the results of Balakrishnan [61].

Theorem 3-2

If the time invariant linear system (3-86), 3-87) is deterministically observable, then it becomes stochastically more observable in the sense that the mutual information $I(\hat{x}_t, y_t)$ increases with time.

Proof

By Theorem 3-1 covariance P_t which is the solution of

$$\dot{P}_t = FP_t + P_tF^T + GQG^T - P_tH^TR^{-1}HP_t, \quad P_{to} = P_o, \tag{3-91}$$

converges uniformly to P_∞ if the system is deterministically observable. Now consider covariance Γ_t where

$$\dot{\Gamma}_t = F\Gamma_t + \Gamma_t F^T + GQG^T, \quad \Gamma_{to} = P_o \quad . \tag{3-92}$$

We want to show next the relation $\Gamma_t \geq P_t$. Differentiation of (3-92) gives

$$\ddot{\Gamma}_t = F\dot{\Gamma}_t + \dot{\Gamma}_t F^T, \tag{3-93}$$

and the solution of (3-93) becomes

$$\dot{\Gamma}_t = \phi(t)\ \dot{\Gamma}_o \phi^T(t), \tag{3-94}$$

where $\dot{\phi}(t) = F\phi(t),\ \phi(0) = I$. (3-95)

Using the same procedures as (3-91) gives

$$\ddot{P}_t = (F - P_t H^T R^{-1} H)\dot{P}_t + \dot{P}_t (F - P_t H^T R^{-1} H)^T , \tag{3-96}$$

and it's solution

$$\dot{P}_t = \psi(t)\dot{P}_o \psi^T(t), \tag{3-97}$$

with

$$\dot{\psi}(t) = (F - P_t H^T R^{-1} H)\psi(t),\ \psi(0) = I . \tag{3-98}$$

$\dot{\Gamma}_o$ in (3-94) and $\dot{P}_o$ in (3-97) are determined by letting t=0 in (3-92) and (3-91), respectively. Let eigenvalues of F in (3-95) be λ_1, λ_2, ... λ_n such that

$\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda n,$

and eigenvalues of $(F-P_t H^T R^{-1} H)$ in (3-98) be ρ_1, ρ_2, ..., ρ_n such that $\rho_1 \geq \rho_2, \ldots, \geq \rho_n$, then, due to the term $P_t H^T R^{-1} H$ in (3-98) the relation

$\lambda_1 > \rho_1$,
$\lambda_2 > \rho_2$,
.
.
.
$\lambda_n > \rho_n$,

holds. Considering (3-94) and (3-97) and both having the same initial conditions

$$D(t) = \Gamma_t - P_t > 0. \tag{3-99}$$

Further the difference D(t) is monotone in time since all the eigenvalues appear as an exponential form in ϕ and ψ by the Caley-Hamilton theorem. So, convergence of P_t to P_∞ and monotonicity of D(t) says that $I(\hat{x}_t, y_t)$ grows monotonically from (3-78). Thus, the system becomes more observable as time progresses. **

More intuitive relations of the two observability concepts can be derived when absence of process noise w_t is assumed. In this case, using a matrix inversion identity [62] for (3-91)

$$P_t^{-1} = -F^T P_t^{-1} - P_t^{-1} F + H^T R^{-1} H, \quad P_{to}^{-1} = P_o^{-1}, \tag{3-100}$$

Then, the solution of (3-100) is

$$\begin{aligned} P_t^{-1} &= \phi^T(t_o.t) P_o^{-1} \phi(t_o,t) + \int_{t_o}^{t} \phi^T(s,t) H^T R^{-1} H \phi(s,t) ds \quad , \\ &= J_i(t,t_o) + J_o(t,t_o). \end{aligned} \tag{3-101}$$

Notice, that P_t^{-1} in (3-101) is exactly the same as the Fisher information matrix $J(t,t_o)$ in (3-65). Using the same procedures for the covariance Γ_t in (3-92), yields

$$\dot{\Gamma}_t^{-1} = -F^T \Gamma_t^{-1} - \Gamma_t^{-1} F, \quad \Gamma_{to}^{-1} = \Gamma_o^{-1} = P_o^{-1}, \tag{3-102}$$

and it's solution

$$\Gamma_t^{-1} = \phi^T(t_o,t) \; \Gamma_o^{-1} \; \phi(t_o,t). \tag{3-103}$$

Assume here that P_o^{-1} is nonsingular, i.e., there is some prior information about all states. If the system is deterministically observable, i.e.,the second term of (3-101) is positive definite, then comparison of (3-101) and (3-103) considering, again, the definition of (3-78),shows that $I(\hat{x}_t, y_t)$ increases until P_t reaches to its limit.

Now consider there exists system noise w_t. Then from (3-92) its solution is

$$\begin{aligned}\Gamma_t &= \phi(t,t_o)\Gamma_o\phi^T(t,t_o) + \int_{t_o}^{t}\phi(t,s)G(s)Q(s)G^T(s)\phi^T(t,s)ds, \\ &= C_i(t,t_o) + C_o(t,t_o).\end{aligned} \tag{3-104}$$

Notice that the matrix $C_o(t,t_o)$ is termed, traditionally, as a stochastic controllability matrix. So, from (3-101), (3-78), the classical concept of stochastic observability and controllability affect the mutual information as follows: $I(\hat{x}_t, y_t)$ is increased by both increased quantity of controllability and observability. Contribution of the stochastic controllability matrix $C_o(t,t_o)$ is made via increasing Γ_t in (3-104), and thus increasing $I(\hat{x}_t, y_t)$ in (3-78).

Contribution of the increased stochastic observability $J_o(t,t_o)$ is via decreasing P_t in (3-101) and thus increases $I(\hat{x}_t, y_t)$ since P_t enters (3-78) into its denominator.

CHAPTER 4: INFORMATION STRUCTURAL ANALYSIS OF BOT AND ARRAY SONAR SYSTEMS

Simulation results of the information structural analysis of two important examples of nonlinear stochastic systems are presented here. System models are taken as the same underwater tracking problems as in Chapter 2 to relate with the deterministic observability conditions. To fit more practical situations in both BOT and array SONAR tracking examples, it is assumed that the information acquisition about the system states is made through the discrete measurement mechanism. However, the evolution of the system states are assumed to be the time-continuous . Thus, the estimation of the system states are implemented by the discrete-observation, continuous-state filter algorithm.

Before presenting this, the following simple linear system results are provided to give a clear understanding of the current approach.

The term "observability" in this chapter , of course , means the degree of the observability in terms of mutual information.

4.1 Falling-body example.

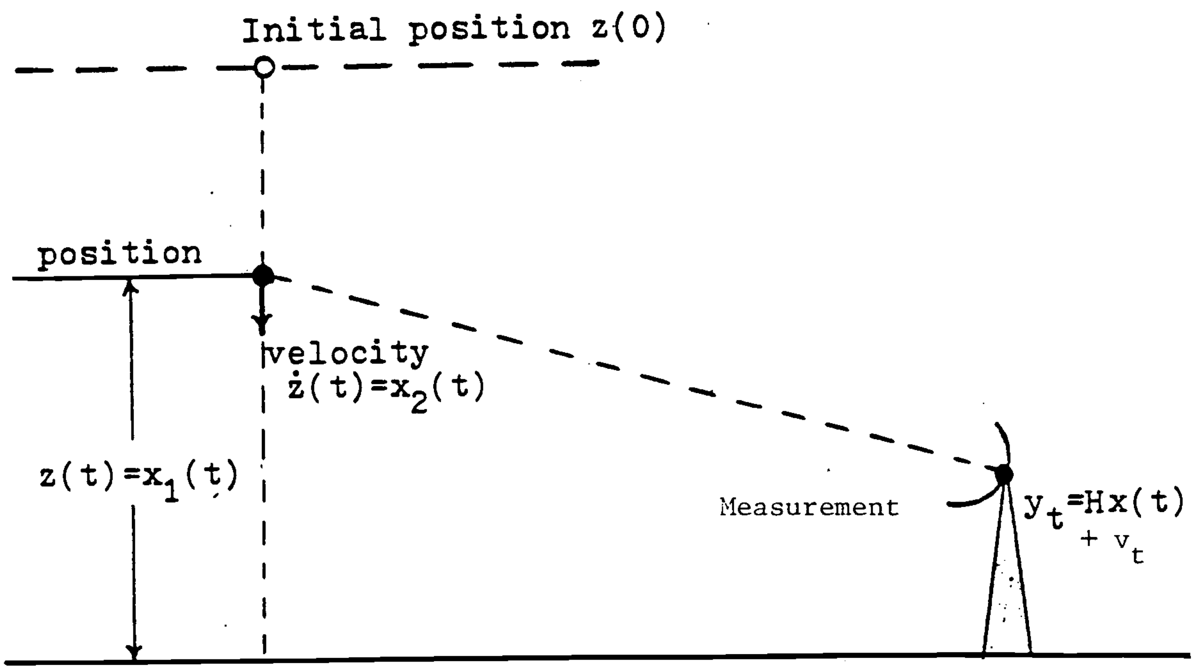


Figure 7, Measurements of falling-body

Consider a noise free second-order system representing a falling-body in a constant gravitational field g (Figure 7).

$$\ddot{z}(t) = -g, \; t \geq 0 \; . \tag{4-1}$$

Let the position variable $x_1=z$, and velocity variable $x_2=\dot{z}$. Then

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ -g \end{bmatrix} \quad , \tag{4-2}$$

This system is observed by the noisy measurement device which can be expressed by

$$y_t = Hx(t) + v_t, \tag{4-3}$$

where random white Gaussian noise has covariance R(t). Simple test shows that the deterministic portion of the system is observable if one observes position x_1 and unobservable if observes velocity x_2. Intuitively this is clear because if one measures x_1, then it's derivative gives velocity x_2. No other information is required to describe the system. However, when one measures velocity x_2, then as integration is required to get position x_1. I.e.,

$$x_1(t) = \int_o^t x_2(t)dt + x_1(0) , \tag{4-4}$$

but $x_1(0)$ can not be determined from any measurement data. So, the system is unobservable in this case.

Using the usual Kalman-Bucy filter with Gaussian noise, mutual information $I(\hat{x}_t, y_t)$ is compared in Table 2. In the deterministically observable case (by measuring position x_1) mutual information of the total system (T in Table 2) grows up to 5.7 from zero at final time 20 sec. Position (p) and velocity (v) grow 4.9 and 1.8 respectively. But for the unobservable case (by measuring x_2) corresponding observability grows: T = 2.8, P = 2.3, v = 1.8. To compare the significance of the logarithm scale a linear scale is also shown. For

Table 2, Observability of the falling-body

	Observable system						Unobservable system					
obs. / t (sec)	Linear scale			Log scale			Linear scale			Log scale		
	O_T	O_P	O_V	O_T	O_P	O_V	O_T	O_P	O_V	O_T	O_P	O_V
0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0
2	3.3	2.8	1.1	1.2	1.0	0.1	1.8	1.2	1.8	0.6	0.2	0.6
4	7.5	5.1	1.7	2.0	1.6	0.5	2.5	1.6	2.4	0.9	0.5	0.9
6	15.3	9.8	2.5	2.7	2.3	0.9	3.3	2.3	3.0	1.2	0.8	1.1
8	28.0	17.7	3.3	3.3	2.9	1.2	4.3	3.2	3.5	1.5	1.2	1.3
10	45.0	28.2	3.9	3.8	3.3	1.4	5.5	4.1	4.0	1.7	1.4	1.4
12	73.0	43.0	4.4	4.3	3.8	1.5	7.1	5.2	4.4	1.9	1.6	1.5
14	109.0	61.0	4.8	4.7	4.1	1.6	9.0	6.3	4.8	2.2	1.8	1.6
16	158.0	82.0	5.2	5.1	4.4	1.7	11.3	7.6	5.2	2.4	2.0	1.7
18	223.	107.	5.6	5.4	4.6	1.7	14.0	9.0	5.6	2.6	2.2	1.7
20	307.	137.	6.0	5.7	4.9	1.8	17.2	10.4	6.0	2.8	2.3	1.8

the unobservable system, only the observed velocity variable keeps the same level of observable system (1.8).

The degree of observability directly affects the filtering error. This is analyzed in Figures 8 and 9. Figure 8 shows the deterministically observable case with initial errors of 20 m in position and 5 m/s in velocity. Since position is measured in this case, its information is dominant and thus the corresponding error decreases rapidly. The velocity error is, also, quite small at the final time since x_2 is also an observable variable. However, Figure 9 is much different than Figure 8 even with the same initial errors. Since velocity is observed here, position is an unobservable variable, and thus carries very large errors up to the final time. The velocity variable (observed quantity here) shows quite satisfactory performance compared to the position variable.

Table 3 shows the effect of initial information $P_o(=\Gamma_o)$ on the observability+ and filtering error. In general, as larger initial information is assumed (smaller P_o) the system obtains smaller final information. Note also that in most cases information acquisition is quite fast in the initial stage. This phenomenon is more significant as P_o increases. It implies that the filter forgets the initial uncertainty very quickly when the assumed initial information is small. This is one of the most desirable features of the Kalman-Bucy filter. Practical experience suggests, however, that in stochastic nonlinear filter design, with non-negligible nonlinearity, it is desirable not to use overly pessimistic initial-error

+Observability again refers to $I(\hat{x}_t, y_t)$ for all the following data.

Fig. 8 Observable Falling Body

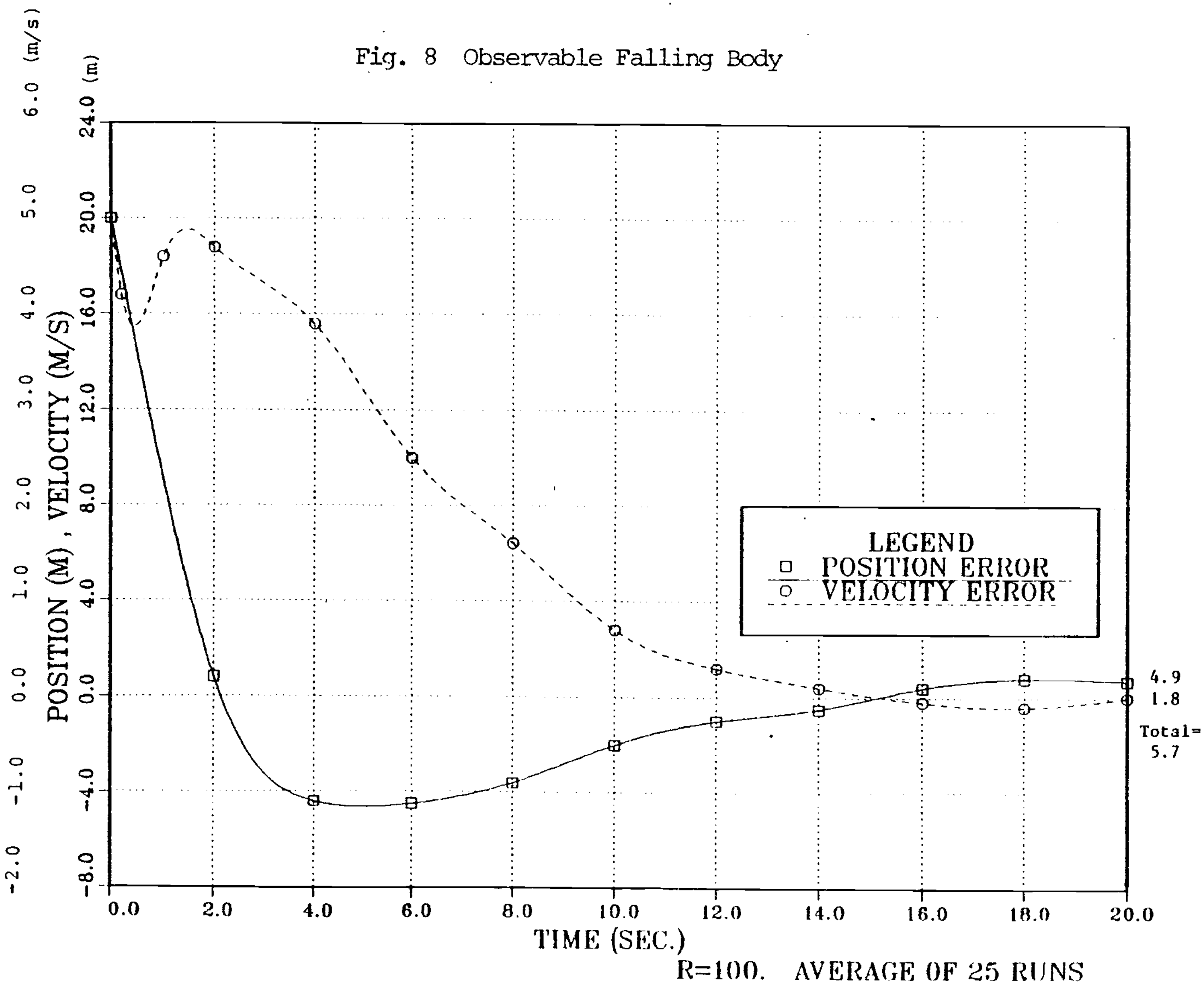


Fig. 9 Unobservable Falling Body

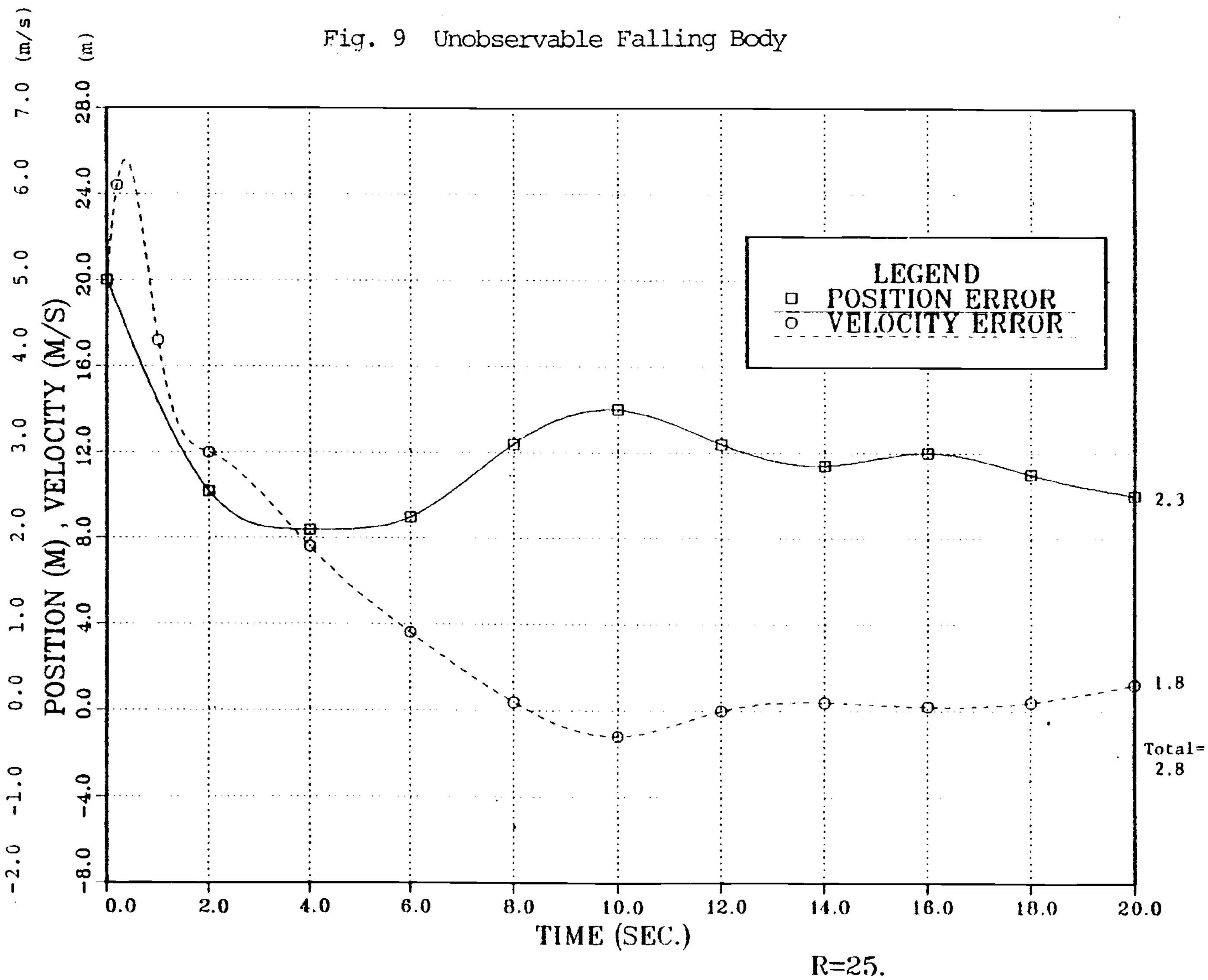


Table 3, Effects of initial information P_o on observability

t	$P_o = \begin{bmatrix} 100 & \\ & 10 \end{bmatrix}$					$P_o = \begin{bmatrix} 500 & \\ & 50 \end{bmatrix}$					$P_o = \begin{bmatrix} 2500 & \\ & 500 \end{bmatrix}$				
(sec)	O_T	O_p	O_v	e_p	e_v	O_T	O_p	O_v	e_p	e_v	O_T	O_p	O_v	e_p	e_v
	Observable system														
0	0.0	0.0	0.0	20.0	5.0	0.0	0.0	0.0	20.0	5.0	0.0	0.0	0.0	20.0	5.0
5	2.4	2.0	0.7	4.6	2.8	4.5	3.2	1.6	1.3	1.0	8.1	5.2	3.7	0.5	0.01
10	3.9	3.4	1.4	2.6	0.7	5.7	4.7	2.4	1.0	0.3	9.1	6.9	4.6	0.4	0.1
15	4.9	4.3	1.6	0.6	0.1	6.5	5.6	2.6	0.9	0.2	9.4	7.7	4.6	1.0	0.2
20	5.7	4.9	1.8	0.1	0.02	7.1	6.1	2.6	0.1	0.1	9.8	8.3	4.6	0.1	0.1
	Unobservable system														
0	0.0	0.0	0.0	20.0	5.0	0.0	0.0	0.0	20.0	5.0	0.0	0.0	0.0	20.0	5.0
5	1.0	0.7	1.0	15.8	0.9	2.1	1.0	2.1	22.0	0.03	4.4	2.7	4.3	27.0	0.3
10	1.7	1.4	1.4	15.8	0.3	2.7	2.0	2.5	23.0	0.05	4.7	2.9	4.6	27.0	0.2
15	2.3	1.9	1.6	15.6	0.1	3.2	2.6	2.6	23.0	0.02	4.9	3.7	4.6	27.0	0.04
20	2.8	2.3	1.8	16.5	0.04	3.6	3.1	2.6	23.4	0.1	5.2	4.2	4.6	28.0	0.1

Average of 25 runs

Unit : e_p; m , e_v; m/s

covariances since a large P_o could excessively dampen the system dynamics and filter gain matrix and thus reject some of the valuable measurement data in spite of fast information pick up from the measurement mechanism[64]. This phenomenon can be found in the position error (e_p) when the system is deterministically unobservable with a high value P_o. An opposite direction, i.e., overly optimistic P_o, sometimes, makes the response of the filter too slow.

As a summary, system observability is strong with strong position and velocity observability when the system is deterministically observable. But it is weak when the system is deterministically unobservable. Since position is an unobservable state in the latter case, its poor observability generates large filtering errors during the observed period.

4-2 BOT system and Information analysis.

It is well known that a BOT system is observable only when relative maneuvering exists. It is checked, again in Chapter Two. using so called, mixed-coordinate system (see also [33],[63]). Here the same problem is used to analyze and compare the observability content in terms of the information theoretic point of view. For comparison, two more popular coordinate systems - rectangular and modified polar (MP) coordinates- are adopted in this section. System description of the individual coordinates are presented in Table 4, with proper dimensional noises. Measurement equations are written in discrete form for future conveniences. Using the same procedures as derived in Chapter Two, deterministic observability for the remaining two coordinates can be checked. Long algebraic manipulation shows also that the system is observable when relative maneuvering exists. This is not surprising since deterministic observability is not affected by the coordinate transformation.

Note in Table 4 that the system equation is linear and the observation equation is nonlinear in rectangular coordinates and vice versa in the modified-polar and mixed coordinates. The variables r, v, a, β represent range, velocity, acceleration, bearing, respectively.

Table 4, System description of different coordinates

	Rectangular	Modified polar	Mixed
State variable	$x(t)=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}r_x\\v_x\\r_y\\v_y\end{bmatrix}$	$x(t)=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}\dot{\beta}\\ \dot{r}/r\\ \beta\\ 1/r\end{bmatrix}$	$x(t)=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}\beta\\r\\v_x\\v_y\end{bmatrix}$
State eqs.	$\dot{x}(t)=\begin{bmatrix}0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{bmatrix}x(t)+\begin{matrix}\\ \\ \\ a_y\end{matrix}$ $+w(t)$	$\dot{x}(t)=\begin{bmatrix}-x_1x_2-a_yx_4\sin(x_3)\\x_1^2-x_2^2+a_yx_4\cos(x_3)\\x_1\\-x_2x_4\end{bmatrix}$ $+\ w(t)$	$\dot{x}(t)=\begin{bmatrix}\dfrac{x_3\cos(x_1)-x_4\sin(x_1)}{x_2}\\x_3\sin(x_1)+x_4\cos(x_1)\\0\\a_y\end{bmatrix}$ $+\ w(t)$
Meas. eqs.	$y_k=h(x_k,k)+v_k$ $=\tan^{-1}\left(\dfrac{x_1}{x_3}\right)+v_k$	$y_k=H\ x_k+v_k$ $=[\,0\ 0\ 1\ 0\,]\ x_k+v_k$	$y_k=H\ x_k+v_k$ $=[\,1\ 0\ 0\ 0\,]\ x_k+v_k$

To implement a not-excessively-complicated nonlinear filter of a continuous-system discrete-observation type, a truncated - second order filter [34], [36] is considered. With the same target and observer (or ownship) configuration as in Figure 2, one-directional maneuvering is assumed as

$$a_x(t) = 0 \ ,$$
$$a_y(t) = -0.25 \cos(0.005t) \quad m/s^2 \ , \tag{4-5}$$

and initial states are assumed Gaussian with proper mean.
Other parameters used are

T (Sampling interval) = 10 sec,

Δt (time update interval between observation) = 1 sec,

r(0) (initial range) = 8000 m,

v_{Tx} (target vel. in x-direction) = 10m/s ~ 20 kt ,

$v_{Ty} = 0$

v_{Ox} (observer vel. in x-direction) = 15 m/s ~ 30 kt ,

$v_{Oy} = 5 \sin(0.005t)$ m/s.

Measurement noise sequence and system noise are assumed to be, also, Gaussian with variance R_k and Q(t), respectively.

Under the assumption of near symmetric form of density and negligible third and higher-order moments, a modified version of the truncated Gaussian second-order filter is implemented.

Continuous-discrete type filter is, commonly, implemented in two stages. The first stage, a measurement-update stage, processes observed data according to the discrete filter. The second stage performs the time propagation integral of the first and second moment (or higher moments if necessary) of the state between the observation interval according to the continuous fashion.

This form of filter is particulary suited to us due to the nature of the underwater SONAR system where the data-acquisition interval is quite long compared to the data-processing rate. The actual algorithm is summarized next [34].

1. Measurement Update

At the sampling instant t_k, abbreviated by k, mean and covariance are computed as

$$\hat{x}_{k+1} = \hat{x}_k + K_k[y_k - h(\hat{x}_k,k) - \hat{b}_m(k)], \tag{4-6}$$

$$P_{k+1} = P_k - K_k H(\hat{x}_k,k)P_k, \tag{4-7}$$

where gain K_k is given by

$$K_k = P_k H^T(\hat{x}_k,k)A_k^{-1}, \tag{4-8}$$

$$A_k = H(\hat{x}_k,k)P_k H^T(\hat{x}_k,k) - \hat{b}_m \hat{b}_m^T + R_k, \tag{4-9}$$

$$H(\hat{x}_k,k) = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}_k}, \text{ h is measurement function,}$$

and where the bias correction term $\hat{b}_m$ is an m-vector with i-th component

$$\hat{b}_{mi}(k) = 0.5\mathrm{tr}\left\{\frac{\partial^2 h_i}{\partial x\,\partial x^T}P_k\right\}\Bigg|_{x=\hat{x}_k}, \quad i=1,2,\ldots,m, \tag{4-10}$$

m is a measurement dimension.

2. Time progagation between observations

Between observation intervals there is no measurement data, so $\hat{x}$ and P progate time forward according to the continuous filter with the initial conditions

$$\hat{x}_{ti} = \hat{x}_{k+1},\ P_{ti} = P_{k+1}.$$

Time integration of $\hat{x}$ and P at t, $t \varepsilon [t_k, t_{k+1}]$ becomes

$$\dot{\hat{x}}_t = f(\hat{x}_t, a_t, t) + \hat{b}_p \quad , \tag{4-11}$$

$$\dot{P}_t = F(\hat{x}_t, t) P_t + P_t F^T(\hat{x}_t, t) + \widehat{G_t Q(t) G_t^T} \ , \tag{4-12}$$

where f(.) is the system function with an extra parameter a_t and

$$F(\hat{x}_t, t) = \left. \frac{\partial f}{\partial x} \right|_{x=\hat{x}_t} .$$

Bias correction term $\hat{b}_p$ is an n-vector with i-th component

$$\hat{b}_{pi}(t) = 1/2 \ tr \left\{ \frac{\partial^2 f_i}{\partial x \partial x^T} P_t \right\} \Bigg|_{x=\hat{x}_t} , \tag{4-13}$$

and for system noise function $G_t(x_t)$

$$(G_t Q(t) G_t^T)_{ij} = \sum_{k,l=1}^{s} \Bigg[G_{ik} Q_{kl} G_{lj}^T + tr \left\{ \left(\frac{\partial G_{ik}}{\partial x} Q_{kl} \frac{\partial G_{lj}}{\partial x} \right) \right\} +$$

$$1/2 \ G_{ik} Q_{kl} tr\left\{ \frac{\partial^2 G_{lj}^T}{\partial x \partial x^T} P_t \right\} + 1/2 \ tr\left\{ P_t \frac{\partial^2 G_{ik}}{\partial x \partial x^T} \right\} Q_{kl} G_{lj}^T \Bigg] , \tag{4-14}$$

s is a dimension of system noise.

Thus at $t = t_{k+1}$, the initial condition of the first stage becomes, again

$$\hat{x}_k = \hat{x}_t,\ P_k = P_t, \tag{4-15}$$

and the same procedures repeat for new observed data.

3. Unobserved system covariance

Another covariance Γt is required to compute $I(\hat{x}_t, y_t)$. This is evaluated according to equation (4-12). Since no measurement is made here the measurement update is not necessary. Of course the reference point should be different with (4-12) except at the initial conditions.

With assigned parameters and algorithms, simulation is conducted for three different coordinates. The following are the results found from the analysis of the simulation for the first 40 minutes.

Tables 5, 6, 7 show the mutual information contents of the three coordinate systems with various parameter changing-system noise Q(t) and maneuvering a_y. Total system observability is most strong in the modified polar coordinates (Table 6). Rectangular and mixed coordinates show almost the same levels (Table 5 and 7). Of course, directly observed variables - bearing(β) in mixed and MP, range (r) in rectangular - exhibit the strongest observability in all cases. Inspection of all three tables show that system observability drops significantly as the maneuvering parameter changes from $a_y \neq 0$ (maneuvering exists) to $a_y = 0$ (non-maneuvering). This can be explained best by the deterministic observability. As seen in Chapter

Two, the sytem is observable deterministically only when manuevering exists.

Another notable observability decrease appears when there exists system noise ($Q \neq 0$). This is due to a change of mutual information quantity from $I_o(\hat{x}_t, y_t)$ to $I(\hat{x}_t, y_t)$ (See Chapter Three for notation).

Notice also that range information is most drastically influenced by the observer maneuvering (3.8 to 0.2 for mixed, 7.4 to -3.0 for modified polar, 4.3 to 0.9 (r_x) and 5.0 to 1.8 (r_y) for rectangular coordinates, respectively). In spite of the strongest total observbility, contribution by the range observability to the total observability is the most negligible in the MP case.

Velocity observability remains very poor, generally, in the non-maneuvering case, or when system noise exists.

The effects of the degree of observability on the range and velocity estimation error are shown in Figures 10 to 13. Range errors (Figures 10 to 12) converge toward zero for the maneuvering and without system noise case (even different convergence rates), but not for other cases. For all three coordinates, range errors seems to diverge when $a_y = 0$ and $Q = 0$. At least, they do not converge to zero in the non-maneuvering case in any sense.

Relative poor observability of the range variable in the MP system may be the reason why the range error exhibits some oscillatory property in Figure 11.

A more desirable convergence is shown by the mixed coordinates if $a_Y \neq 0$, $Q = 0$ (Figure 10). Note that the vertical scale in the MP system is different than the other two coordinates.

Careful comparison of the observability tables and corresponding estimation error figures shows that they are very closely related, i.e., the fast information growth interval corresponds to the abrupt error descreasing interval. Figure 13 shows that velocity errors converge to zero nicely for both mixed and rectangular coordinates when maneuvering exists. This may be due to the strong observability of these variables. Note that initial velocity error (1 m/s ~ 2kts) does not decrease satisfactorily when $a_Y = 0$ for both coordinates. The velocity variables are not available exclusively for the MP coordinates.

Table 5, Observability (Effects of Q and a_y) : Mixed

t (min)	$Q = 0.$										$Q \neq 0.$				
	$a_y \neq 0.$					$a_y = 0.$					$a_y \neq 0.$				
	Tot	β	r	v_x	v_y	Tot	β	r	v_x	v_y	Tot	β	r	v_x	v_y
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2.5	4.5	2.8	0.8	0.3	0.0	4.4	2.8	0.8	0.2	0.0	2.8	2.5	0.5	0.2	0.0
5	7.3	4.6	1.3	0.3	0.0	7.0	4.5	1.1	0.4	0.0	4.1	3.8	0.8	0.2	0.0
10	10.4	6.8	1.9	0.7	0.01	9.6	6.3	1.0	0.4	0.01	5.8	5.3	1.1	0.2	0.01
15	12.8	6.9	2.0	1.1	0.1	11.2	7.3	0.9	0.4	0.04	5.8	5.5	1.0	0.4	0.03
20	15.2	7.3	3.1	2.7	0.1	12.3	7.9	0.7	0.5	0.1	6.2	5.5	0.8	0.5	0.04
25	16.3	8.1	3.0	3.0	0.2	13.1	8.2	0.5	0.5	0.3	6.4	5.7	0.7	0.5	0.1
30	17.8	7.6	3.4	3.5	1.0	13.8	8.3	0.3	0.5	0.6	6.8	5.9	0.7	0.6	0.1
35	18.8	8.6	3.7	3.9	1.4	14.4	8.4	0.2	0.5	1.0	6.7	5.8	0.6	0.6	0.2
40	19.5	8.2	3.8	4.0	1.7	14.8	8.4	0.2	0.5	1.2	6.7	5.7	0.7	0.7	0.2

$R = (1^o)^2$, $T = 10$ sec.

Table 6, Observability (Effects of Q and a_y) : MP

t	Q = 0.										Q ≠ 0.				
	a_y ≠ 0.					a_y = 0.					a_y ≠ 0.				
	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2.5	10.2	5.3	-0.1	7.8	-0.1	9.9	5.9	-0.1	7.7	-0.2	7.3	4.8	-0.2	7.2	-0.2
5	13.9	5.3	1.4	8.9	0.5	12.6	5.4	0.4	9.0	-0.6	8.2	4.1	0.3	8.4	0.2
10	18.5	1.3	0.2	9.8	1.2	15.3	5.9	1.8	10.5	-1.4	8.8	0.5	-1.0	9.3	0.1
15	24.8	5.0	5.5	10.5	4.4	16.1	5.9	2.5	11.3	-1.9	11.5	2.5	2.4	9.4	1.8
20	27.8	5.3	5.3	11.4	5.2	16.6	5.9	3.0	11.8	-2.3	12.0	1.1	1.1	9.4	1.9
25	28.4	3.2	5.0	10.8	5.4	17.1	6.0	3.5	12.2	-2.6	12.2	1.0	1.4	9.3	2.2
30	30.9	3.2	6.1	11.3	6.8	17.4	6.1	3.8	12.5	-2.7	12.3	1.0	-0.5	9.5	2.3
35	32.6	7.2	8.5	12.2	7.2	17.7	6.2	4.2	12.7	-2.9	13.0	2.0	3.6	9.3	2.5
40	33.8	8.2	8.9	12.0	7.4	17.9	6.3	4.5	13.0	-3.0	13.4	1.9	3.0	9.2	2.6

$R = (1^o)^2$, $T = 10$ sec.

Table 7, Observability (Effects of Q and a_y) : Rec.

t	Q = 0.										Q ≠ 0.				
	$a_y \neq 0.$					$a_y = 0.$					$a_y \neq 0.$				
	Tot	r_x	v_x	r_y	v_y	Tot	r_x	v_x	r_y	v_y	Tot	r_x	v_x	r_y	v_y
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2.5	8.6	4.9	0.3	0.7	0.0	8.3	5.2	0.2	0.8	0.0	7.8	4.8	0.3	0.7	0.0
5	11.4	3.6	0.4	1.2	0.0	10.9	4.1	0.3	1.4	0.0	9.7	3.6	0.4	1.1	0.0
10	15.0	2.7	0.7	1.9	0.1	13.6	2.8	0.3	1.6	0.1	11.5	2.5	0.5	1.7	0.1
15	17.0	2.4	1.1	2.3	0.6	15.2	2.1	0.3	1.6	0.4	11.3	2.1	0.7	1.8	0.3
20	19.5	3.1	2.1	3.0	0.7	16.5	1.7	0.3	1.6	1.0	11.5	2.3	1.3	2.0	0.3
25	21.0	3.1	2.3	3.0	1.2	17.5	1.4	0.3	1.7	1.6	11.5	2.2	1.3	2.2	0.6
30	23.3	4.0	3.5	4.3	2.7	18.5	1.2	0.3	1.7	2.3	12.1	2.5	1.5	3.1	0.9
35	24.5	4.2	3.8	4.9	3.3	19.2	1.0	0.4	1.7	2.8	11.9	2.4	1.5	3.2	1.1
40	25.2	4.3	3.9	5.0	3.7	19.5	0.9	0.4	1.8	3.2	11.9	2.6	1.7	3.2	1.1

$R = (1^{\circ})^2$, $T = 10$ sec.

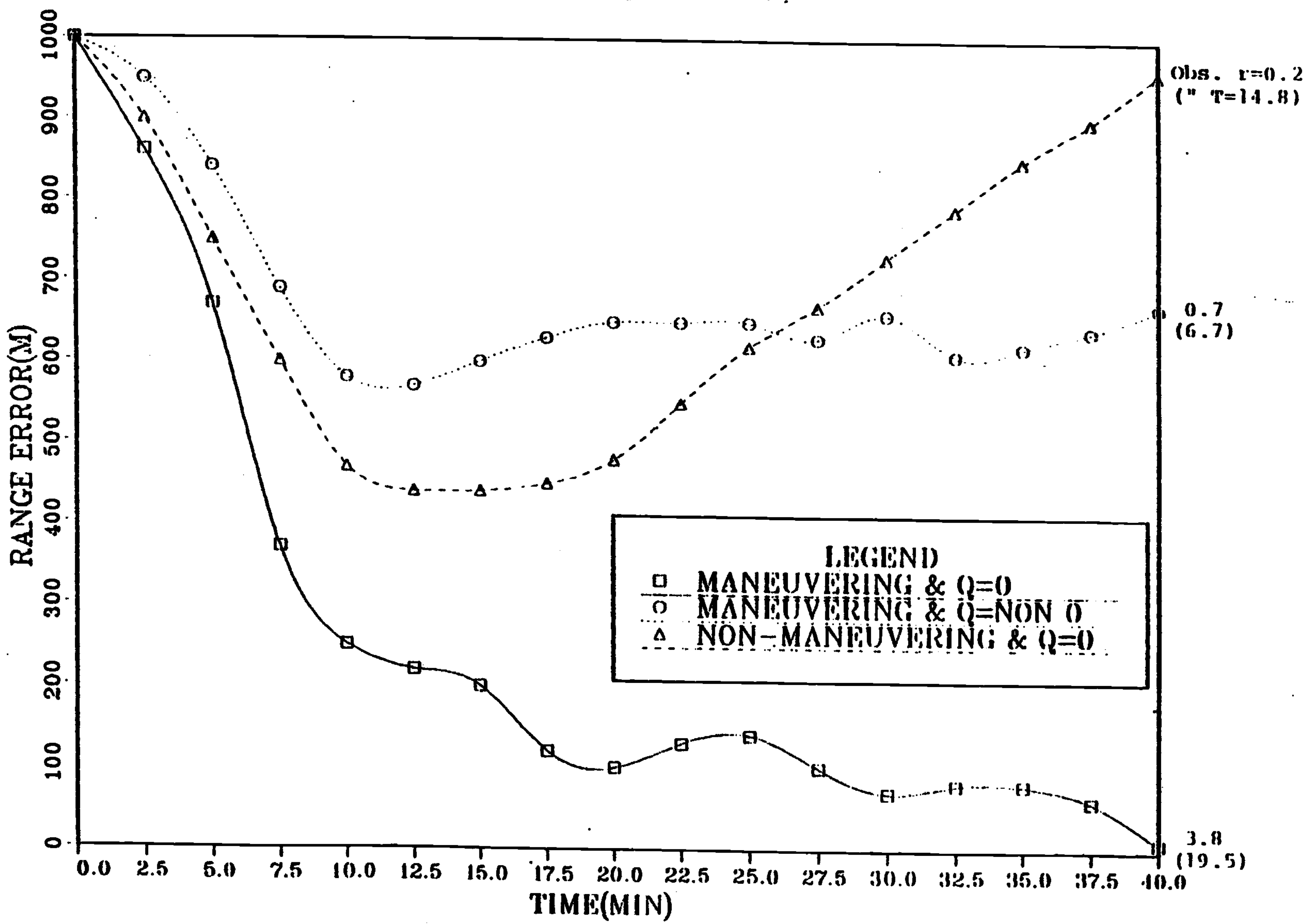


Fig. 10 Obs. and Range Error (Mixed)

Fig 11. Obs. and Range Error (MP)

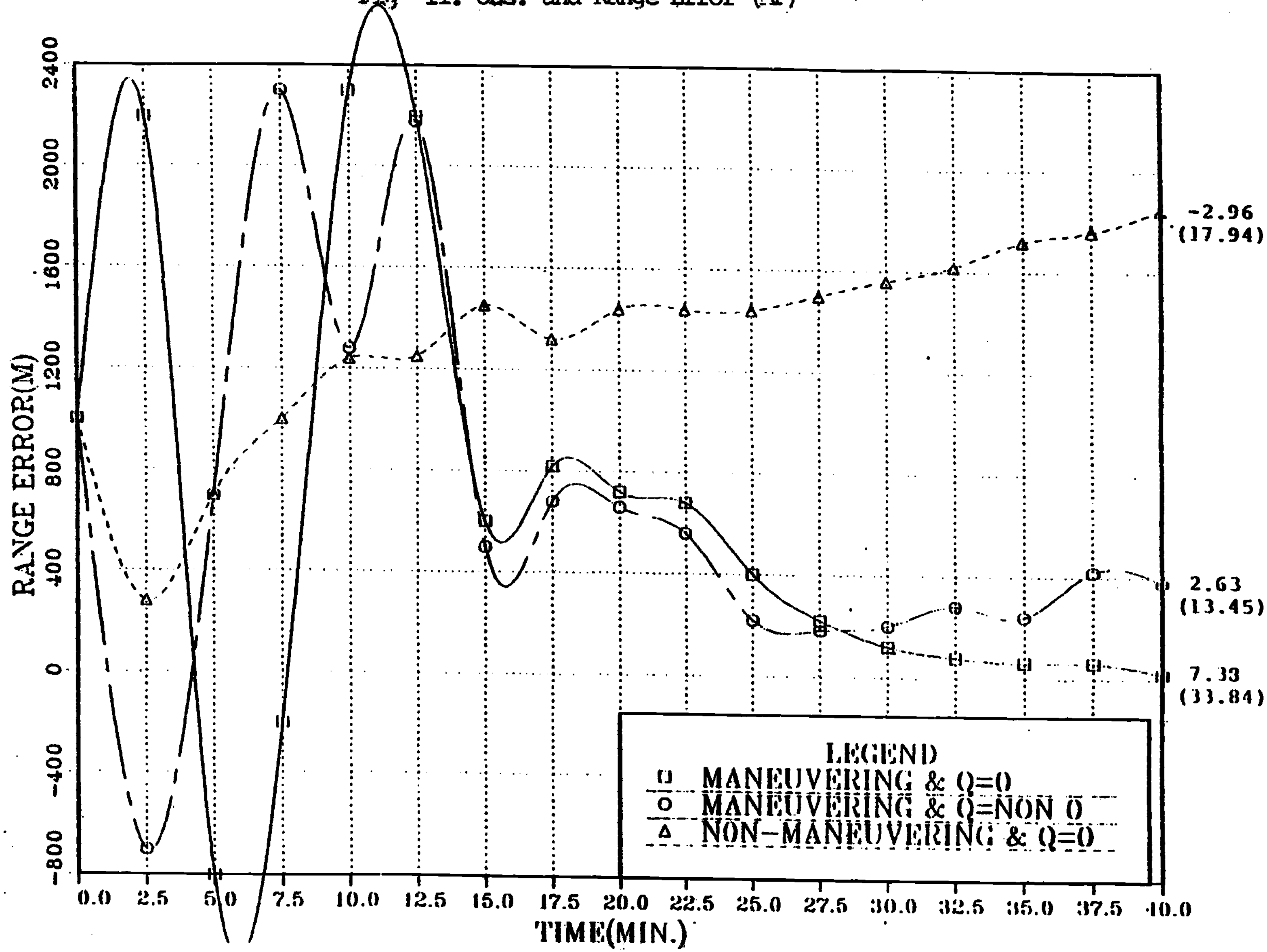


Fig. 12 Obs. and Range Error (REC.)

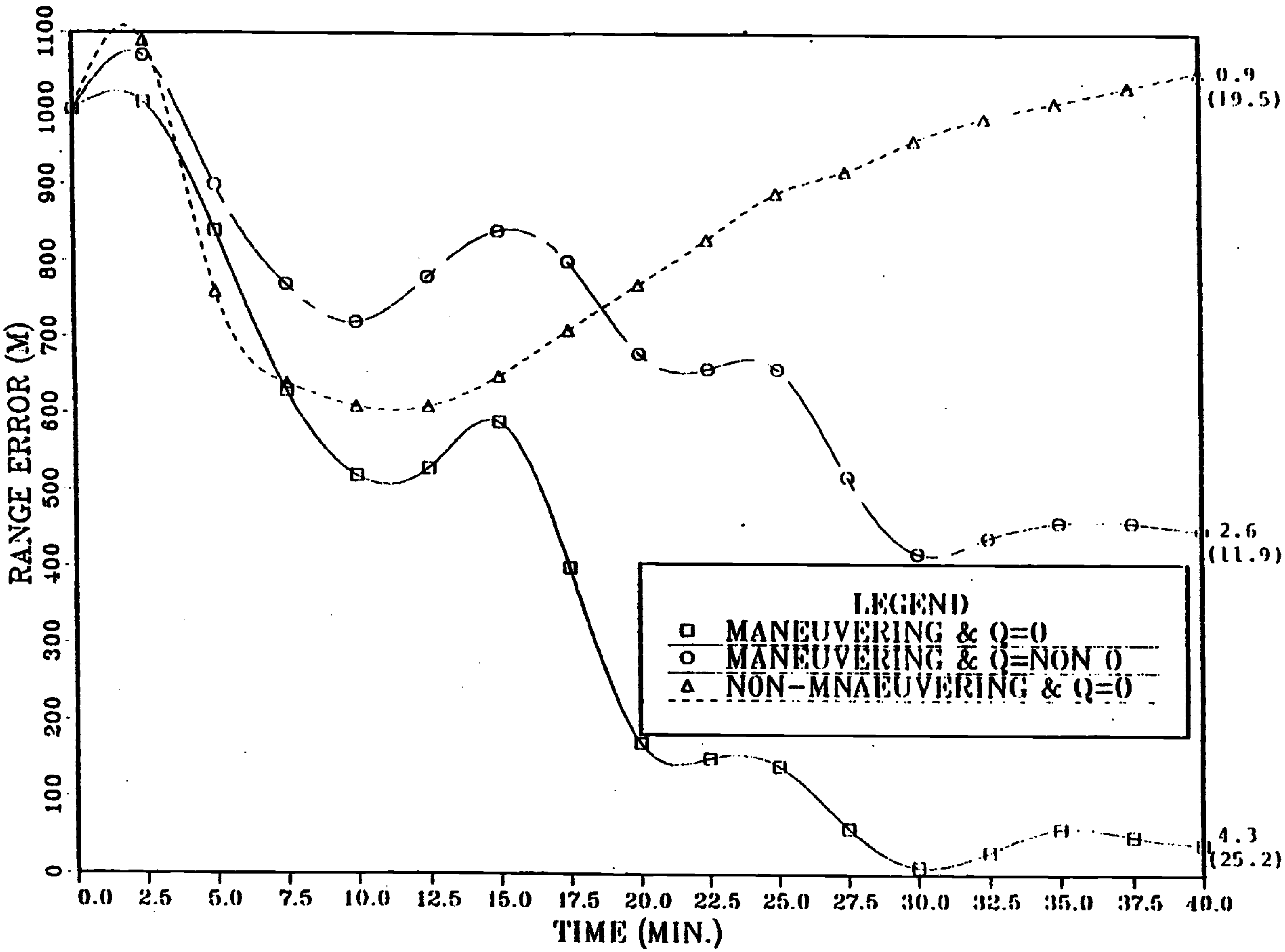
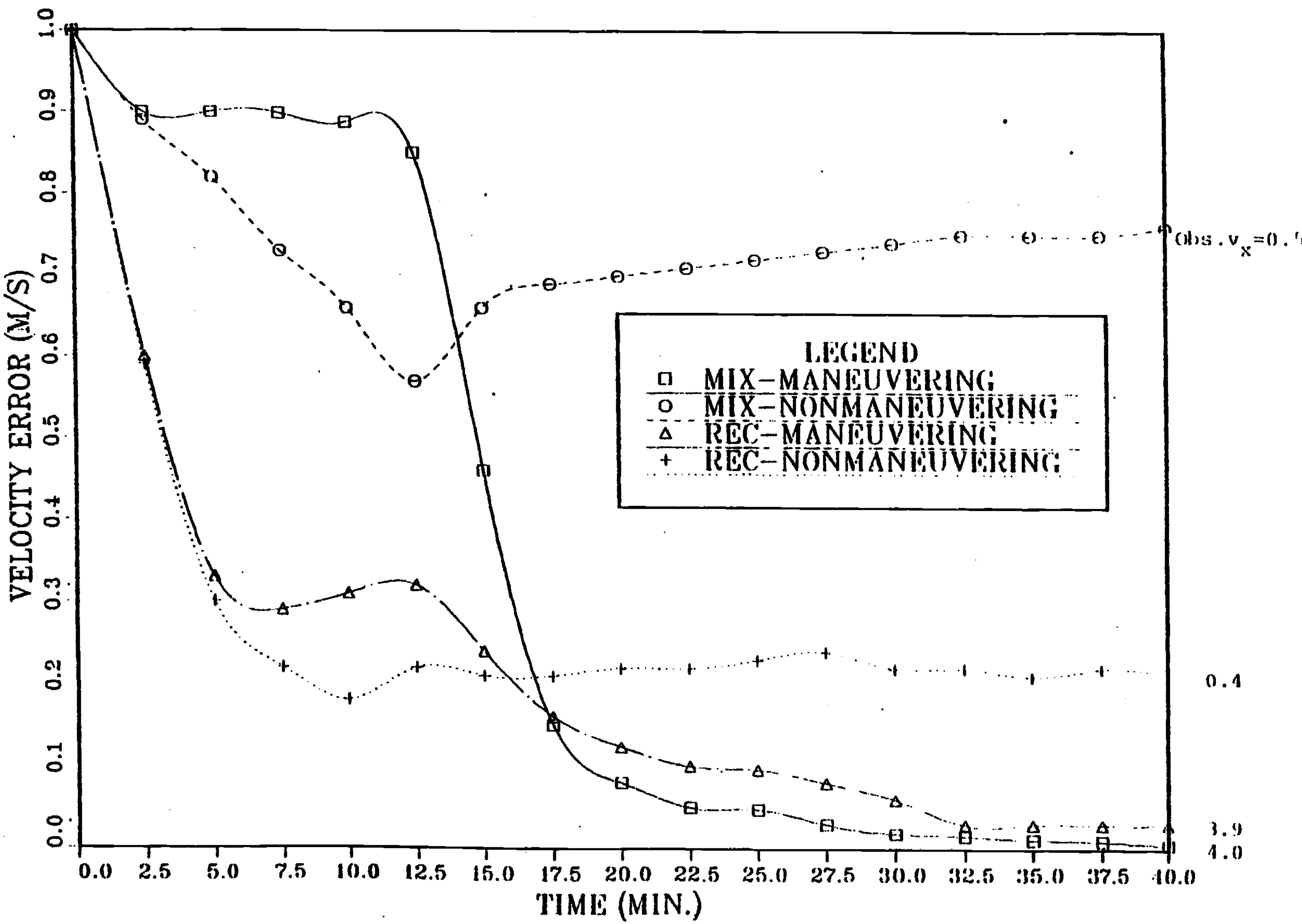


Fig 13. Obs. and Velocity Error



Observability analysis in terms of measurement noise R is shown in Tables 8, 9 and 10 with corresponding range error (e_r). As expected, observability decreases as the noise level increases. Particularily observability of the target speed becomes very poor when high noise level is presented.

Comparing the range errors for all coordinates shows that mixed coordinates exhibit the smallest errors even with the high noise level. This may be an extremely important characteristic from practical point of view.

Due to the fast information pick-up in the early stage, range error drops very quickly for the mixed coordinates. For example, within one minute, e_r drops around 10% of its initial error and stays within that value in low noise ($R=(0.2^{\circ})^2$). However, the rectangular coordinate case takes five minutes and the MP takes more than twenty minutes. Even though the system observability is high in the MP coordinates, range error shows quite unstable behavior. This trend takes longer as the noise level becomes higher (Table 9). Analysis shows that the instability is due to the to the instability in P_t.

Table 11 shows the effect of the data sampling interval for the mixed coordinates. From a standard 10-second interval, it is extended to 20 seconds or is shortened to 2 seconds. More frequent measurement (shorter T) makes the system more ovbservable. Specifically, observability of the speed variable in the maneuvering direction (y-direction here) improved significantly.

However, due to data processing speed limitation of the on-board processor as well as other limitations, ping interval for active SONAR or random process correlation time, for example, one cannot practically decrease the sampling interval arbitrarily in the underwater tracking system.

One more point which has not appeared here is the effect of the magnitude of the maneuvering. Sensitivity analysis shows that once maneuvering exists its magnitude does not give any significant influence to the information content. This also may be a very valuable finding from the economic and tactical standpoint.

Table 8, Effects of measurement noise R

(Mixed, Q=0.0 , $a_y \neq 0.0$)

t (min)	$R = (0.2^o)^2$						$R = (1.0^o)^2$						$R = (6.0^o)^2$					
	Tot. obs.	β	r	v_x	v_y	e_r	Tot	β	r	v_x	v_y	e_r	Tot	β	r	v_x	v_y	e_r
0.0	0.0	0.0	0.0	0.0	0.0	1000 (m)	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000
1.0	4.1	2.4	2.2	0.0	0.0	120	2.1	1.6	0.2	0.0	0.0	740	1.8	1.8	0.0	0.0	0.0	990
2.5	7.6	5.2	4.4	0.0	0.0	90	4.3	2.5	1.5	0.0	0.0	340	2.9	2.6	0.1	0.0	0.0	900
5.0	10.7	7.4	5.5	0.01	0.01	60	7.0	4.3	3.5	0.0	0.0	130	4.1	2.8	0.6	0.0	0.0	860
10.0	14.6	9.6	5.9	0.1	0.3	30	10.4	6.6	5.2	0.01	0.01	30	6.7	3.7	2.3	0.0	0.0	440
15.0	17.7	9.7	5.9	0.5	1.3	40	11.8	7.6	5.2	0.02	0.1	90	8.1	4.6	3.0	0.0	0.0	480
20.0	19.8	10.2	5.9	1.0	1.6	100	12.6	7.6	5.0	0.14	0.3	100	8.6	5.0	3.2	0.01	0.01	420
25.0	21.5	10.2	5.8	1.2	2.1	50	13.3	7.8	4.5	0.2	0.6	90	9.1	5.0	3.3	0.01	0.03	310
30.0	24.0	10.5	6.2	2.3	3.8	80	14.2	7.9	4.2	0.3	1.0	30	9.6	5.2	3.3	0.01	0.05	210
35.0	25.3	11.3	6.3	2.6	4.4	80	15.0	8.3	4.0	0.4	1.4	70	10.1	5.5	3.3	0.01	0.1	200
40.0	25.9	11.1	6.2	2.7	4.8	40	15.5	8.1	3.9	0.4	1.7	40	10.3	5.6	3.2	0.02	0.1	190

Table 9, Effects of measurement noise R

(MP , Q = 0.0 , $a_y \neq 0.0$)

t	$R = (0.2^o)^2$						$R = (1^o)^2$						$R = (6^o)^2$					
	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r	e_r	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r	e_r	Tot	$\dot{\beta}$	$\dot{r}/r$	β	1/r	e_r
0.0	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000
1.0	11.2	6.6	0.0	7.9	0.0	860	6.4	4.2	0.0	5.0	0.0	480	3.0	1.2	0.0	2.0	0.0	510
2.5	16.8	6.9	1.0	10.1	0.3	4500	10.2	5.3	-0.1	7.8	-0.1	5000	5.7	3.2	-0.1	4.4	-0.1	680
5	23.3	8.0	4.2	11.9	1.1	5000	13.9	5.3	1.4	8.9	0.5	2300	6.5	2.8	-0.7	5.9	-0.4	3900
10	30.4	4.4	3.1	12.8	4.2	470	18.5	1.3	0.2	9.8	1.2	4900	8.6	-0.1	-1.4	7.0	-0.2	35000
15	37.0	8.1	8.6	13.7	7.5	110	24.8	5.0	5.5	10.5	4.4	600	12.4	1.6	2.0	7.0	1.1	2000
20	40.0	8.3	8.3	14.6	8.4	510	27.8	5.3	5.3	11.4	5.2	700	15.5	2.1	2.4	7.9	2.0	12300
25	41.0	6.4	8.1	14.1	8.5	90	28.4	3.2	5.0	10.8	5.4	340	16.5	0.4	1.8	7.6	2.1	8400
30	43.0	5.9	9.1	14.5	9.9	100	30.9	3.2	6.1	11.3	6.8	120	18.0	-0.3	2.8	7.8	3.2	470
35	44.7	10.3	11.6	15.4	10.3	20	32.6	7.2	8.5	12.2	7.2	20	19.5	3.7	5.0	8.8	3.6	320
40	46.0	11.2	12.0	15.2	10.5	20	33.8	8.2	8.9	12.0	7.4	20	20.8	4.7	5.5	8.6	3.8	390

Table 10, Effects of measurement noise R

(Rec., $Q = 0.0$, $a_y \neq 0.0$)

t	$R = (0.2^o)^2$						$R = (1^o)^2$						$R = (6^o)^2$					
	Tot	r_x	v_x	r_y	v_y	e_r	Tot	r_x	v_x	r_y	v_y	e_r	Tot	r_x	v_x	r_y	v_y	e_r
0	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000
1	9.8	8.8	0.01	0.4	0.0	800	6.2	6.2	0.0	0.0	0.0	1000	2.7	2.7	0.0	0.0	0.0	1000
2.5	12.5	8.5	0.04	1.8	0.0	170	7.3	6.7	0.01	0.2	0.0	920	3.6	3.5	0.0	0.0	0.0	1100
5	15.5	7.7	0.05	2.9	0.01	130	9.1	6.6	0.03	1.1	0.02	350	4.3	4.1	0.0	0.06	0.0	1000
10	19.0	6.5	0.1	3.7	0.2	50	12.3	6.2	0.03	2.8	0.02	90	5.6	4.2	0.01	0.5	0.0	700
15	21.6	6.0	0.4	3.9	1.2	40	13.7	5.4	0.05	3.2	0.1	130	6.6	4.2	0.02	1.1	0.01	580
20	23.7	6.0	0.9	4.0	1.5	100	14.6	5.1	0.2	3.2	0.3	100	7.0	4.0	0.04	1.3	0.03	520
25	25.0	5.7	1.1	4.1	2.1	20	15.4	4.7	0.2	3.1	0.6	60	7.5	3.9	0.05	1.5	0.05	400
30	27.5	6.5	2.3	5.5	3.8	90	16.4	4.5	0.3	3.4	1.0	70	7.9	3.7	0.05	1.9	0.1	290
35	28.6	6.5	2.6	6.0	4.3	80	17.2	4.3	0.4	3.7	1.4	100	8.3	3.6	0.05	2.1	0.1	260
40	29.2	6.3	2.7	6.0	4.7	20	17.8	4.0	0.4	3.7	1.7	80	8.5	3.4	0.05	2.3	0.2	250

Table 11, Effects of sampling interval T (Mixed, Q = 0. , $a_y \neq 0.$)

t	T = 20 sec.						T = 10 sec.						T = 2 sec.					
	Tot	β	r	v_x	v_y	e_r	Tot	β	r	v_x	v_y	e_r	Tot	β	r	v_x	v_y	e_r
0	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000	0.0	0.0	0.0	0.0	0.0	1000
2.5	3.0	1.7	1.0	0.0	0.0	200	4.3	2.5	1.5	0.0	0.0	340	6.9	3.9	2.6	0.0	0.0	380
5	5.9	3.6	3.0	0.0	0.0	160	7.0	4.3	3.5	0.0	0.0	130	10.2	5.9	4.6	0.01	0.01	190
10	9.1	6.0	4.9	0.0	0.0	40	10.4	6.6	5.2	0.01	0.01	30	13.6	8.2	5.6	0.04	0.05	10
15	10.5	7.0	5.0	0.01	0.1	110	11.8	7.6	5.2	0.02	0.1	90	15.4	8.8	5.4	0.1	0.4	0
20	11.3	7.1	4.8	0.1	0.2	130	12.6	7.6	5.0	0.1	0.3	100	16.9	8.8	5.3	0.4	0.8	10
25	12.0	7.3	4.4	0.1	0.4	110	13.3	7.8	4.5	0.2	0.6	90	18.2	9.0	4.9	0.5	1.3	20
30	12.8	7.4	4.1	0.2	0.7	50	14.2	7.9	4.2	0.3	1.0	30	19.9	9.2	4.9	1.0	2.3	0
35	13.5	7.7	3.8	0.2	1.0	80	15.0	8.3	4.0	0.4	1.4	70	21.0	9.8	4.9	1.2	2.8	20
40	14.0	7.6	3.7	0.2	1.2	60	15.5	8.1	3.9	0.4	1.7	40	21.6	9.5	4.8	1.3	3.1	10

$R = (1^o)^2$

4-3. Information and Sensor number, Measurement Policy in Array SONAR Tracking

Another application area where system observability is crucially important in the ocean environment is the underwater SONAR tracking problem. Here, one is interested in determination of the number of sensors and their deployment configuration such that the system is deterministically observable as well as stochastically more strongly observable. One also wants to decide what kind of quantity should be measured to maximize the collected information with the given conditions. The last point is more important for our purpose here since even with the same number of sensors and with the same deployment structure, measurement of different quantities results with different degrees of observability.

We have already analyzed the same problem from the determinisitic point of view in Chapter Two. We observed that the system is observable except when we measured one absolute time delay with one sensor (1S1abs.D). The other cases are all observable at least in a wide sense. See Figure 3 for the sensor-target configuration. We observed, also, that Doppler measurement increases the measurement quantity with a factor of f_c (carrier frequency) compared to the delay measurement.

Here the same problem is analyzed stochastically. Seven measurement policies are chosen as in Chapter Two for the linearily deployed sensors. The standard extended Kalman filter of the discrete type is used [36].

The other parameters used are as follows:

measurement sampling interval; $T = 15$ sec.

initial condition of x (when no initial noise is added)

$$\hat{x}(0) = \begin{bmatrix} \hat{r}_x(0) \\ \hat{v}_x(0) \\ \hat{r}_y(0) \\ \hat{v}_y(0) \\ \hat{c}_1(0) \\ \hat{c}_2(0) \end{bmatrix} = \begin{bmatrix} 10000 \text{ m} \\ -15.433 \text{ m/s (~30 knots, approaching)} \\ 4000 \text{ m} \\ 0 \text{ m/s} \\ 1500 \text{ m/s} \\ 1500 \text{ m/s} \end{bmatrix},$$

where $x_i(0)$ is assumed $N(\hat{x}_i(0), \sigma_i)$, $i = 1, \ldots 6$, such that

$$\begin{aligned}
\sigma_1 &= \sigma_x = 100 \text{ m}, \\
\sigma_2 &= \sigma_{Vx} = 0.15 \text{ m/s}, \\
\sigma_3 &= \sigma_y = 40 \text{ m}, \\
\sigma_4 &= \sigma_{Vy} = 0.1 \text{ m/s}, \\
\sigma_5 &= \sigma_{C1} = 5 \text{ m/s}, \\
\sigma_6 &= \sigma_{C2} = 5 \text{ m/s}.
\end{aligned}$$

The measurement noise assumed is also a Gaussian sequence with covariance

$$\begin{aligned}
\sigma_{\tau 12} &= 0.019 \text{ sec}, \\
\sigma_{\tau 23} &= 0.026 \text{ sec}, \\
\sigma_{\tau 13} &= 0.016 \text{ sec}, \\
\sigma_{absD} &= 0.359 \text{ sec}, \\
\sigma_{f12} &= 0.1875 \text{ Hz},
\end{aligned}$$

and $P_o = \Gamma_o$ is assumed to be

$$P_o = \begin{bmatrix} \sigma_x^2(0) \quad x \quad 10^4 \\ \sigma_{Vx}^2(0) \quad x \quad 5x10^2 \\ \sigma_y^2(0) \quad x \quad 10^4 \\ \sigma_{Vy}^2(0) \quad x \quad 5x10^2 \\ \sigma_{C1}(0) \quad x \quad 10^2 \\ \sigma_{C2}(0) \quad x \quad 10^2 \end{bmatrix} .$$

f_c = 3500 Hz (modulation carrier frequency) ,

Z_1 = 2000 m (intersensor distance of s_1 and s_3),

Z_2 = 2000 m (intersensor distance of s_1, and s_2).

With the above parameter 20 runs are averaged. Table 12 shows the mutual information content for the whole system for various measurement schemes. Clearly an increased number of deployed sensors yields stronger observability. In spite of the largest observation magnitude (notice that absolute delay is much larger than relative delay magnitude for far-field observation) 1S1abs.D system shows the weakest observability due to the unobservable state x_5 ($=c_1$).

Inspection of the table shows also that the degree of the observability can be approximately categorized in three groups.

1. 1S1abs.D (Obs. = 9.2)
   2S1D (10.5)
   2S1P (13.4)

2.	2S1D1P	(20.6)
	3S2D	(19.7)
	3S3D	(21.2)
3.	3S2D1P	(30.8)

When only delay or Doppler is measured for one or two sensors, the system still remains in a weakly observable status even when the system is deterministically observable (the first group).

Stronger information is obtained when measuring more than one quantity, i.e., both delay and Doppler with two sensors (2S1D1P), or when one more sensor is added to the measurement of only one quantity (3S21D, 3S3D) (the second group).

Information does not increase, appreciably, with the addition of the same kind of measurement quantity as can be seen. This may be caused by the fact that the third delay depends entirel on the first two delays. Only two delays are independent in the three-sensor delay measurement.

Stronger and more significant information is obtained when one observes both delay and Doppler with three sensors (the last case).

It is also of interest that most of the information is collected during the very early stages of the observation, i.e., when the first few sets of measurement data are processed.

Information content for the individual measurement policies is shown in Table 13 through 19. In the case of 1S1abs.D (Table 13) mutual information about c_1 is zero due to the unobservability of this variable. Observability of v_x, v_y and c_2 is relatively poor compared to the range variables r_x and r_y.

Here one can easily understand the obvious advantage of the mutual information approach (in Shannon's sense) compared to the Fisher information matrix approach. In the current method, the information content of the deterministically observable individual state estimate is calculated as well as the total system information even if some states are unobservable. This is not possible in the Fisher information approach when the information matrix is singular (Compare Table 12 and Table 20).

Table 12, System observability of array SONAR

t (min.) \ meas.	1S 1abs.D	2S1D	2S1P	2S1D1P	3S2D	3S3D	3S2D1P
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.5	5.0	6.1	7.2	13.5	13.4	14.0	21.7
1.0	5.7	7.6	8.6	15.5	16.3	17.9	24.7
1.5	6.1	8.2	9.5	16.9	17.3	18.8	26.1
2.0	7.0	8.6	10.1	17.8	17.9	19.4	27.1
2.5	8.1	8.9	10.6	18.4	18.2	19.7	27.8
3.0	8.6	9.2	11.3	18.9	18.4	20.0	28.5
3.5	8.8	9.7	11.7	19.4	18.8	20.3	29.1
4.0	8.9	9.9	12.4	19.8	19.1	20.6	29.7
4.5	9.1	10.2	12.9	20.2	19.4	20.8	30.2
5.0	9.2	10.5	13.4	20.6	19.7	21.2	30.8

Table 13, Observability : 1S1abs.D

t (min)	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	4.35	3.11	0.00	0.03	0.00	0.0	0.01
0.50	5.03	3.27	0.00	0.03	0.00	0.0	0.01
0.75	5.42	3.32	0.00	0.03	0.00	0.0	0.01
1.00	5.70	3.33	0.01	0.04	0.00	0.0	0.01
1.25	5.92	3.34	0.01	0.06	0.00	0.0	0.01
1.50	6.14	3.35	0.02	0.13	0.00	0.0	0.01
1.75	6.48	3.44	0.03	0.36	0.00	0.0	0.01
2.00	6.99	3.65	0.04	0.78	0.01	0.0	0.02
2.25	7.57	3.94	0.04	1.27	0.01	0.0	0.02
2.50	8.12	4.19	0.04	1.73	0.01	0.0	0.03
2.75	8.52	4.33	0.04	2.04	0.01	0.0	0.03
3.00	8.64	4.29	0.05	2.10	0.01	0.0	0.03
3.25	8.72	4.26	0.06	2.13	0.02	0.0	0.03
3.50	8.80	4.22	0.08	2.16	0.02	0.0	0.04
3.75	8.87	4.19	0.10	2.19	0.03	0.0	0.04
4.00	8.94	4.15	0.12	2.21	0.04	0.0	0.04
4.25	9.00	4.11	0.13	2.23	0.05	0.0	0.04
4.50	9.06	4.07	0.15	2.26	0.05	0.0	0.05
4.75	9.12	4.03	0.16	2.28	0.06	0.0	0.05
5.00	9.17	3.99	0.16	2.31	0.08	0.0	0.05

Table 14, Observability : 2S1D

t	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	5.67	0.19	0.00	0.34	0.00	0.29	0.33
0.50	6.07	0.23	0.00	0.36	0.01	0.30	0.33
0.75	6.62	0.60	0.00	0.37	0.02	0.35	0.39
1.00	7.58	1.07	0.01	0.39	0.03	0.40	0.47
1.25	7.99	1.26	0.01	0.46	0.04	0.43	0.50
1.50	8.15	1.33	0.01	0.50	0.06	0.44	0.51
1.75	8.34	1.45	0.02	0.52	0.07	0.44	0.51
2.00	8.63	1.54	0.02	0.57	0.09	0.46	0.54
2.25	8.83	1.66	0.03	0.66	0.44	0.47	0.55
2.50	8.95	1.69	0.04	0.72	0.13	0.47	0.55
2.75	9.11	1.76	0.04	0.81	0.15	0.47	0.55
3.00	9.22	1.82	0.05	0.86	0.17	0.47	0.55
3.25	9.47	1.91	0.07	1.06	0.19	0.48	0.57
3.50	9.66	1.97	0.08	1.20	0.21	0.49	0.57
3.75	9.86	2.06	0.09	1.32	0.24	0.49	0.58
4.00	9.96	2.11	0.11	1.34	0.26	0.49	0.58
4.25	10.05	2.15	0.13	1.36	0.28	0.49	0.58
4.50	10.15	2.16	0.14	1.38	0.29	0.50	0.58
4.75	10.27	2.19	0.16	1.41	0.31	0.49	0.57
5.00	10.48	2.23	0.18	1.50	0.33	0.50	0.58

Table 15, Observability : 2S1P

t	Total	r_x	v_x	r_Y	v_Y	c_1	c_2
0.25	6.65	0.59	0.03	0.15	0.00	0.23	0.20
0.50	7.22	0.84	0.05	00.16	0.00	0.26	0.25
0.75	7.94	1.28	0.11	0.18	0.01	0.39	0.34
1.00	8.57	1.67	0.15	0.21	0.01	0.45	0.37
1.25	9.12	1.82	0.20	0.27	0.02	0.49	0.40
1.50	9.47	1.98	0.24	0.34	0.03	0.50	0.41
1.75	9.81	2.13	0.27	0.47	0.03	0.50	0.41
2.00	10.09	2.24	0.31	0.64	0.04	0.53	0.43
2.25	10.32	2.34	0.36	0.81	0.05	0.53	0.43
2.50	10.56	2.46	0.40	0.90	0.05	0.55	0.44
2.75	10.81	2.56	0.44	0.97	0.06	0.56	0.45
3.00	11.25	2.70	0.48	1.18	0.07	0.56	0.46
3.25	11.45	2.79	0.51	1.29	0.07	0.56	0.46
3.50	11.69	2.90	0.54	1.36	0.08	0.56	0.46
3.75	12.12	3.02	0.59	1.48	0.09	0.56	0.46
4.00	12.36	3.15	0.63	1.55	0.10	0.56	0.46
4.25	12.63	3.25	0.67	1.63	0.11	0.56	0.46
4.50	12.89	3.32	0.72	1.70	0.13	0.56	0.46
4.75	13.11	3.39	0.78	1.73	0.14	0.56	0.46
5.00	13.44	3.48	0.84	1.31	0.15	0.56	0.46

Table 16, Observability : 2S1D1P

t	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	12.31	0.93	0.03	0.55	0.00	0.53	0.65
0.50	13.52	1.19	0.07	0.79	0.02	0.64	0.65
0.75	14.57	1.47	0.21	1.12	0.06	0.65	0.67
1.00	15.53	1.78	0.32	1.47	0.14	0.66	0.68
1.25	16.40	2.01	0.46	1.71	0.25	0.66	0.69
1.50	16.91	2.13	0.52	1.82	1.82	0.36	0.70
1.75	17.34	2.25	0.58	1.88	0.46	0.66	0.71
2.00	17.75	2.37	0.65	1.95	0.53	0.67	0.71
2.25	18.15	2.48	0.73	2.00	0.57	0.67	0.72
2.50	18.41	2.55	0.75	2.00	0.69	0.67	0.72
2.75	18.65	2.62	0.79	1.99	0.62	0.67	0.72
3.00	18.90	2.70	0.82	2.00	0.65	0.67	0.72
3.25	19.20	2.81	0.83	2.07	0.73	0.67	0.73
3.50	19.40	2.90	0.83	2.08	0.77	0.67	0.72
3.75	19.58	2.98	0.84	2.09	9.80	0.66	0.72
4.00	19.70	3.09	0.65	2.12	0.84	0.66	0.72
4.25	19.99	3.20	0.86	2.15	0.89	0.66	0.72
4.50	20.18	3.30	0.87	2.18	0.93	0.66	0.72
4.75	20.39	3.42	0.90	2.22	1.00	0.66	0.72
5.00	20.64	3.53	0.94	2.27	1.09	0.66	0.72

Table 17, Observability : 3S2D

t	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	12.48	0.21	0.00	1.36	0.00	1.48	0.35
0.50	13.40	0.37	0.00	1.43	0.04	1.50	0.45
0.75	15.37	1.12	0.01	1.74	0.11	1.76	0.97
1.00	16.28	1.64	0.01	2.07	0.18	1.95	1.31
1.25	16.95	1.92	0.02	2.31	0.29	2.15	1.54
1.50	17.34	2.08	0.03	2.46	0.41	2.28	1.86
1.75	17.57	2.18	0.04	2.54	0.52	2.30	1.73
2.00	17.92	2.27	0.05	2.63	0.64	2.34	1.79
2.25	18.04	2.30	0.06	2.66	0.73	2.33	1.79
2.50	18.18	2.32	0.07	2.68	0.82	2.30	1.78
2.75	18.30	2.35	0.08	2.73	0.90	2.26	1.77
3.00	18.43	2.37	0.09	2.77	0.98	2.24	1.76
3.25	18.58	2.42	0.10	2.83	1.06	2.20	1.76
3.50	18.75	8.48	0.12	2.89	1.13	2.17	1.76
3.75	18.96	2.55	0.14	2.96	1.18	2.15	1.77
4.00	19.11	2.60	0.16	3.01	1.24	2.12	1.77
4.25	19.28	2.64	0.18	3.06	1.28	20.6	1.77
4.50	19.43	2.67	0.20	3.09	1.33	2.05	1.77
4.75	19.56	2.71	0.22	3.13	1.37	2.01	1.76
5.00	19.74	2.78	0.24	3.19	1.41	1.98	1.75

Table 18, Observability : 3S3D

t	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	13.10	0.21	0.00	0.36	0.00	1.48	0.35
0.50	14.00	0.40	0.00	1.44	0.06	1.51	0.46
0.75	17.02	1.58	0.01	1.97	0.12	1.91	1.28
1.00	17.86	1.96	0.01	2.25	0.20	2.12	1.54
1.25	18.43	2.26	0.03	2.50	0.35	2.30	1.81
1.50	18.79	2.39	0.04	2.65	0.47	2.46	1.92
1.75	18.94	2.42	0.05	2.69	0.58	2.43	1.93
2.00	19.39	2.72	0.06	2.94	0.68	2.47	2.06
2.25	19.60	2.74	0.07	2.97	0.77	2.46	2.05
2.50	19.74	2.75	0.08	3.00	0.87	2.43	2.03
2.75	19.86	2.76	0.09	3.02	0.96	2.39	2.01
3.00	19.99	2.78	0.11	3.06	1.04	2.35	2.00
3.25	20.14	2.80	0.13	3.09	1.12	2.30	1.98
3.50	20.25	2.81	0.14	3.11	1.19	2.26	1.96
3.75	20.43	2.85	0.17	3.15	1.24	2.21	1.95
4.00	20.59	2.90	0.19	3.21	1.29	2.19	1.94
4.25	20.77	3.01	0.21	3.29	1.34	2.16	1.93
4.50	20.93	3.04	0.24	3.32	1.39	2.15	1.93
4.75	21.06	3.06	0.26	3.35	1.43	2.11	1.92
5.00	21.20	3.13	0.29	3.41	1.48	2.10	1.92

Table 19, Observability : 3S2D1P

t	Total	r_x	v_x	r_y	v_y	c_1	c_2
0.25	21.74	1.73	0.06	1.46	0.00	1.48	0.84
0.50	22.86	1.93	0.12	1.64	0.18	1.61	0.94
0.75	23.82	2.13	0.19	1.87	0.39	1.81	1.16
1.00	24.70	2.41	0.25	2.21	0.59	2.05	1.38
1.25	25.50	2.63	0.35	2.47	0.77	2.25	1.64
1.50	26.09	2.78	0.44	2.62	0.94	2.35	1.78
1.75	26.58	2.94	0.50	2.77	1.10	2.42	1.84
2.00	27.07	3.08	0.59	2.89	1.23	2.46	1.98
2.25	27.43	3.19	0.67	2.97	1.34	2.46	1.98
2.50	27.84	3.29	0.74	3.06	1.44	2.47	2.00
2.75	28.15	3.39	0.81	3.14	1.54	2.45	2.01
3.00	28.49	3.51	0.85	3.24	1.63	2.45	2.02
3.25	28.81	3.64	0.87	3.36	1.72	2.47	2.03
3.50	29.10	3.78	0.83	3.49	1.81	2.49	2.04
3.75	29.38	3.9	0.90	3.60	1.88	2.49	2.06
4.00	29.68	4.05	0.92	3.74	1.95	2.50	2.09
4.25	29.96	4.19	0.95	3.88	2.01	2.51	2.11
4.50	30.22	4.32	0.98	3.98	2.07	2.48	2.12
4.75	30.49	4.46	1.03	4.09	2.13	2.45	2.14
5.00	30.76	4.60	1.10	4.20	2.20	2.42	2.15

some unobservable states. In this case the total or individual state information cannot be computed because of this singularity. In that case even identification of unobservable states is not generally possible. To identify those unobservable states usually intuition, experience or trial and error are used. All the other six measurement policies (Tables 14 to 19) show that the system is observable even though the degree of observability is different. As the number of sensors increases unobservable or weakly observable states become more strongly observable. Specifically, information growth for the sound speed variables c_1 and c_2 is significant when the three-sensor policy is used regardless of measured quantity.

Strong system observability for the 3S2D1P case is due to the strong individual state information for all six states.

The effects of filtering errors due to the different degrees of information content is seen from Figure 14 through 16 for range r_x, target speed v_y, and sound speed c_2, respectively.

Roughly, increasing the number of measured quantities with more sensors gives a smaller filtering error because of the stronger observability. With an initially given 1,000 m range error, combine the measurement of delay and Doppler yields significantly small errors. The errors stay within few ten meters in 5 minutes final time for both 2S1D1P and 3S2D1P cases. 3S2D1P case, particularly, shows very desirable characteristics as can be seen from Figure 14. It is important to note here that very undesirable properties (in the sense that large error or oscillation of range error results) appear when measuring only time delay. The same figure, also, shows large (more than initial error) errors in the case of 2S1D, 3S3D, and some overshoot appears for 1S1D even with reasonably good range information. Notice that 1S1D has only limited usage, e.g., target-sensor synchronization or in case of active SONAR situation.

One now can say that Doppler measurement which is combined with proper delay measurement is crucial for good range estimation in SONAR tracking.

Figure 15 shows target velocity error with an initial 2 m/s (~4 knots) error. Here one can observe some different aspects as compared with the range error. I.e., no matter what quantity is measured, the system exerts less velocity error when more sensors are included with increased number of measured variables. Figure 15, also, shows that the magnitude of this error can be divided in three groups,exactly, as the total system observability is divided. The first group (1S1D, 2S1JD, 2S1P) again shows the poorest performance and the third group (3S2D1P) is the superior group.

1S1D shows some oscillatory properties here, also. Extended observation beyond five minutes showed that the error in 2S1D1P case decreases from around 5 1/2 minutes.

Figure 16 shows the evolution of the sound speed error for an initially given 50 m/s. This value may be slightly larger than the practical situation.

- However, one can easily recognize three distinct groups of error trends. These three groups exactly coincide with the groups which are made in the system observability. I.e., 1S1D, 2S1D, 2S1P group shows the poorest performance and 2S1D1P, 3S2D, 3S3D group shows the medium range error and, again, 3S2D1P shows the superior performance. 1S1D case shows a mild overshoot with the weakest information.

For comparison, a discrete version of the Fisher information matrix (3-65) is computed for the selected five observation policies.

$$I(k,1) = \sum_{i=1}^{k} \Phi^{-T}(k,i)H^{T}(\hat{x}(i))R_{i}^{-1} H(\hat{x}(i)) \Phi^{-1} (k,i), \tag{4-16}$$

Here, iterative modification of (3-67),

$$\begin{aligned} I(k+1,1) = {} & \Phi^{-T} (k+1,k) I(k,1) \Phi^{-1} (k+1,k) + \\ & H^{T}(\hat{x}(k+1)) R^{-1}(k+1) H(\hat{x}(k+1)), \end{aligned} \tag{4-17}$$

is used instead of (4-16). This is shown at Table 20. Matrix I(k,1) remains singular over the entire observation period for the 1S1D measurement case and remains nonsingular with shown magnitude of determinant in other cases.

Comparison of this table with the total information contents (Table 12) will reveal that the two approaches exactly correspond to each other for the chosen five measured policies.

Superiority of the measurement 3S2D1P system is shown here, also, Thus one can conclude this section as follows: at least two sensors are required for the system to be observable. 3S2D1P measurement gives the most desirable performance in all cases. If only two sensors are available, a combination of delay and Doppler (2S1D1P) measurement is strongly recommended. For small range error, Doppler measurement is crucial. For small target velocity and sound speed errors, include as many sensors as possible to make strong system observability.

1S1D policy is not recommended except in special cases as in the experimentally well synchronized case [64] or in an active SONAR system.

Fig. 14 Range Error

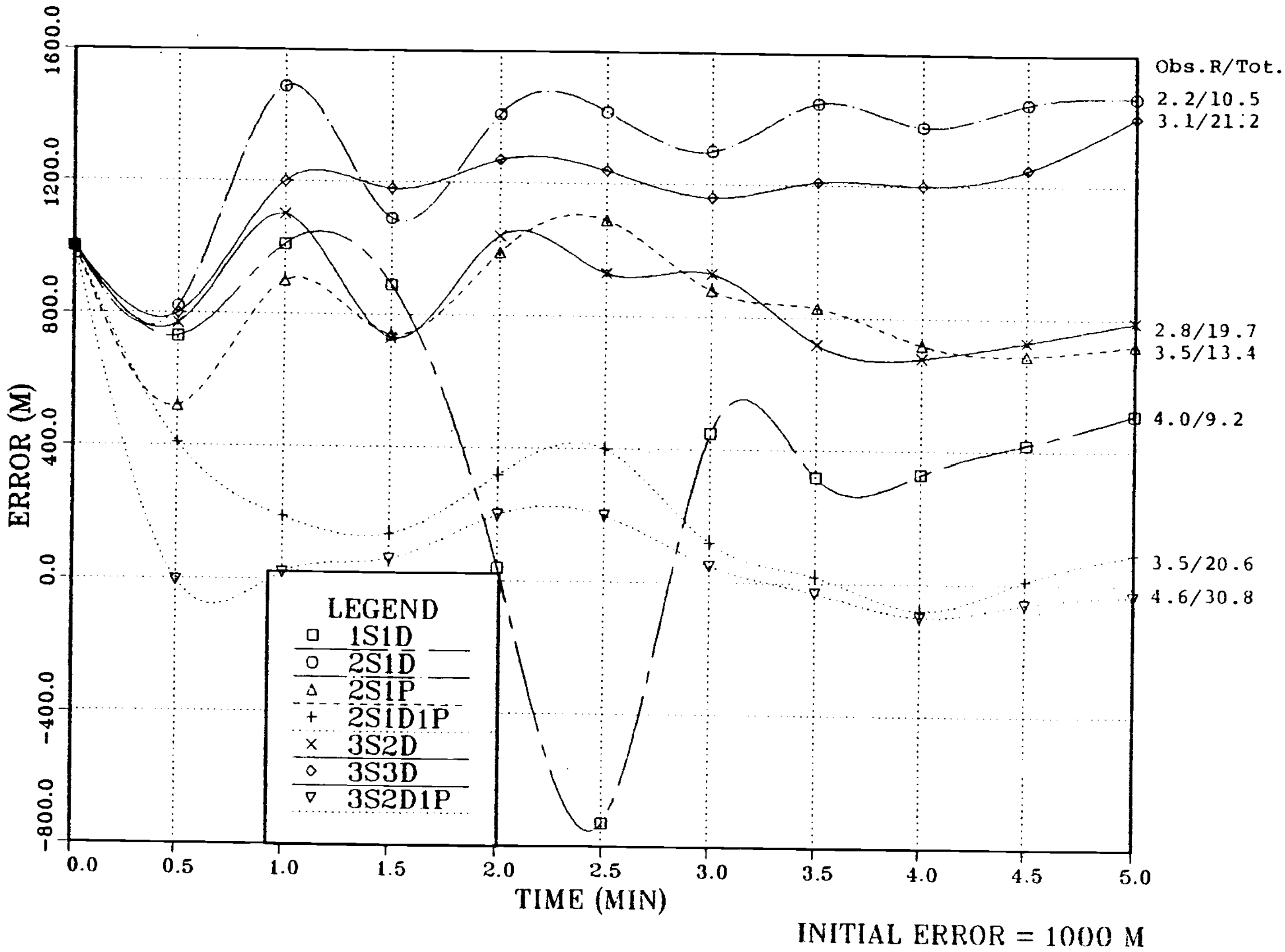


Fig. 15 Velocity Error

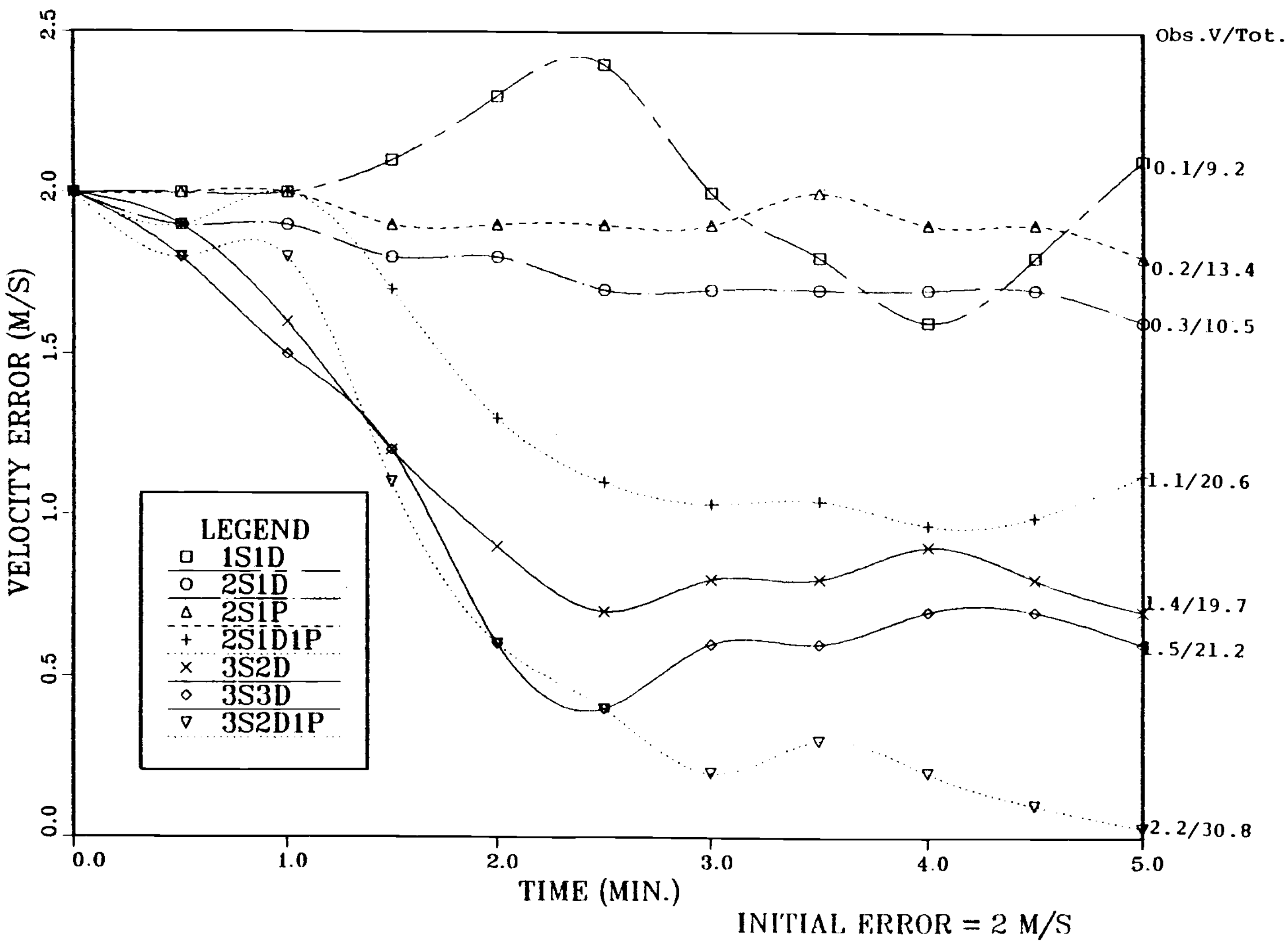
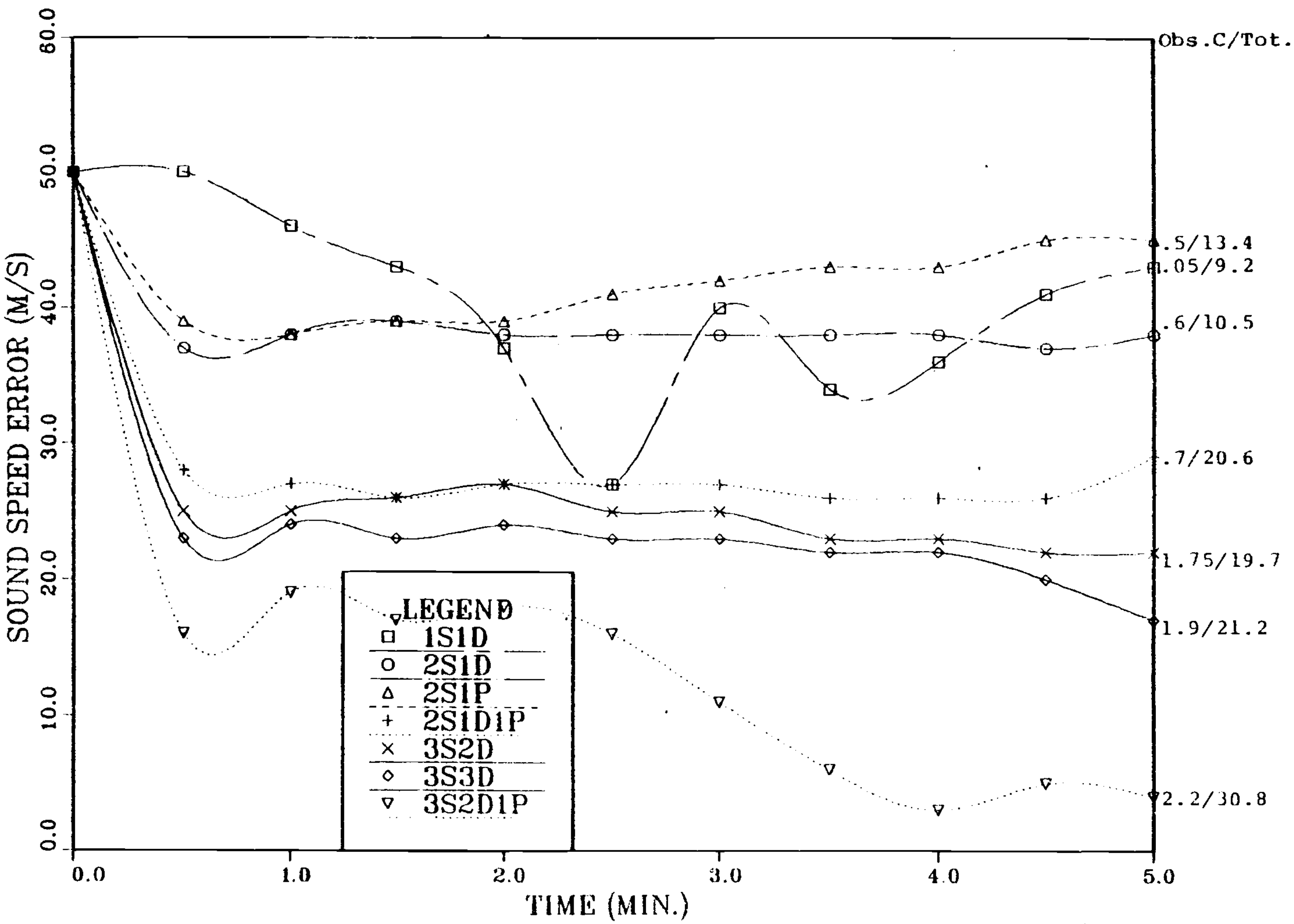


Fig. 16 Sound Speed Error



INITIAL ERROR = 50 M/S

Table 20,

Observability (singularity of Fisher information matrix)

meas. / t (min.)	1S 1abs.D	2S1D1P	3S2D	3S3D	3S2D1P
0.0	↑	↕	↕	↕	↕
0.5		4.9×10^{-21}	8.1×10^{-22}	3.8×10^{-21}	5.3×10^{-17}
1.0					
1.5					
2.0					
2.5		4.0×10^{-15}	2.6×10^{-11}	2.0×10^{-10}	1.1×10^{-9}
3.0					
3.5					
4.0					
4.5					
5.0	↓	3.7×10^{-9}	7.6×10^{-8}	6.3×10^{-7}	1.2×10^{-5}

* ↔ implies unobservable period

CHAPTER 5: SUMMARY AND CONCLUSION

In this dissertation, system observability is studied for both deterministic and stochastic cases.

Since nonlinear observability for deterministic systems is a geometric nonlinear functional property, the inverse and implicit function theorems are useful. By modifying the global implicit-function theorem, sufficient conditions for the given nonlinear function to be globally homeomorphic are derived. From an applicational point of view, the nonzero Jacobian condition, which can be related to n-1 dimensions for the special case, provides the connectedness condition for every state to be connected to the measurement space. However, a finite-covering condition must be tightened to a one-covering condition then by which univalence of the connectedness can be guaranteed.

Before these two conditions can be applied to the system equations, differentiation of the system observation equations with respect to t, and substitution of the lower-order derivatives of observation equations to the higher order up to (n-1)-th derivatives must be preceded.

Depending on the satisfication of the conditions, observability in the strict sense, observability in the wide sense, and unobservable states are determined.

Application of this method is demonstrated by several examples. Especially, two practical problems in the ocean environment where the observability is very important are dealt with here.

Bearing-only-target tracking problem, which is described by, a so-called, mixed-coordinate system, is analyzed. The fact that this system is only observable when relative maneuvering exists and unobservable when non-maneuvering is proven again, and special cases of interest are studied. In the linear-array SONAR problem, at least two sensors are necessary for system observability. Doppler measurement scales up the delay measurement quantity by the factor of modulation carrier frequency.

For stochastic system observability, a new approach is attempted. Instead of using the classical Fisher information matrix, mutual information (in the Shannon sense) is computed and utilized as a criterion for determination of the degree of the observability. Computed here is the amount of information about one random process (state x_t) contained in another random process (observation y_t). Since the mutual information is defined as the uncertainty or entropy difference between the sender and the receiver of the information, from information theory, it is required to know two entropies $H(x_t)$ and $H(x_t|y_t)$.

Fortunately, entropy and variance have one-to-one relationship except in a few special cases. So, mutual information can be computed from two covariances - a priori and a posteriori statistical covariances - as far as both are available. Since, in practice, higher moments are required in the evaluation of the second moment of the density approximation must be used to obtain it. Any well-developed approximated nonlinear filter algorithm can be used. Since, covariance in this case does not fully characterize the statistics of the state x_t mutual information $I(\hat{x}_t, y_t)$ is used instead of $I(x_t, y_t)$. The relationship between the deterministic observability rank condition and the stochastic observability in terms of mutual information is discussed for the linear system.

Obvious advantages of the mutual information approach over the Fisher information approach in connection with practical application aspects are:

1) System observability computation is possible even in the case that some states are unobservable. This is not possible in the Fisher information due to the singularity of the observability matrix.

2) Identification of unobservable states is immediate by just indicating the states whose information do not grow. But this is very difficult in the Fisher information where they can only be done by empirical guessing or trial and error [71].

3) Both mutual information and Fisher information consider both system and measurement noise effects, theoretically. But the Fisher information matrix in the applicational form only accommodates measurement noise.

4) The Fisher information matrix for the nonlinear system, traditionally, uses the first-order linearization. But Shannon

information seems more readily calculated according to higher-order state estimation approximation.

The relationship between the Fisher information and the Shannon mutual information is discussed for a special form of nonlinear system.

According to the result of the mutual information derivation, it is computed and compared for the three practical examples. Simple linear example, falling body, is followed by the two nonlinear simulation results which are the same system model used in the deterministic observability.

In the BOT system, three coordinates; rectangular, modified polar and mixed coordinate system are compared.

Information structure analysis shows that both range and target speed are weakly observable when the observer does not maneuver relative to the target for all three coordinates. Once meneuvering exists in any direction its magnitude has not much effect on observability. It is observed that system dynamic noise reduces collected information significantly.

Measurement noise and data sampling intervals.also have certain effects on observability. Their effects are analyzed.

Analysis always shows that poor system observability is followed by large filtering error and vice versa. In spite of no specific superiority in its state observability, mixed coordinates show the most desireable performance in all cases.

Modified polar coordinates show some unstable characteristics in spite of its strong observability magnitude.

At least two sensors are required for every state to be observable in the array SONAR tracking problem.

3S2D1P measurement policy is the most recommendable if up to three sensor deployment is available.

If only two sensors are available, a combination of delay and Doppler (2S1D1P) measurement is the most recommendable policy. For only small range, Doppler measurement is crucially important. On the other hand, for small target velocity and sound-speed errors, include as many sensors as possible to make the system more strongly observable since those errors are porportional to the whole system observability.

1S1D is not recommendable except for particulary well-synchronized experimental cases.

As a result, for the deterministic observability problem, two simple and convenient conditions - connectedness and univalence - are developed.

For stochastic observability, it is found that the mutual information approach is a valid alternative which seemingly can determine the degree of observability more completely than the classical Fisher information matrix.

The effect of the deterministic observability to the stochastic observability and related topics are analyzed for the BOT and array SONAR tracking simulation.

REFERENCES

1. Palais, R. S., Natural Operation of Differential Forms, Trans. Amer. Math. Soc., Volume 92, pp 125 - 141, July 1959.

2. Athans, M., Falb, P.L., Optimal Control, McGraw-Hill Inc., N.Y., 1966.

3. Chen, C. T., Introduction to Linear System Theory, Holt, Rinehart and Winston, N.Y., 1980.

4. Lee, E.B., Marcus, L., Foundations of Optimal Control Theory, John Wiley & Sons, N.Y., 1967.

5. Hwang, M., Seinfeld, J. H., Observability of Nonlinear Systems, J. of Opt. Theory and Application, Volume 10, No. 2, pp 67-77, 1972.

6. Hermann, R., Krener, A.J., Nonlinear Controllability and Observability, IEEE Tr. Auto Con. Volume AC-22, No. 5, pp 728-740, Oct. 1977.

7. Nijeijer, H., Observability of Autonomous Discrete Time Nonlinear Systems. A Geometric Approach, Int. J. Control, Volume 36, No. 5, pp 867-874, 1982.

8. Fliess, M., A Remark on Nonlinear Observability, IEEE Tr. Auto. Con., Volume AC-27, No. 2, pp 489-490, April 1982.

9. Schoenwandt, S., On Observability of Nonlinear Systems, 2nd IFAC Symposium, Prague, Czechoslavakia, June 1970.

10. Kostyukovskii, Y. M., Observability of Nonlinear Controlled System, Automation Remote Control, Volume 9, pp 1384-1396, 1968.

11. Kostyukovskii, Y.M., Simple Conditions of Observability of Nonlinear Controlled System, Automation Remote Control, Volume 10, pp 1575-1584, 1968.

12. Griffith, E.W., Kunmar, K.S.P., On the Observability of Nonlinear Systems: I, J. of Math. Ana. and Appl., Volume 35, pp 135 - 147, 1971.

13. Kou, S. R., Elliott, D.L., Tarn.T.J., Observability of Nonlinear Systems, Info. and Control, Volume 22, pp 89-99, 1973.

14. Singh, S. N., Observability in Nonlinear Systems with Unmeasurable Inputs, Int. J. Sys. and Science, Volume 6, No. 8, pp 723-732, 1975.

15. Fujisawa, T., Kuh, E.S., Some Results on Existence and Uniqueness of Solution of Nonlinear Networks, IEEE Tr. on Circuit Th., Volume CT-18, No. 5, pp 501 - 506, 1971.

16. Fitts, J. M., On the Observability of Nonlinear Systems with Applications to Nonlinear Regression Analysis, Info. Science, Volume 4, pp 129-156, 1972.

17. Galperin, Ye. A., On the Observabiity of Nonlinear Systems, Eng. Cybernetics, Volume 1, pp 338-345, 1972.

18. Sen, P., Chidambara, M.R., Observability of a Class of Nonlinear Systems, IEEE Tr. Auto. Con., Volume AC-25, No. 6, pp 1236-1237, 1980.

19. Kuh, S., Hajj, I. N., Nonlinear Circuit Theory; Resistive Networks, Proc. of IEEE, Volume 59, No. 3, pp 340 - 355, 1971.

20. Chua, L.O., Lam, Y.F., Global Homeomorphism of Vector-Valued Function, J. of Math Ana. and Appl., Volume 39, pp 600 - 624, 1972.

21. Kaplan, W., Advanced Mathematics for Engineers, Addison-Wesley Pub. Company, Calif., 1981.

22. Gale, G., Nikaido, H., The Jacobian Matrix and Global Univalence of Mapping, Math. Annalen, Volume 159, pp 81-93, 1965.

23. Kalman, R.E., On the General Theory of Control Systems, Proc. of the 1st IFAC Congress, London, 1961.

24. Oden, J.T., Applied Functional Analysis, Prentice-Hall, NJ, 1979.

25. Ortega, J. M., Rheinboldt, W.C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, N.Y., 1971.

26. Burkill, J.C., Burkill, H., A Second Course in Mathematical Analysis, Cambridge Univ. Press, 1970.

27. Nikaido, H., Convex Structure and Economic Theory, Academic - Press, N.Y., 1972.

28. Royden, H.L., Real Analysis-2nd Edition, MacMillan Publishing Company, N.Y., 1968.

29. Chevalley, C., Theory of Lie Groups, Princeton University Press, N.J. 1948.

30. Browder, F.E., The Solvability of Nonlinear Functional Equations, Duke Math J., Volume 30, pp 557-566, 1963.

31. Sutherlandd, W.S., Introduction to Metric and Topological Space, Oxford University Press, London, 1981.

32. Duren, D.L., Univalent Functions, Springer-Verlag, N.Y., 1963.

33. Hwang, C.S., Mohler, R.R., Nonlinear Observability and Mixed Coordinate Bearing Only Signal Analysis, Proc. 23rd Conference on Decision and Control, Las Vegas, 1984.

34. Maybeck, P.S., Stochastic Models, Estimation and Control, Vol. 2 Academic Press, N.Y., 1982.

35. Aoki, M., Optimization of Stochastic Systems, Academic Press, N.Y., 1967.

36. Jazwinski, A.H., Stochastic Process and Filtering Theory, Academic Press, 1970.

37. Ham, F.M., Brown, R.G., Observability, Eigenvalues, and Kalman Filtering, IEEE Tr. on Aero. and Elect. Sys., Volume AES-19, No. 2, pp 269-273, 1983.

38. Taylor, J.H., The Cramer-Rao Estimation Error Lower Bound Computation for Deterministic Nonlinear Systems, IEEE Tr. on Auto. Con., Volume AC-24, No. 2, pp 343-344, 1979.

39. Chang, C.B., Two Lower Bounds on the Covariance for Nonlinear Estimation Problems, IEEE Tr. on Auto. Con., Volume AC-26, No. 6, pp 1294-1297, 1981.

40. Hartley, R.V.L., Transmission of Information, Bell Sys. Tech. J., Volume 7, pp538, 1928.

41. Hamming, R.W., Coding and Information Theory, Prentice-Hall, N.J., 1980.

42. Schroeder, M.R., Linear Prediction, Entropy and Signal Analysis, IEEE ASSP Magazine, pp 3-11, 1984.

43. Stam, A.J., Some Inequality Satisfied by the Quantities of Information of Fisher and shannon, Info. and Control, Volume 2, pp 102-112, 1959.

44. Hammard, P., Shannon's Information and Fisher's Information for Diffusion Processes, IFAC Workshop on Info. and Sys., pp 41-49, France, 1977.

45. Duncan, T.E., On the Calculation of Mutual Information, SIAM J. Appl. Math., Volume 19, No. 1, pp 215-220, 1970.

46. Ting-Ho, Lo, J., A General Bayes Rule and its Application to Nonlinear Estimation, Info. Sci., Volume 8, pp 189-198, 1975.

47. Eykhoff, P., System Identification, John-Wiley & Sons, N.Y., 1977.

48. Kalata, P., Priemer, P., Linear Prediction, Filter and Smoothing: An Information-Theortic Approach, Info. Sci., Volume 17, pp 1-14, 1979.

49. Shannon, C.E., Weaver, W., The Mathematical Theory of Communication, The Univ. of Illinois Press, Urbana, 1949.

50. Kullback, S., Information Theory and Statistics, John-Wiley & Sons, N.Y., 1959.

51. Weidemann, H.L., Stear, E.B., Entropy Analysis of Parameter Estimation, Info. and Control, Vol. 14, pp 493-506, 1969.

52. Kalata, P., Priemer, R., On Minimal Error Entropy Stochastic Approximation, Int. J. Sys. Sci., Volume 5, No. 9, pp 895-906, 1974.

53. Ohe, S., Tomita, Y., Omatu, S., Soeda, T., Information-Theoretical Optimal Smoothing Estimator, Info. Sci., Volume 22, pp 201-215, 1980.

54. Omatu, S., Tomita, Y., Soeda, T., An Alternative Expression of the Mutual Information for Gaussian Processes, IEEE Tr. Info. Theory, Vol. IT-22, pp 593-595, 1976.

55. Gel' Fand, I.M., Yaglom, A.M., Calculation of the Amount of Information About a Random Function Contained in Another Such Function, Amer. Math. Soc. Trans, Volume 2, No. 12, pp 199-246, 1959.

56. Liptser, R.S., Shiryayev, A.N., Statistics of Random Processes II, Applications, Springer-Verlag, N.Y., 1978.

57. Wang, E., Stochastic Process in Information and Dynamical Systems McGraw Hill, N.Y., 1971.

58. Lipster, R.S., Shiryayev, A.N., Statistics of Random Processes I, Springer-Verlag, N.Y., 1978.

59. Sunahara, Y, Aihara, S., Shiraiwa, M., The Stochastic Observability for Noisy Nonlinear Systems, Int. J. on Control, volume 22, No. 4, pp 461-480, 1975.

60. Bar-Shalom, Y., Multitarget Tracking-Lecture Notes in Advanced Topics in Modern Control, Naval Postgraduate School, 1984, (to be published by Academic Press).

61. Balakrishnan, A.V., Stochastic Filtering and Control, Optimization Software, Los Angeles, 1983.

62. Bryson, A.E., Ho, Y.C., Applied Optimal Control, John-Wiley & Sons, N.Y., 1975.

63. Mohler, R.R., Hwang, C.S., On Observation Model Analysis for Information, DARPA Undersea Surveillance Symposium, Monterey, Calif., 1985.

64. Helton, R.A., Oceanographic and Acoustic Characteristics of the Dabob Bay Range-Report 1300, Naval Torpedo Station, Keyport, WA., 1976.

65. Hwang, C.S., Mohler, R.R., Undersea Sound Speed and Range Estimation, Report of Dept. of Elect. and Computer Eng., Oregon State Univ., 1983.

66. Van-Tree, H.L., Detection, Estimation and Modulation Theory-Part III, John-Wiley & Sons, N.Y., 1971.

67. Cozzolino, J.M., Zahner, M.J., The Maximum Entropy Distribution of the Future Market Price of the Stock, Operation Research, Volume 21, pp 1200-1211, 1973.

68. Carter, G.C., Variance Bounds for Passively Locating an Acoustic Source with Symmetric Linear Array, J. Acoustic Soc. Amer., Volume 62, pp 922-926, 1977.

69. Alspach, D.L., Mohnkern, G.L., Lobbia, R.N., Sound speed Estimation as a Means of Improving Target Tracking Performance, AD-A086, 603/8, Orincon Co., La Jolla, CA, 1980.

70. Urick, R.J., Principles of Underwater Sound (2nd.Ed.), McGraw Hill, N.Y., 1975.

71. Gallager, R.G., Information Theory and Reliable Communication, John-Wiley & Sons, N.Y., 1968.

Appendices

APPENDIX A: FUNCTIONS AND FUNCTIONAL DEPENDENCE [24],[31]

Definitions

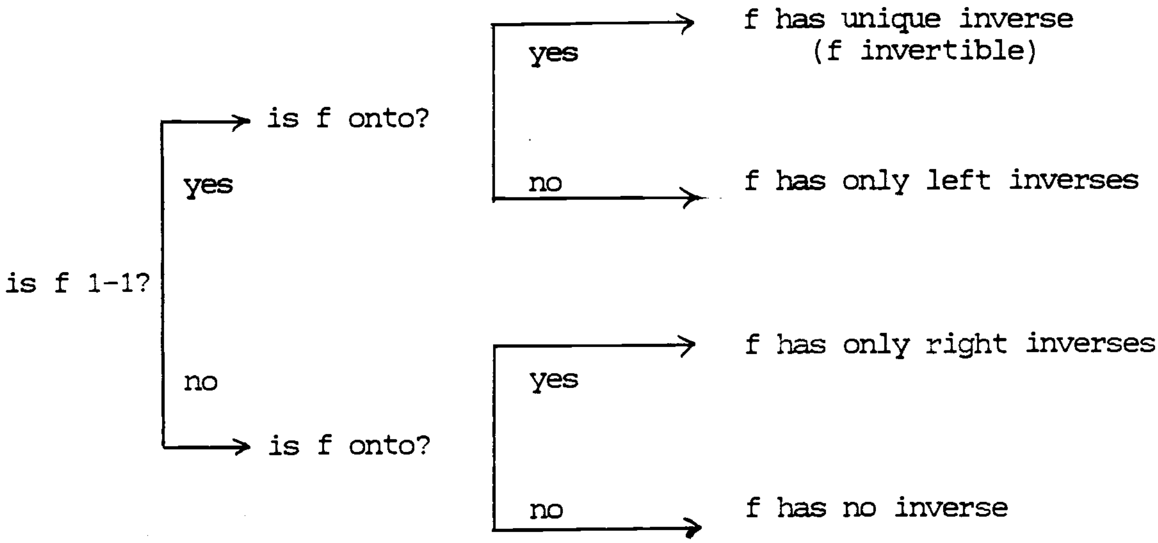
Consider an n real valued continuous function f: $R^n \to R^n$. Function f: $X \to Y$, $X \varepsilon R^n$, $Y \varepsilon R^n$ is one-to-one from X into Y if for every $y \varepsilon R(f)$, range of f, there is exactly one $x \varepsilon X$ such that $y=f(x)$. Then f has a left-inverse g if and only if f is one-to-one, i.e., there exists a function g: $Y \to X$ such that

$$gof = g(f(X)) = I_X, \tag{A-1}$$

where I_X is the identity function for X. If every $y \varepsilon Y$ is the image of at least one $x \varepsilon X$, then f is an onto function. In this case f has a right-inverse g such that

$$fog = f(g(Y)) = I_Y. \tag{A-2}$$

If f is one-to-one and onto, then it is said to be an one-to-one correspondence. A f is invertible if and only if it is one-to-one correspondence, and thereby has a left and right inverses which are equal. A function f is a homeomorphism if it is one-to-one correspondence and has continuous inverse f^{-1}. Further, if f is continuously differentiable, i.e., C^1 function, then f is called diffeomorphism. C^1 diffeomorphism means that the inverse f^{-1} exists and is also of class C^1 . So, the invertability property of f can be diagrammed as follows:



Functional dependence

Real valued n function $f = (f_1, f_2, \ldots, f_n)^T$ is functionally dependent in an open subset G of X if there exists a function ψ from R^n to R^1 such that

$$\psi(f_1, f_2, \ldots, f_n) = 0 \text{ for } X \varepsilon G \qquad \text{(A-3)}$$

Now introduce following theorems. Proof can be found in many standard analysis text, for example [21], [24], [26].

Theorem A-1 [26]

Let X be a subset of R^n. If f: $X \to R^n$ is C^1 function in an open set $G \varepsilon X$ and the Jacobian J of f is not identically zero for $X \varepsilon G$, then $f_1, f_2, \ldots, f_n$ are functionally independent in G.

Remarks

$\mathrm{DetJf}(x) \neq 0$ for some x implies [15], [32] that for any interior point y of R(f) there exists a neighborhood of y where the C^1 inverse f^{-1} of f is defined. Further, if f^{-1} is unique globally, then f is a C^1-diffeomorphism and f maps R^n onto itself as a one-to-one correspondence. If f^{-1} is not unique globally, then the next inverse function theorem may be used to restrict the domain X on which f is one-to-one.

Theorem A-2 Inverse Function theorem [26]

Let x be an interior point of a set X in R^n and suppose the function $f:X \to Y$, $Y \in R^n$ satisfies the following:

i) f is a class C^1.

ii) $\det Jf(x) \neq 0$,

then, there exists an open set U containing x such that the restriction of f to U, $f_{|U}$ is one-to-one. The inverse f^{-1} is also C^1 on the open set $V=f(U)$.

A generalization of the inverse function theorem to the function of the form $f:R^n x R^r \to R^m$, m is not necessarily equal to n, is the following implicit function theorem.

Theorem A-3 Implicit Function theorem [21]

Let $(x,v)^T$, $x \in R^n$ be an interior point of a set E in $R^n x R^r$ and suppose that the function $f: E \to R^n$ satisfies the following conditions

i) $f(x,v) = y$,

ii) $f(.)$ is C^1 at $(x,v)^T$,

iii) $\det Jf(x,v) \neq 0$

Then, there exists neighborhood N, R of x, v given by

$N = [x-a]x[x+a]$,

$R = [v-b]x[v+b]$,

where a, b are proper real constant vectors, and a C^1 function g: R->N such that

$X = g(Y,V)$,

is the only solution lying in NxR of

$f(X,V) = Y$.

APPENDIX B: DETERMINATION OF THE MAXIMUM ENRTOPY DENSITY

Determination of the maximum entropy density function is derived next. This is useful in the computation of upper bound of the the information contents which is contained in the arbitrary random variable or random process.

Consider a scalar random variable x which has density p(x), but the form of p(x) is not known. Then from (3-4),

$$H(x) = -\int_{-\infty}^{\infty} p(x)\ln p(x)dx. \tag{B-1}$$

One wants to find p(x) which maximizes (B-1) under some constraints. Since maximum entropy density function p(x) is changed as the range of x and constraints are changed. Suppose first that;

maximize (B-1) with

$x \epsilon [0,a]$,

$$\int_{o}^{a} p(x)dx = 1 ,$$

then by the help of the calculus of variations, one can compute maximum entropy density p(x) as

$$p(x) = \begin{cases} 1/a, & 0 \leq x \leq a \\ 0, & \text{elsewhere} \end{cases} \tag{B-3}$$

i.e., uniform density gives the maximum entropy in this case.

But if the range of x is change to

$x \epsilon [0, \infty]$,

and constraints are changed to

$$\int_{0}^{\infty} p(x)dx = 1,$$

$$\int_{0}^{\infty} xp(x)dx = E[x] = m, \tag{B-4}$$

then the result is

$$p(x) = \frac{1}{m} \exp\{-\frac{x}{m}\} , \qquad 0 \leq x < \infty \quad , \tag{B-5}$$

i.e., one-sided exponential density yields the maximum entropy density.

More generally, if

$x \epsilon (-\infty, \infty)$,

with the constraints

$$\int_{-\infty}^{\infty} p(x)dx = 1,$$

$$\int_{-\infty}^{\infty} xp(x)dx = m,$$

$$\int_{-\infty}^{\infty} (x-m)^2 p(x)dx = \text{var } x = \sigma^2, \tag{B-6}$$

then one can prove the following important result [42], [49], [67].

Theorem B-1

For $x \in (-\infty, \infty)$ with the constraint (B-6), the maximum entropy density function p(x) is a Guassian density.

Proof

Problem is to show that the solution p(x) which maximizes (B-1), i.e.,

$$\max.\{-\int_{-\infty}^{\infty} p(x)\, \mathrm{lmp}(x)\,dx\}$$

with the given range and constraints have the form

$$p(x) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{(x-m)^2}{2\sigma^2}\} \quad . \qquad \text{(B-7)}$$

The Lagrangian M for this problem is

$$M = -\int_{\infty}^{\infty} p(x)\ln p(x)dx + \lambda\,[1-\int_{-\infty}^{\infty} p(x)dx]$$
$$+ \mu\,[m - \int_{-\infty}^{\infty} xp(x)dx] + \beta\,[\sigma^2 - \int_{-\infty}^{\infty} (x-m)^2 p(x)dx], \qquad \text{(B-8)}$$

where λ , μ , β are Lagrangian multipliers. Using calculus of variations with

$$\delta\phi(p) = (d\phi/dp)\delta p,$$

$$\delta M = -\int\{\ln p(x) + 1 + \lambda + \mu x + \beta(x-m)^2\}dx.dp \tag{B-9}$$

By setting $\delta M=0$, (B-9) gives

$$\ln p(x) + 1 + \lambda + \mu x + \beta(x-m)^2 = 0, \tag{B-10}$$

or

$$p(x) = \exp\{-1 - \lambda - \mu x - \beta(x-m)^2\}. \tag{B-11}$$

Substitution of (B-11) into (B-6) and solving for λ , μ , β yields Guassian density

$$p(x) = \frac{1}{2\pi\sigma^2} \exp\{- \frac{(x-m)^2}{2\sigma^2}\}. \tag{B-12}$$

So, for any distribution of x next relation holds

$$\begin{aligned} H(x) &\leq H_G(x), \\ &= 1/2 \ln(2\pi e\sigma_x^2). \end{aligned} \tag{B-13}$$

where $H_G(x)$ is a Guassian entropy.

Table B-1 shows H(x) of commonly used density functions for fixed variance. Note that the Gaussian density has the largest entropy.

Table B-1. Entropy of common density functions.

$\sigma_x^2 = 1$

Dist.	H(x)
Gaussian	1.4189
Uniform	1.2425
Triangular	1.3959
Exponential	1.0
Double Exponential	1.3466
Rayleigh	1.3649
Poission (n=10)	1.3879