Chemical Engineering Science Vol. 36, *pp.* 259—272
Pergamon Press Ltd., 1981. Printed in Great Britain

OBSERVABILITY AND REDUNDANCY IN PROCESS DATA ESTIMATION

G. M. STANLEY† and R. S. H. MAH*
Northwestern University, Evanston, IL 60201, U.S.A.

(*Received* 12 *October* 1979; *accepted* 16 *June* 1980)

Abstract—By analogy to the development for dynamic systems, concepts of observability and redundancy may be developed with respect to a steady state system. These concepts differ from their counterparts for dynamic systems in that they can be used to characterize individual variables and local behavior as well as system and global behavior. Relations between local observability, global observability, calculability and redundancy are established and explored in this paper. It is shown that these concepts are useful in characterizing the performance of process data estimators with regard to bias and uniqueness of an estimate, convergence of estimation procedures and the feasibility and implications of problem decomposition.

INTRODUCTION

In our previous investigations [1, 2] we developed data reconciliation techniques for steady state and quasi-steady state (QSS) systems with specific reference to the estimation of temperatures, material and energy flows. We showed that for QSS systems we can construct estimators (discrete Kalman filters) which can take advantage of both temporal and spatial redundancies. Several questions remain unanswered, however. First of all, when will the filter perform adequately? Are there situations in which it will fail? What is the effect of measurement placement on estimator performance? Redundancy has already been shown to be useful, but how does one determine if a measurement is redundant? These questions are clearly of importance in selecting a measurement strategy.

To answer these questions we develop a theory of observability and redundancy in this paper and demonstrate the importance of these concepts in predicting qualitative estimator performance, not only for the QSS filter, but also for constrained least-squares estimators and others.

Observability determines if knowledge of a given set of measurements uniquely determines the state of a system. Originally, observability was defined for dynamic systems [3, 4]. In this paper we define observability as a property of a steady state system that is described by algebraic equations. However, the fundamental question of observability is the same in steady state and dynamic systems: we still want to determine if a given set of measurements can be used to determine the state of the system. In order to render this discussion more precise and concise we need to introduce some definitions and notations. We shall begin with the definition of a steady state system.

*~~Author to whom correspondence should be addressed~~.

†~~Present address: Exxon Chemical Company, Linden, New Jersey, U.S.A.~~.

STEADY STATE OBSERVABILITY

Concepts and definitions

A steady state system of dimension n with l measurements is defined as the triplet (S, $\mathbf{h}$, V), where S is a subset of R^n and $\mathbf{h}$ is a function that maps $\bar{S}$ into R^l. S is called the feasible set and $\mathbf{h}$ is called the measurement function. V is the set from which a particular value of "measurement noise" may be obtained. S is always defined by sets of equality, inequality, or strict inequality constraints. It will be dense in itself. The variables $\mathbf{x} \in R^n$ are called the state variables.

The measurement function $\mathbf{h}$ is used to obtain a set of l measurements $\mathbf{z}$ with additive noise $\mathbf{v}$, i.e.

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{v}, \quad \mathbf{v} \in V, \mathbf{x} \in \bar{S}. \tag{1}$$

If $V = \{\mathbf{0}\}$, then $\mathbf{v} = \mathbf{0}$, $\mathbf{z} = \mathbf{h(x)}$, and we say that the measurements are perfect. If $V \neq \{\mathbf{0}\}$, then the measurements are "noisy". We will make no assumptions on the noise statistics. It may be zero or a constant, or Gaussian, for instance. Figure 1 shows an information flow diagram for a steady state system. Several simple examples of process systems are given in the Appendix.

Suppose we wish to determine the value of a variable x_i by taking measurements using eqn (1). If we know the constraints defining S, the measurement function $\mathbf{h}$, and a set of measurement values $\mathbf{z}^0$, will we be able to determine x_i? This is the question to be answered by our study of observability. Loosely speaking, if the answer to the question is "yes", then x_i is observable. If every x_i, $i = 1, \ldots, n$ is observable, the entire system is observable. Since $\mathbf{g}$ or $\mathbf{h}$ or both may be nonlinear, we define observability as a local property, i.e. as one which depends on the values of the problem variables $\mathbf{x}$. Just as a function can be differentiated at some points but not others, a system can be observable at some points but not others. Since it is easier to define "unobservability", we shall define local unobservability at a point in the system (S, $\mathbf{h}$, V) as a property of the deterministic system (S, $\mathbf{h}$, $\mathbf{0}$).

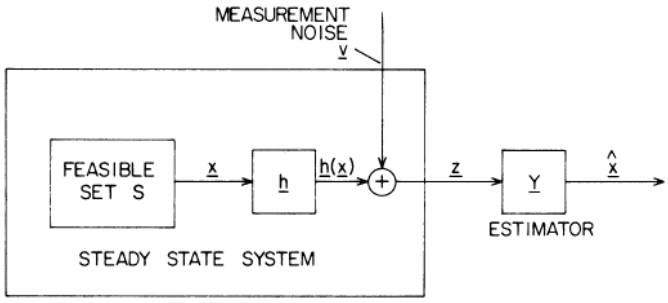

Fig. 1. A steady state system and its estimator.

Definition. In a system (S, $\mathbf{h}$, V), let $\mathbf{x}^0 \in \bar{S}$ and I be an index set. $\mathbf{x}_I$ is *locally unobservable* at $\mathbf{x}^0$ if there exists a sequence $\{\mathbf{x}^k\}_{k=1}^{\infty}$ such that

$$\mathbf{x}^k \to \mathbf{x}^0, \quad \text{as } k \to \infty \tag{2}$$

$$\left.\begin{aligned} &\mathbf{x}^k \in S && (3)\\ &\mathbf{h}(\mathbf{x}^k) = \mathbf{h}(\mathbf{x}^0) && (4)\\ &\delta\mathbf{x}_I{}^k = \mathbf{x}_I{}^k - \mathbf{x}_I{}^0 \neq \mathbf{0} && (5) \end{aligned}\right\} \text{for all } k$$

The $\delta\mathbf{x}^k$ are called *feasible unmeasurable perturbations* about $\mathbf{x}^0$. The perturbations $\delta\mathbf{x}^k$ are "feasible" because the perturbed variables $\mathbf{x}^k = \mathbf{x}^0 + \delta\mathbf{x}^k$ are feasible, and "unmeasurable" because by eqn (4) the perturbations do not affect the measurement values.

Local unobservability at any point is undesirable for variables $\mathbf{x}_I$. Even with perfect measurements, $\mathbf{x}^0$ cannot be distinguished from $\mathbf{x}^k$ since both are feasible, yet both result in the same measurement values. Moreover, it should not be assumed that local unobservability can occur only at isolated points (see Appendix). All other observability concepts except global observability are defined in terms of local unobservability at a point.

Definition. In a system (S, $\mathbf{h}$, V), let $\mathbf{x}^0 \in \bar{S}$. $\mathbf{x}_I$ is *locally observable at* $\mathbf{x}^0$ if it is not unobservable.

Definition. In a system (S, $\mathbf{h}$, V) let $S_1 \subset \bar{S}$. $\mathbf{x}_I$ is *locally unobservable on* S_1 if $\mathbf{x}_I$ is locally unobservable at any point in S_1. $\mathbf{x}_I$ is *locally observable on* S_1 if $\mathbf{x}_I$ is locally observable at each point in S_1. $\mathbf{x}_I$ is *globally unobservable* if $\mathbf{x}_I$ is locally unobservable at any point in S.

For each of the properties we have just defined, if the index set I includes all $i = 1, \ldots, n$, then we say that the vector $\mathbf{x}$ or the system has the given property. Theorem 1 states an alternative formulation of observability that only applies to systems, rather than individual variables.

Theorem 1. Let $\mathbf{x}^0 \in S$ in a system (S, $\mathbf{h}$, $\mathbf{0}$), and let $\mathbf{z}^0 = \mathbf{h}(\mathbf{x}^0)$. The system is locally observable at $\mathbf{x}^0$ if and only if there exists a set $S_1 \subset S$ such that S_1 contains more than one point and $\mathbf{x}^0$ is the unique vector in S_1 satisfying $\mathbf{z}^0 = \mathbf{h}(\mathbf{x})$ and $\mathbf{x} \in S$.

Proof. We prove the theorem in terms of local unobservability. First suppose no such set S_1 exists. Let $\{S_k\}$ be a sequence of nested decreasing sets containing $\mathbf{x}^0$, i.e. $S_k \supset S_{k+1} \supset S_{k+2} \supset \ldots \supset \mathbf{x}^0$. Then for each k, there exists $\mathbf{x}^k \in S_k$ such that $\mathbf{x}^k \neq \mathbf{x}^0$ and $\mathbf{h}(\mathbf{x}^k) = \mathbf{h}(\mathbf{x}^0)$. Thus, $\mathbf{x}^k \to \mathbf{x}^0$ and $\{\mathbf{x}^k\}$ satisfies all the conditions in the definition of local unobservability. Hence (S, $\mathbf{h}$, V) is locally unobservable at $\mathbf{x}^0$.

Conversely, suppose (S, $\mathbf{h}$, V) is locally unobservable. Then a sequence $\{\mathbf{x}^k\}$ as just described exists, i.e. with $\mathbf{x}^k \to \mathbf{x}^0$, $\mathbf{x}^k \neq \mathbf{x}^0$, $\mathbf{h}(\mathbf{x}^k) = \mathbf{z}^0 = \mathbf{h}(\mathbf{x}^0)$ and $\mathbf{x}^k \in S$. Hence solutions $\mathbf{x}^k$ to $\mathbf{z}^0 = \mathbf{h}(\mathbf{x})$ and $\mathbf{x} \in S$ exist arbitrarily close to $\mathbf{x}^0$ and no set S_1 described in the theorem can exist. ■

In other words, local observability at $\mathbf{x}^0$ is equivalent to local uniqueness of the solution $\mathbf{x}^0$ to $\mathbf{z}^0 = \mathbf{h}(\mathbf{x})$, $\mathbf{x} \in S$. Note that in particular, if the system is locally observable, then $\mathbf{h}(\mathbf{x}) \neq \mathbf{h}(\mathbf{x}^0)$ for all feasible points in some neighborhood of $\mathbf{x}^0$. In other words, $\mathbf{x}^0$ can be distinguished from all nearby points in S because the resulting measurement values are different.

We need to define an additional concept closely related to observability:

Definition. The system (S, $\mathbf{h}$, V) is *calculable* on $S_1 \subset \bar{S}$ if $\mathbf{h}$ is one-to-one on S_1. $\mathbf{x}_I$ is *globally observable* on S (or $\bar{S}$) if it is calculable on S (or $\bar{S}$).

Calculability is a very desirable property because it implies that, given a perfect measurement $\mathbf{z}^0$, then the system state can be found as $\mathbf{x}^0 = \mathbf{h}^{-1}(\mathbf{z}^0)$. Note that a system may be locally observable at a point $\mathbf{x}^0$ and yet $\mathbf{h}$ might not be a one-to-one function on any set containing $\mathbf{x}^0$. This point will be brought out in an example. The relationships between the definitions given above are illustrated in Fig. 2.

To illustrate the definitions, consider the following examples. The feasible set for the first two examples is sketched in Fig. 3, and that of the third example is illustrated in Fig. 4.

Example 1. Consider the problem $z = x_1$; $g_1(x_1, x_2) = x_2 - \sin x_1 = 0$; $g_2(x_1, x_2) = x_1 > 0$. This system is globally observable on $\bar{S}$, as are x_1 and x_2.

Example 2. Now let us make $z = x_2$. This example differs from the previous one only in that x_2 is measured rather than x_1. The system is locally observable at each point in S, since a small enough neighborhood can always be placed around a point (such as $\mathbf{x}^1$) so that $x_1^1 \neq x_1$ for any $\mathbf{x}$ in that neighborhood. The system is also calculable on S_2 indicated in Fig. 3, since $\mathbf{h}$ is clearly a one-to-one function on $\bar{S}_2$. On the other hand, if S_1 is any set containing $\mathbf{x}^1$, then neither the system nor x_1 can be calculable on S_1. Neither the system nor x_1, are globally observable on S. However, x_2 is globally observable.

Example 3. In this problem we have
measurements: $z_1 = x_1$, $z_2 = x_2$

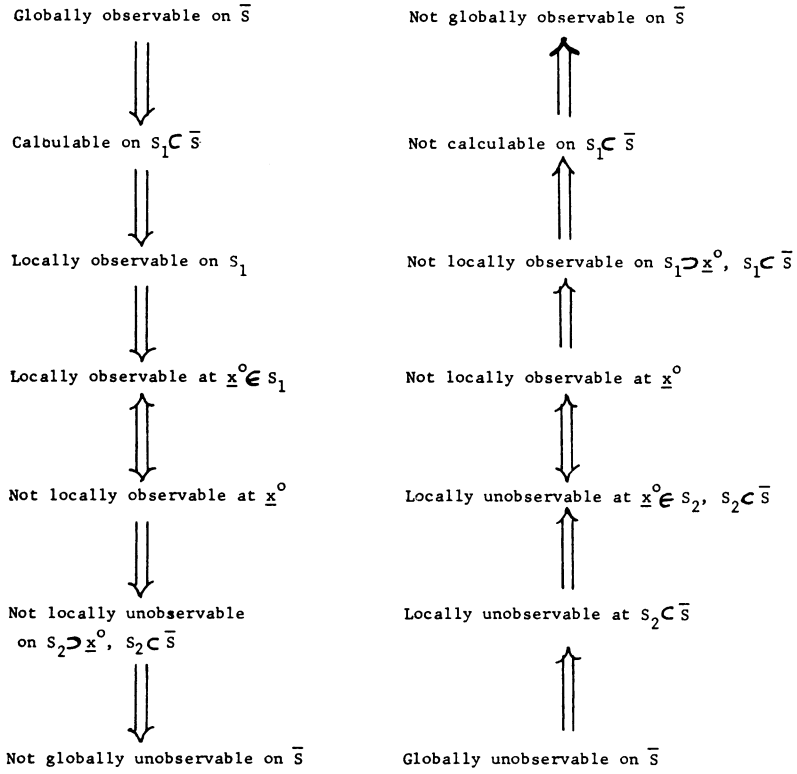

Fig. 2. Relations between observability definitions.

constraints: $(x_1 - x_2)x_3 = 0$
$x_3 \sin(\pi/x_2) = 0$
$0 < x_1 < 1, \quad 0 < x_2 < 1, \quad 0 \le x_3 \le 1$.

S in this case is the union of a bounded plane at $x_3 = 0$ and a series of vertical line segments perpendicular to the line $x_1 = x_2$ as shown in Fig. 4. Note that for $x_3 > 0$, we must have $x_2 \in \{1/2, 1/3, 1/4, \ldots\}$. Let the index set $I = \{1, 2\}$. $\boldsymbol{x_I}$ is globally observable on $\bar{S}$. x_3 is locally unobservable on the vertical line segments (including their intersections with the plane $x_3 = 0$) and is also locally unobservable on the "limiting" line segment from (0,0,0) to (0,0,1). x_3 is locally observable at each of the remaining points. Both x_3 and the system are calculable on any subset of $\bar{S}$ that does not contain any locally unobservable points.

Comparison with previous work

The concept of observability was introduced by Kalman [3,4] for linear dynamic systems. The first observability conditions for nonlinear dynamic systems were developed by Kostyukovskii [*5,* 6], but were later shown to be incorrect [7]. Observability criteria for nonlinear systems have been discovered by Griffith and Kumar[7], Kou, Elliott and Tarn[8], and Singh[9].

At this point, we draw comparisons between the new definition of steady state observability and definitions applying to nonlinear dynamic systems. Many definitions

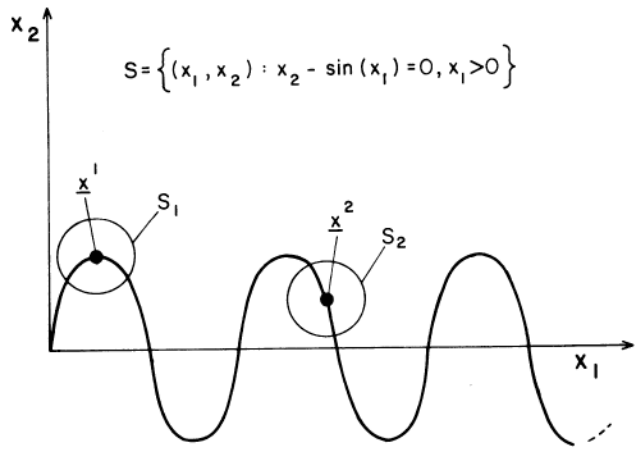

Fig. 3. A feasible set.

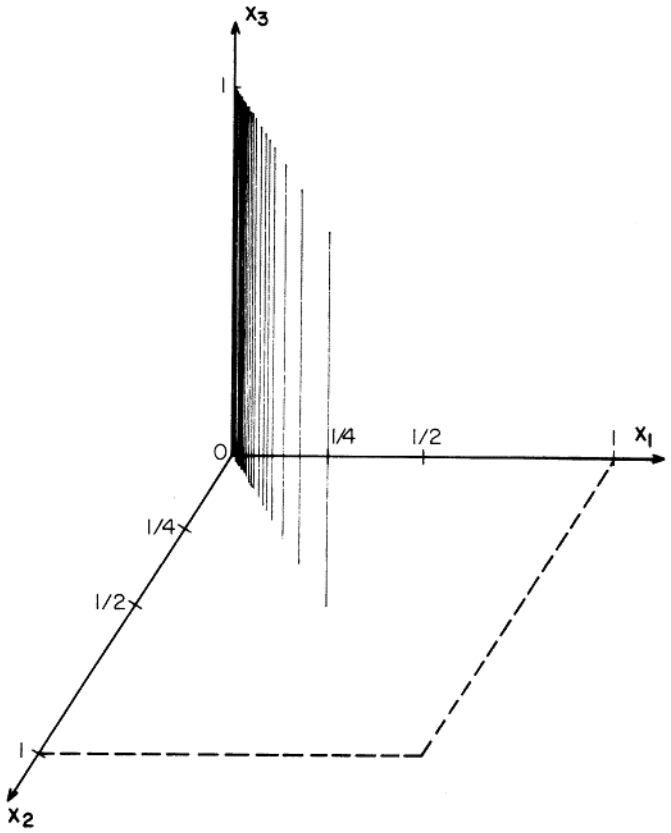

Fig. 4. A feasible set.

are available. For simplicity, we quote an elementary definition from Kou, Elliott and Tarn[8]: The system dynamics are

$$d\mathbf{x}/dt = \mathbf{f}(t,\mathbf{x}). \tag{6}$$

The measurement equation is

$$\mathbf{z} = \mathbf{h}(t,\mathbf{x}). \tag{7}$$

The initial state $\mathbf{x}(t_0)$ is unknown and belongs to set S. This system is (completely) observable in S on the time interval $[t_0, t_1]$ if there exists a one-to-one correspondence between the set S of initial states and the set of trajectories of the measured output $\mathbf{z}(t)$ for t in $[t_0, t_1]$. Note that this definition is global, since it requires that the one-to-one correspondence exists on the entire set S.

Instead of a measurement trajectory, only a measurement vector is available for steady state systems. Calculability is the corresponding concept for steady state systems, since there will be a one-to-one function between S and the range of $\mathbf{h}$ if and only if $\mathbf{h}$ is one-to-one. However, to obtain meaningful estimates, it will not be necessary for $\mathbf{h}$ to be one-to-one on any set. Instead, it will be seen that local observability will be adequate. We replace global criteria with local criteria, and do not require that $\mathbf{h}$ be one-to-one on any set.

Another important aspect of steady state observability is that we have defined it in terms of individual variables. In dynamic systems, unobservable modes will usually affect many variables, and this preciseness might not be necessary. On the other hand we need to know the observability properties of individual variables in steady state flow networks. In dynamic systems, the system observability is all that is usually studied. One notable exception is the work in linear, discrete-time systems by Yoshikawa and Bhattacharyya [10] where "partial observability", or "observability with respect to matrix $\mathbf{T}$" is defined. Instead of attempting to estimate the entire state vector $\mathbf{x}^0$, the authors consider the problem of determining $\mathbf{T}\mathbf{x}^0$, given some matrix $\mathbf{T}$.

In fact, partial observability could have been defined for steady state systems. The definition of local unobservability of $\mathbf{x}_I$ at $\mathbf{x}^0$ could be restated as a definition of system local unobservability at $\mathbf{x}^0$ with respect to $\mathbf{T}$ by replacing eqn (5) with

$$\mathbf{T}(\boldsymbol{x}^k - \mathbf{x}^0) \neq \mathbf{0}. \tag{8}$$

For instance, if $\mathbf{x}$ is partitioned as $\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2)$, define $\mathbf{T} = [\mathbf{I} \vdots \mathbf{0}]$ where the identity matrix corresponds to $\mathbf{x}^1$. Then, where we would state that $\mathbf{x}^1$ is locally observable, the new definition would state that the system is locally observable with respect to $[\mathbf{I} \vdots \mathbf{0}]$. A definition of this form may appear more general than our definition. However, this generality was not needed for the systems studied in this work.

Definitions of observability in nonlinear dynamic systems with inputs are also available in Singh[1] and Griffith and Kumar[11]. The usual definitions of observability

assume either known inputs or unknown and unmeasurable inputs. In reality, some inputs may be measured with measurement error, and some might not be measured, as is the case for the state variables. No distinction is made between state and input variables in this paper. As a result, separate theorems are not required for cases where input variables are partially measured.

In linear dynamic systems the concept of "structural controllability" which is closely related to observability has been developed as an extension of the standard theory [12-14]. This concept does not depend on the numerical values of parameters that are known only approximately. Hence it permits non-numerical controllability classification for linear systems, i.e. classifying matrix model parameters as "zeros" (fixed) and "non-zeros" (approximate). One could introduce a similar definition of steady state structural observability that should be useful for linear systems. However, as the examples in the Appendix will show, local unobservability does not always occur just at isolated points. For such nonlinear systems structural observability is a less useful concept than the observability defined in this paper. We will show in the sequel to this paper that both local and global observability can be classified in certain process systems using non-numerical tests.

We note in passing that the term "identifiable" appears in the statistical literature for linear least-squares estimation[15]. After defining the least-squares estimators in a later section, we will relate identifiability to observability.

Observability classification theorems

We shall now develop several theorems for classifying observability in steady state systems. To simplify statements of theorems and proofs we shall refer to a steady state system (S, $\mathbf{h}$, V) as being in "standard form" if

$$S = \{\mathbf{X}: \mathbf{g}(\mathbf{x})=\mathbf{0}, \mathbf{x}\in K\} \tag{9}$$

where K is the union of a finite number of convex sets $K \subset R^n$, and $\mathbf{g}$ maps R^n into R^n, p being the number of equations, and if,

$$\mathbf{h}(\mathbf{x}) = \mathbf{H}\mathbf{x}+\mathbf{c} \tag{10}$$

where $\mathbf{H}$ is an l x n matrix and $\mathbf{c}$ is a constant vector.

We always assume rank $[\mathbf{H}] < n$, for otherwise the measurements alone could be used to determine $\mathbf{x}$, without even considering S, and the system would be globally observable. In cases where $\mathbf{g}$ is assumed to be differentiable at a point $\mathbf{x}^0$, we define the p x n matrix $\mathbf{G}(\mathbf{x}^0)$,

$$\mathbf{G}(\mathbf{x}^0) = \mathbf{g}'(\mathbf{x}^0) = \partial\mathbf{g}(\mathbf{x}^0)/\partial\mathbf{x}. \tag{11}$$

In the special case where $\mathbf{G}$ is constant, i.e.

$$\mathbf{g}(\mathbf{x}) = \mathbf{G}\mathbf{x}+\mathbf{a} = \mathbf{0} \tag{12}$$

we shall refer to the system as being in standard form with linear (affine) constraints.

The system in the standard form is more general than it may appear at first glance. If $\mathbf{h}$ is nonlinear, the measurements may be linearized by augmenting the state variables by $\mathbf{y}$,

$$\mathbf{z} = \mathbf{y} + \mathbf{v} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix} + \mathbf{v} \tag{13}$$

so that the equality constraints are now

$$\mathbf{g}(\mathbf{y},\mathbf{x}) = \mathbf{g}(\mathbf{x}) = \mathbf{0} \tag{14}$$

$$\mathbf{h}(\mathbf{x}) - \mathbf{y} = \mathbf{0}. \tag{15}$$

Most inequality constraints can be used to define the convex sets K_i. Whenever any doubt arises as to the convexity of a region defined by inequalities, we can rewrite the inequality constraint,

$$g_i(\mathbf{x}) \geq 0 \tag{16}$$

as an equality constraint using a slack variable x_a,

$$g_i(\mathbf{x}) - x_a^2 = 0. \tag{17}$$

Strict inequality constraints can usually be included in this formulation by using them to define K_i.

Note that whenever some equality constraints are derived originally from inequality constraints, the square of a slack variable will be present in that constraint. As a result, while the system may be locally observable, it can never be globally observable because there are two roots for x_a^2. It is still possible for $\mathbf{x}_I$ to be observable, where I is the index set for the original (nonslack) state variables.

We shall now develop criteria for determining whether or not a system in the standard form is observable.

Theorem 2. (First order sufficient condition for calculability, and hence for local observability). For a system with feasible set defined by eqn (9), $\mathbf{x}^0\in S$ and let $\mathbf{g}$, $\mathbf{h}$ be continuously differentiable in some neighborhood of R^n containing $\mathbf{x}^0$. If rank $\begin{bmatrix}\mathbf{g}'(\mathbf{x}^0)\\ \mathbf{h}'(\mathbf{x}^0)\end{bmatrix} = n$, then the system is calculable on some set $S_0 \subset S$ containing $\mathbf{x}^0$.

Proof. Consider the mapping $\begin{bmatrix}\mathbf{g}\\ \mathbf{h}\end{bmatrix}$ which maps R^n into R^{p+l}. By the inverse function theorem[16], $\begin{bmatrix}\mathbf{g}\\ \mathbf{h}\end{bmatrix}$ is a one-to-one map on some neighborhood S_1 of R^n containing $\mathbf{x}^0$. Let $S_0 = S_1 \cap S$. Then, since $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x}\in S_0$, $\mathbf{h}$ must be one-to-one on S_0, and hence the system is calculable on S_0. ■

To illustrate the use of Theorem 2 let us again consider Example 2. Expressed in the standard form we have $g(\mathbf{x}) = -\sin x_1 + x_2 = 0$; $K=\{\mathbf{x}:x_1>0\}$; $\mathbf{g}'(\mathbf{x}^0) = [-\cos x_1^0 \quad 1]$; $\mathbf{h}'(\mathbf{x}^0) = \mathbf{H} = [0\ 1]$ and $\mathbf{c} = \mathbf{0}$. Hence,

$$\text{Rank}\begin{bmatrix}\mathbf{g}'(\mathbf{x}^0)\\ \mathbf{h}'(\mathbf{x}^0)\end{bmatrix} = n = 2, \qquad x_1^0 \neq k\pi/2, \quad k = 1,2,3, \ldots$$

When $x_1 = k\pi/2$, the first order sufficient condition is not met and no conclusions can be drawn. But the test indicates that the system is observable for all other values of $\mathbf{x}$.

We shall now state a related result, which follows directly from the implicit function theorem [16] and Theorem 1.

Theorem 3. For a system with feasible set defined by eqn (9) let $\mathbf{g}$ be continuously differentiable at $\mathbf{x}^0$, and let all the convex sets K_i containing $\mathbf{x}^0$ be open. The system is locally unobservable at $\mathbf{x}^0$ if rank $\begin{bmatrix}\mathbf{g'(x^0)}\\ \mathbf{h'(x^0)}\end{bmatrix} = p+l < n$.

Theorem 4. For a system in standard form with linear constraints and K being open in R^n, if rank $\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix} = n$, then the system is globally observable on S. Conversely, if rank $\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix} < n$, then the system is globally unobservable on S.

Proof. Consider the mapping $\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix}$ from R^n into R^{p+l}. If the rank is n, then $\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix}$ is one-to-one on R^n, and hence one-to-one on S. But $\mathbf{Gx} = -\mathbf{a}$ for $\mathbf{x}$ in S, so $\mathbf{H}$ must be one-to-one on S. Thus, by definition, the system is calculable on S, i.e. globally observable on S.

Conversely, suppose the rank is less than n. Then there exists a vector $\delta\mathbf{x} \neq \mathbf{0}$ such that

$$\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix}\delta\mathbf{x} = \mathbf{0}. \tag{18}$$

Let $\mathbf{x}^0$ be an arbitrary point in S. By eqn (18), for any k,

$$\mathbf{g}(\mathbf{x}^0 + \frac{1}{k}\delta\mathbf{x}) = \mathbf{G}\mathbf{x}^0 + \mathbf{a} + \frac{1}{k}\mathbf{G}\delta\mathbf{x} = \mathbf{0} \tag{19}$$

$$\mathbf{h}(\mathbf{x}^0 + \frac{1}{k}\delta\mathbf{x}) = \mathbf{H}\mathbf{x}^0 + \mathbf{c} + \frac{1}{k}\mathbf{H}\delta\mathbf{x} = \mathbf{h}(\mathbf{x}^0). \tag{20}$$

Define

$$\mathbf{x}^k = \mathbf{x}^0 + \frac{1}{k}\delta\mathbf{x}\,. \tag{21}$$

Since $\mathbf{x}^0$ lies in a convex, relatively open subset of S, the point $\mathbf{x}^k$ must lie in S for $k > N$ for a sufficiently large N. Hence, the sequence $\{\mathbf{x}^k\}$ satisfies the requirements in the definition of local unobservability at $\mathbf{x}^0$ with the sequence of feasible unmeasurable perturbations $\{\frac{1}{k}\delta\mathbf{x}\}$. Since $\mathbf{x}^0$ was arbitrary in S, the system is globally unobservable. ■

As we have shown in Example 2**,** the test based on the first order sufficient condition is not always conclusive at all points. The second order sufficient conditions developed in the next theorem are particularly useful in these situations. Before stating the theorem we will introduce two notations. Let $\mathbf{D}$ be a matrix whose columns form a basis for the solutions to $\begin{bmatrix}\mathbf{g'(x^0)}\\ \mathbf{H}\end{bmatrix}\delta\mathbf{x} = \mathbf{0}$ and let $\mathbf{B}_j(\mathbf{x}^0) = \partial^2 g_j(\mathbf{x}^0)/\partial\mathbf{x}^2$ be the Hessian matrix of g_j, $j = 1,2, \ldots, p$, evaluated at $\mathbf{x}^0$.

Theorem 5. (Second order sufficient conditions for local observability). For a system in a standard form let $\mathbf{x}^0 \in \mathbf{S}$, let $\mathbf{g}$ be twice continuously differentiable in some neighborhood $S_1 \subset R^n$ containing $\mathbf{x}^0$, and let rank $\begin{bmatrix}\mathbf{g'(x^0)}\\ \mathbf{H}\end{bmatrix} < n$. Then the system is locally observable at all feasible points in some set $S_0 \subset S_1$ containing $\mathbf{x}^0$, if for any j, (a) $\mathbf{B}_j(\mathbf{x}^0) > 0$, or (b) $\mathbf{B}_j(\mathbf{x}^0) < 0$, or (c) $\mathbf{D}^T\mathbf{B}_j(\mathbf{x}^0)\mathbf{D} > \mathbf{0}$, or (d) $\mathbf{D}^T\mathbf{B}_j(\mathbf{x}^0)\mathbf{D} < \mathbf{0}$.

Proof. Let K_i be any of the convex sets of K that contains $\mathbf{x}^0$. Let $\mathbf{d} \neq \mathbf{0}$ be an arbitrary "direction vector" for which $\mathbf{Hd} = \mathbf{0}$ and $\mathbf{x}^0 + \mathbf{d} \in K_i$. Any feasible unmeasurable perturbation must satisfy these conditions for at least one i. First suppose $\mathbf{g}'(\mathbf{x}^0)\mathbf{d} \neq \mathbf{0}$. Then it can be shown that any perturbation along this direction is infeasible if it is small enough (for that case, the theorem is proven). Hence, we assume $\mathbf{g}'(\mathbf{x}^0)\mathbf{d} = \mathbf{0}$.

Using a second order Taylor expansion (with integral remainder) for a single function g_j [16],

$$g_j(\mathbf{x}^0 + s\mathbf{d}) = g_j(\mathbf{x}^0) + \mathbf{G}_j(\mathbf{x}^0)s\mathbf{d} + \int_0^1 (s\mathbf{d})^T\mathbf{B}_j(\mathbf{x}^0 + ts\mathbf{d})s\mathbf{d}(1-t)\,dt \tag{22}$$

where $\mathbf{G}_j(\mathbf{x}^0)$ is the j^{th} row of $\mathbf{G}(\mathbf{x}^0)$ and s is a scalar multiplier. Note that the first two items on the right-hand side vanish and that the integrand is a continuous mapping of $s \in R^1$ into R^1 because it is a composite of continuous functions.

If $\mathbf{B}_j > 0$ or $\mathbf{B}_j < 0$, then $\mathbf{d}^T\mathbf{B}_j(\mathbf{x}^0)\mathbf{d} \neq 0$ for any $\mathbf{d} \neq \mathbf{0}$. Because of the continuity property just mentioned, it follows that

$$g_j(\mathbf{x}^0 + s\mathbf{d}) \neq 0 \tag{23}$$

for small enough s. Hence no feasible unmeasurable perturbations can exist and the theorem is proven for those cases. Recall that we only need to consider directions such that both $\mathbf{G}(\mathbf{x}^0)\mathbf{d} = \mathbf{0}$ and $\mathbf{Hd} = \mathbf{0}$, i.e. $\mathbf{d} = \mathbf{Dy}$ for some vector $\mathbf{y}$. Substituting $\mathbf{d}$ in the r.h.s. of eqn (22), it becomes apparent that if $\mathbf{D}^T\mathbf{B}_j(\mathbf{x}^0)\mathbf{D} > 0$ or $\mathbf{D}^T\mathbf{B}_j(\mathbf{x}^0)\mathbf{D} < 0$, we can again argue that $g_j(\mathbf{x}^0 + s\mathbf{d}) \neq 0$ for small enough s. Again, no feasible unmeasurable perturbations can occur, and the theorem is proven. Repeat the argument for each of the convex sets K_i. ■

As an application of Theorem 5, consider the points in Example 2, at which first-order sufficiency conditions fail, i.e. $x_1^0 = k\pi/2$, $k = 1,2,3, \ldots$. The solutions to $\begin{bmatrix}\mathbf{g'(x^0)}\\ \mathbf{H}\end{bmatrix}\mathbf{d} = \mathbf{0}$ are multiples of $\begin{bmatrix}1\\0\end{bmatrix}$. Hence $\mathbf{D} = \begin{bmatrix}1\\0\end{bmatrix}$, $\mathbf{B}(\mathbf{x}^0) = \begin{bmatrix}\sin x_1^0 & 0\\ 0 & 0\end{bmatrix}$, $\mathbf{D}^T\mathbf{B}(\mathbf{x}^0)\mathbf{D} = \sin x_1^0 \neq 0$ for $x_1^0 = k\pi/2$ and hence the test indicates local observability at these points. Using both first and second order conditions, we have classified every feasible point as observable.

EFFECTS OF OBSERVABILITY ON STEADY STATE ESTIMATION

Observability properties depend only on S and $\mathbf{h}$. Defining observability and detecting it has been done without reference to any particular estimation technique. However, as might be expected, observability is closely

linked to the performance of any estimator, even in the presence of measurement noise. We now study the relations between observability and estimation. Readers may wish to refer to Deutsch [17]. and Rhodes [18] for the background material on estimation theory.

Referring to Fig. 1, we use our knowledge of S and $\mathbf{h}$ to construct an estimator that will take the "noisy" measurements $\mathbf{z}$ and produce an estimate $\hat{\mathbf{x}}$ of the state variable $\mathbf{x}$. We use the term "estimator" in a very general sense. It may be derived from statistical principles, or it may simply be a calculation procedure based on some measurements and on the solution of some of the equations in the deterministic system. We can denote the estimation process by

$$\hat{\mathbf{x}} = \mathbf{Y}(\mathbf{z}) \tag{24}$$

where the $\mathbf{Y}$ indicates the mapping from the measurements $\mathbf{z}$ into the estimates. For instance, if all the variables $\mathbf{x}$ were measured directly, $\mathbf{Y}$ might be the identity map, or a more sophisticated estimator might be developed to guarantee that the constraints (9) are met. In some cases to be discussed, e.g. constrained least-squares estimation of an unobservable system, $\mathbf{Y}$ might be a "point-to-set map", i.e. one that produces a set of values of $\hat{\mathbf{x}}$ rather than a single estimate.

For a given measurement noise vector $\mathbf{v}$, the states $\mathbf{x}$ and S are related to the estimate $\hat{\mathbf{x}}$ by the composite map $\mathbf{Y}\circ\mathbf{h}_v$:

$$\hat{\mathbf{x}} = \mathbf{Y}(\mathbf{h}_v(\mathbf{x})) = (\mathbf{Y}\circ\mathbf{h}_v)(\mathbf{x}), \quad \mathbf{x}\in S \tag{25}$$

where

$$\mathbf{h}_v(\mathbf{x}) = \mathbf{h}(\mathbf{x})+\mathbf{v}, \quad \mathbf{v}\in V \tag{26}$$

Ideally, the map $\mathbf{Y}\circ\mathbf{h}_v$ would be one-to-one on S so that two distinct states $\mathbf{x}^1$ and $\mathbf{x}^2$ in S yield different, unique estimates regardless of the value of the measurement noise vector $\mathbf{v}$. If $\mathbf{Y}\circ\mathbf{h}_v$ is one-to-many some estimates will not be unique. If $\mathbf{Y}\circ\mathbf{h}_v$ is many-to-one, then different states may result in the same estimate. $\mathbf{Y}\circ\mathbf{h}_v$ will fail to be one-to-one if either $\mathbf{Y}$ or $\mathbf{h}_v$ fails to be one-to-one. We will study these two cases separately. First we treat the effect of unobservability on the estimates.

When a system is unobservable at some point, even perfect measurements and constraints are not sufficient to distinguish between two feasible $\mathbf{x}$ values. The addition of measurement noise cannot improve upon this situation. This property, which follows directly from the definition of local unobservability, is stated below.

Property 1. If $\mathbf{x}_I$ is locally unobservable at $\mathbf{x}^0\in S$, then any estimator for (S, $\mathbf{h}$, V) will fail to distinguish between $\mathbf{x}^0$ and all members of some sequence $\{\mathbf{x}^k\}$ where $\mathbf{x}_I{}^k \neq \mathbf{x}_I{}^0$ and $\mathbf{x}^k \rightarrow \mathbf{x}^0$, as $k \rightarrow \infty$, whether the measurements are perfect or noisy.

Property 1 implies that if the system is locally unobservable at some point, then $\mathbf{h}$ is many-to-one, and hence so is $\mathbf{h}_v$. No matter how cleverly an estimator is constructed, there is no way to determine the true state of the system. The basic problem is that either insufficient equations are known about the system, or the measurements are inadequate.

One feature of Property 1 should be noted. It assumes that the models of S and $\mathbf{h}$ truly represent physical phenomena. In reality, the constraints and measurement functions may be inexact; additional equations describing the physical phenomena may have been omitted. Thus, some of the feasible unmeasurable perturbations that can occur in our simple model may be physically impossible. For this reason, we should also study the effect of observability on particular estimation techniques. An estimator is artificially constructed as a set of equations that can be implemented exactly as written on a computer within round-off errors. As such, there is no uncertainty about its equations, and one can make precise statements about its behavior that do not depend on the physical phenomena.

Property 1 predicts failure of any estimator if the measurements are inadequate and result in unobservability. On the other hand, will any form of observability guarantee the "success" (in some sense) of any estimator? The measurements may contain enough information but one cannot make general statements that observability results in good estimates. Estimator performance will depend on the particular form of the estimator. We shall now consider the effects of observability on some particular estimators for particular systems.

Linear estimation

Since we shall be concerned with noisy measurements from this point on, we shall replace eqn (10) by

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{c} + \mathbf{v} \tag{27}$$

and refer to the steady state system so defined with the general constraint set S as being in the modified standard form. Let us define a linear (affine) estimator by the equation

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{z}+\mathbf{b} \tag{28}$$

where $\mathbf{b}$ and $\mathbf{W}$ can be chosen arbitrarily.

Theorem 6. For a system in modified standard form, it is possible to construct an unbiased linear estimator, if and only if the system is calculable on S.

Proof. Let $E(\mathbf{v}) = \bar{\mathbf{v}}$, and let $\mathbf{x}^0$ be any point in S. Then for a set of measurements $\mathbf{z}^0$ we have

$$E(\hat{\mathbf{x}}) = E(\mathbf{W}\mathbf{z}^0 + \mathbf{b}) = \mathbf{W}\mathbf{H}\mathbf{x}^0 + \mathbf{W}\mathbf{c} + \mathbf{W}\bar{\mathbf{v}} + \mathbf{b}. \tag{29}$$

It is easily seen that the estimator gives unbiased estimates for all points $\mathbf{x}^0\in S$, if and only if

$$\mathbf{b} = -\mathbf{W}\mathbf{c} - \mathbf{W}\bar{\mathbf{v}} \tag{30}$$

and

$$\mathbf{W}\mathbf{H}\mathbf{x}^0 = \mathbf{x}^0, \ \mathbf{x}^0\in S. \tag{31}$$

In other words, **WH** must be the identity map on S, but not on R^n if $l < n$.

Note that given **W** we can always choose **b** to satisfy eqn. (30). However, it is not always possible to choose **W** to satisfy eqn (31), as we now show.

First, suppose the system is not calculable on S. Then there exists $\mathbf{x}^1$ and $\mathbf{x}^2$ in S such that $\mathbf{x}^1 \neq \mathbf{x}^2$ and $\mathbf{H}\mathbf{x}^1 + \mathbf{c} = \mathbf{H}\mathbf{x}^2 + \mathbf{c}$. Hence $\mathbf{H}\mathbf{x}^1 = \mathbf{H}\mathbf{x}^2$ and $\mathbf{W}\mathbf{H}\mathbf{x}^1 = \mathbf{W}\mathbf{H}\mathbf{x}^2$ where $\mathbf{x}^1 \neq \mathbf{x}^2$. It follows that **WH** cannot be the identity map on S for any choice of **W**. Hence any linear estimator is biased on S when the system is not calculable on S.

On the other hand, suppose the system is calculable on S. Then **H** is a one-to-one function on S by definition. Let U be the minimal subspace containing S, generated by taking all linear combinations of vectors in S. Then clearly **H** is one-to-one on U, hence a left inverse **W** exists, i.e. there exists a matrix **W** such that $\mathbf{W}\mathbf{H}\mathbf{x} = \mathbf{x}$ for $\mathbf{x} \in U$. But $S \subset U$, and hence we have found a **W** to construct an unbiased estimator. ■

Constrained least-squares estimation

For a set of measurements z let us define

$$\omega(\mathbf{x}) = [\mathbf{z} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \tag{32}$$

where **R** is a weighting matrix frequently chosen as an approximation to the covariance matrix of measurement noise. In this discussion we shall assume **R** to be positive definite. Constrained least-squares estimates are obtained by solving the following problem:

$$\min_{\mathbf{x}} \omega(\mathbf{x}), \quad \mathbf{x} \in S \tag{33}$$

for which a solution is assumed to exist. Numerical procedures may be used to solve this problem. Most algorithms only guarantee convergence to a local minimum point, and hence the following theorems will be stated in terms of local solutions.

Theorem 7. For a system (S, **h**, V), let $\hat{\mathbf{x}}$ be a constrained least squares estimate. Suppose the system is locally unobservable at $\hat{\mathbf{x}}$. Then $\hat{\mathbf{x}}$ is not a unique local solution to the least squares estimation problem.

Proof. By local unobservability, there exists a sequence $\{\mathbf{x}^k\}$, $\mathbf{x}^k \in S$ such that $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{h}(\mathbf{x}^k)$, $\mathbf{x}^k \to \hat{\mathbf{x}}$ and $\mathbf{x}^k \neq \hat{\mathbf{x}}$. Then, by eqn (32), $\omega(\mathbf{x}^k) = \omega(\hat{\mathbf{x}})$. Since each $\mathbf{x}^k$ is feasible in the least squares estimation problem, and since we can pick $\mathbf{x}^k$ arbitrarily close to $\hat{\mathbf{x}}$, $\hat{\mathbf{x}}$ is not a unique solution. ■

In this case, **Y** is a point-to-set map, and hence not one-to-one. If the estimation problem for an unobservable system is to be solved using a nonlinear optimization algorithm, an estimate may be obtained, but it will not be unique. If the problem is to be solved analytically, difficulties may arise. For instance, a solution in closed form will not be possible without additional constraints. If the constraints and the measurement function are linear, as in the case of Kuehn and Davidson[19], a singular matrix would have to be inverted.

While the proof of Theorem 7 is trivial, the implications are important. Note that the uniqueness of a solution depends on the observability at the value of the estimate $\hat{\mathbf{x}}$ rather than the observability at the true value in the system. The value of the estimate depends on the value of the measurement noise **v**, so that even if the deterministic system is locally observable on a large subset $S_1 \subset S$ containing the true value of **x**, $\hat{\mathbf{x}}$ may be far away from any values in S_1 and may be nonunique. For instance, with a constant measurement bias, the set S_2 of **x** values that result in unique estimates will be a "shifted" and "distorted" version of S_1. On the other hand if a system is locally unobservable on a subset S_1, with random noise, the closer the estimate is to S_1, the more likely nonunique answers will result.

From the proof of the theorem, it should be noted that multiple solutions to the estimation problem (for an unobservable system) occur arbitrarily close to $\hat{\mathbf{x}}$. This is an important point. In nonlinear systems, it is not surprising when multiple solutions to equations exist. In ordinary situations, however, many solutions can be ruled out on physical grounds, or ruled out based on prior estimates. Thus, isolated multiple solutions may not be so serious. By starting a nonlinear iterative procedure near a good prior estimate, isolated multiple solutions may not ever be noticed. However, the proof shows that the multiple solutions in an unobservable system are arbitrarily close together. In most cases, there are surfaces on which the system is unobservable. In solving the optimization problem, an algorithm may locate any point on that surface and even search along that surface. As long as a system is observable (in some way), there is always the hope that by improving the measuring devices to reduce measurement noise, steady state solutions can be separated by the measurements. However, when a system is locally unobservable at some point, problems will arise even if the measurements are perfect.

For the special case of estimation with linear measurements and constraints, there is no distinction between local and global properties. Hence, if the system is unobservable, the estimates will always be nonunique at any $\hat{\mathbf{x}}$.

Note that Theorem 7 applies to a general steady state system. If the system can be put in the standard form, more detailed results can be given. Before doing this, the following lemma must be proven:

Lemma. For a system in standard form let $\hat{\mathbf{x}}$ be a constrained least-squares estimate. Then there exists a neighborhood of zero such that $(\boldsymbol{\omega}'(\hat{\mathbf{x}}))\delta\mathbf{x} \geq 0$ for all $\delta\mathbf{x}$ in that neighborhood for which $\hat{\mathbf{x}} + \delta\mathbf{x} \in \bar{S}$.

Proof. Suppose the lemma is false. Then there exists a sequence $\{\mathbf{x}^k\}$ such that $\mathbf{x}^k \to \hat{\mathbf{x}}$, $\mathbf{x}^k \in \bar{S}$ and $(\boldsymbol{\omega}'(\hat{\mathbf{x}}))\delta\mathbf{x}_k < 0$ where $\delta\mathbf{x}_k = \mathbf{x}^k - \hat{\mathbf{x}}$. Since $\hat{\mathbf{x}}$ is a relative minimum point, there is some $S_1 \subset \bar{S}$ for which $0 \leq \omega(\mathbf{x}) - \omega(\hat{\mathbf{x}})$, $\mathbf{x} \in S_1$. Let $\{\mathbf{d}_k\}$ be a sequence of direction vectors and $\{a_k\}$ be a sequence of scalars such that $\mathbf{x}^k = \hat{\mathbf{x}} + a_k\mathbf{d}_k$, $\|\mathbf{d}_k\| = 1$, $a_k > 0$, and $a_k \to 0$, as $k \to \infty$. Since $\{\mathbf{d}_k\}$ is a bounded sequence, it must have a convergent sub-sequence. Let $\{a_k\mathbf{d}_k\}$ be such a convergent subsequence. Then by the Taylor expansion which is exact for the quadratic, ω,

$$0 \leq \omega(\mathbf{x}^k) - \omega(\hat{\mathbf{x}}) = a_k(\boldsymbol{\omega}'(\hat{\mathbf{x}}))\mathbf{d}_k + a_k^2 \mathbf{d}_k^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{d}_k. \tag{34}$$

Since $a_k \to 0$, we must have $(\boldsymbol{\omega}'(\hat{\mathbf{x}}))\mathbf{d}_k \geq 0$, $k > N$ for a sufficiently large N. But this contradicts the hypothesis. Hence the Lemma is true. ■

Note that the Lemma was not a trivial application of directional derivatives as it may have appeared at first glance. The reason is that S was not assumed convex, and hence we could not take derivatives along a line connecting $\hat{\mathbf{x}}$ to $\hat{\mathbf{x}} + \delta\mathbf{x}_k$ and maintain feasibility. In fact, no information about S was used at all — S might not have been connected, or there might have been no smooth curves in S. Now the theorem relating observability and estimation for systems in standard form can be proven.

Theorem 8. For a system in standard form, let $\hat{\mathbf{x}}$ be a constrained least-squares estimate. It is locally unique, if and only if the system is locally observable at $\hat{\mathbf{x}}$.

Proof. If the system is locally unobservable, then the conclusion follows from the more general Theorem 7. We must prove the other half of the theorem.

Let $(S, \mathbf{h}, V)$ be locally observable at $\hat{\mathbf{x}}$. Let $\{\mathbf{x}^k\}$ be any sequence with $\mathbf{x}^k \to \mathbf{x}^0$, $\mathbf{x}^k \in \overline{S}$, and define $\delta\mathbf{x}_k = \mathbf{x}^k - \hat{\mathbf{x}}$. Then by local observability, $\mathbf{H}\delta\mathbf{x}_k \neq \mathbf{0}$, $k > N$ for a sufficiently large N. But since ω is a quadratic function, its Taylor series representation is exact:

$$\omega(\hat{\mathbf{x}} + \delta\mathbf{x}_k) = \omega(\hat{\mathbf{x}}) + (\boldsymbol{\omega}'(\hat{\mathbf{x}}))\delta\mathbf{x}_k + \delta\mathbf{x}_k{}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{x}_k \quad (35)$$

By the Lemma, $(\boldsymbol{\omega}'(\hat{\mathbf{x}}))\delta\mathbf{x}_k \geq 0$, and since $\mathbf{R} > 0$, $\delta\mathbf{x}_k{}^T\mathbf{H}^T\mathbf{R}\mathbf{H}\delta\mathbf{x}_k > 0$. Hence $\omega(\hat{\mathbf{x}} + \delta\mathbf{x}_k) > \omega(\hat{\mathbf{x}})$. Since the sequence $\{\mathbf{x}^k\}$ was arbitrary, other solutions to the minimization problem cannot be arbitrarily close to $\hat{\mathbf{x}}$, and hence $\hat{\mathbf{x}}$ is a local minimum point. ■

In the theory of linear least-squares estimation, the concept of identifiability[15] is closely related to observability. If the unconstrained linear least-squares estimation problem has multiple solutions, the states are said to be nonidentifiable. Then, linear constraints are arbitrarily introduced to remove the ambiguity. If the constraints render the solution unique, they are said to be "suitable for identifiability". For our purposes, the problem is not to find constraints that are suitable for identifiability: the constraints are already known. Observability is a general property of an individual variable in a nonlinear system, whereas identifiability conditions are mathematical conveniences for the unconstrained linear least-squares estimation problem for a system.

Decomposition according to observability

We shall now state a decomposition theorem that applies to systems with linear constraints.

Theorem 9. For a system in standard form with linear constraints if $\operatorname{rank}\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix} = j < n$, then there exists a nonsingular $n \times n$ matrix $\mathbf{T}$ such that for the change of coordinates $\mathbf{x} = \mathbf{T}\tilde{\mathbf{x}}$,

$$\tilde{\mathbf{G}} = \mathbf{G}\mathbf{T} = [\mathbf{G}_1 \vdots \mathbf{0}] \quad (36)$$

and

$$\tilde{\mathbf{H}} = \mathbf{H}\mathbf{T} = [\mathbf{H}_1 \vdots \mathbf{0}] \quad (37)$$

where $\mathbf{G}_1$ and $\mathbf{H}_1$ each have j columns and $\operatorname{rank}\begin{bmatrix}\mathbf{G}\\ \mathbf{H}\end{bmatrix} = j$.

The implication of this theorem is that in the new coordinates $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}^1, \tilde{\mathbf{x}}^2)$ the subsystem $\mathbf{G}_1\tilde{\mathbf{x}}^1 + \mathbf{a} = \mathbf{0}$ and $\mathbf{z} = \mathbf{H}_1\tilde{\mathbf{x}}^1 + \mathbf{c}$ in R^j is observable, but the unconstrained subsystem in R^{n-j} is unobservable. Kalman[4] first suggested the decomposition of linear dynamic systems into observable and unobservable, as well as controllable and uncontrollable, subsystems.

The proof follows readily from reduction to the column-echelon form [20]. Alternatively, $\mathbf{T}$ could be found using a variety of standard methods from linear algebra such as Gram-Schmidt orthogonalization, singular value decomposition, or pseudo inverse. Rather than dwelling on the proof, we shall examine its application to the constrained least-squares estimation with prior distribution:

$$\min_{\mathbf{x}} \{(\mathbf{x} - \mathbf{x}_0)^T\mathbf{W}(\mathbf{x} - \mathbf{x}_0) + (\mathbf{z} - \mathbf{H}\mathbf{x} - \mathbf{c})^T\mathrm{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{x} - \mathbf{c})\} \quad (38)$$

subject to

$$\mathbf{G}\mathbf{x} + \mathbf{a} = \mathbf{0} \quad (39)$$

where $\mathbf{x}_0$ is a prior estimate and $\mathbf{W} > 0$. $\mathbf{W}^{-1}$ is usually taken as an approximation to the covariance matrix of $\mathbf{x}_0$. Since both the objective function and the constraints are convex, the estimate will be unique.

Applying the coordinate transformation we get

$$\min_{\tilde{\mathbf{x}}} \{(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_0)^T\mathbf{T}^T\mathbf{W}\mathbf{T}(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_0) + (\mathbf{z} - \mathbf{H}_1\tilde{\mathbf{x}}^1 - \mathbf{c})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}_1\tilde{\mathbf{x}}^1 - \mathbf{c})\} \quad (40)$$

subject to

$$\mathbf{G}_1\tilde{\mathbf{x}}^1 + \mathbf{a} = \mathbf{0}. \quad (41)$$

Since the variables $\tilde{\mathbf{x}}^2$ are unconstrained, and $\mathbf{T}^T\mathbf{W}\mathbf{T} > 0$, $\hat{\tilde{\mathbf{x}}}^2 = \tilde{\mathbf{x}}_0^2$ is a minimizing solution for $\tilde{\mathbf{x}}^2$. Then, the problem reduces to that of finding a solution to the constrained problem in $\tilde{\mathbf{x}}^1$. The important point here is that the measurements contributed no information to the determination of $\tilde{\mathbf{x}}^2$. A unique solution can be obtained only because of the prior estimate, but that part of the solution $\hat{\mathbf{x}} = \mathbf{T}\hat{\tilde{\mathbf{x}}}$ which depends on $\hat{\tilde{\mathbf{x}}}^2$ is only as good as the prior estimate.

In some systems, it may happen that the variables $\tilde{\mathbf{x}}^1$ have physical meaning, i.e. we might be interested in their values. In this case, we have just shown that a decomposition procedure can be used: Estimate only the observable variables $\tilde{\mathbf{x}}^1$. This procedure was used in mass flow networks [1]. Note that, if no prior estimate is available ($\mathbf{W} \to \mathbf{0}$), then the problem reduces to the usual least-squares problem, and $\hat{\tilde{\mathbf{x}}}^2$ is arbitrary. A similar decomposition has been done for the Luenberger observer to isolate the effect of unmeasurable input disturbances [21].

OBSERVABILITY AND FILTERING IN A QSS SYSTEM

A quasi-steady state (QSS) system [2] consists of a set of p steady state algebraic constraints in n variables at each time t_k,

$$\mathbf{g}(\mathbf{x}(k)) = \mathbf{0}, \qquad k=0,1,... \tag{42}$$

a set of l measurements taken at each time t_k,

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)) + \mathbf{v}(k), \qquad k=0,1,... \tag{43}$$

and a set of transition equations with "process noise" $\mathbf{w}(k)$,

$$\mathbf{x}^2(k+1) = \mathbf{x}^2(k) + \mathbf{w}(k), \qquad k=0,1,... \tag{44}$$

where $\mathbf{x}(k)$ is partitioned as

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}^1(k) \\ \mathbf{x}^2(k) \end{bmatrix}, \quad \mathbf{x}^1(k) \in R^p, \quad \mathbf{x}^2(k) \in R^{n-p} \tag{45}$$

It will be assumed that eqn (42) can be used to solve uniquely for $\mathbf{x}^1(k)$ in terms of $\mathbf{x}^2(k)$, either analytically or numerically.

We shall now consider the QSS system derived from the system with linear constraints, eqn (12), and linear measurements, eqn (27). We shall assume that the redundant rows of $\mathbf{G}$ have been deleted and $\mathbf{G} = [\mathbf{G}_1 \vdots \mathbf{G}_2]$ has been permuted so that $\mathbf{G}_1$ is square and invertible and $\mathbf{x}$ is similarly partitioned into $[\mathbf{x}^1 \vdots \mathbf{x}^2]$ as in eqn (45). Define

$$\mathbf{\Gamma} = \begin{bmatrix} -\mathbf{G}_1^{-1}\mathbf{G}_2 \\ \mathbf{I} \end{bmatrix} \tag{46}$$

and

$$\mathbf{H}^* = \mathbf{H}\mathbf{\Gamma}. \tag{47}$$

Theorem 10. The steady state system with linear constraints and measurements is observable, if and only if the unconstrained system $\mathbf{z} = \mathbf{H}^*\mathbf{x}^2 + \mathbf{v}$ is observable. Furthermore, $\mathbf{H}^*$ is of full rank n-p, if and only if either system is observable.

Proof. Note that $\mathbf{G}\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{\Gamma}\mathbf{x}^2$. Let the first system be unobservable. Then there exists a $\delta\mathbf{x}$ such that $\mathbf{H}\delta\mathbf{x} = 0$ and $\delta\mathbf{x} = \mathbf{\Gamma}\delta\mathbf{x}^2$. Hence $\mathbf{0} = \mathbf{H}\delta\mathbf{x} = \mathbf{H}\mathbf{\Gamma}\delta\mathbf{x}^2 = \mathbf{H}^*\delta\mathbf{x}^2$. It follows that feasible unmeasurable perturbations can exist in the second system so it is unobservable. Conversely, let the second system be unobservable. Then there exists a $\delta\mathbf{x}^2$ with $\mathbf{0} = \mathbf{H}^*\delta\mathbf{x}^2 = \mathbf{H}\mathbf{\Gamma}\delta\mathbf{x}^2$. Hence feasible unmeasurable perturbations $\delta\mathbf{x} = \mathbf{\Gamma}\delta\mathbf{x}^2$ can exist in the first system, so it is unobservable. The rank of $\mathbf{H}^*$ follows from Theorem 4. ■

It follows from Theorem 10 that without any loss of generality we can restrict our considerations to the unconstrained QSS system on the assumption that partition and transformation prescribed by eqns (46) and (47) have been previously carried out. The QSS system with statistical assumptions may then be described by

$$\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k) \tag{48}$$

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{w}(k) \tag{49}$$

$$\mathbf{v}(k) \sim N(\mathbf{0},\mathbf{R}) \tag{50}$$

$$\mathbf{w}(k) \sim N(\mathbf{0},\mathbf{Q}) \tag{51}$$

$$\mathbf{x}(0) \sim N(\bar{\mathbf{x}}(0), \mathbf{P}_0) \tag{52}$$

and the Kalman filter for this system by

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}(k-1) + \mathbf{K}_k[\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k-1)] \tag{53}$$

where $\mathbf{K}_k$ is the gain matrix [2].

Theorem 11. A linear QSS system is observable in the classical (dynamic systems) sense, if and only if the corresponding steady state system is observable.

Proof. The proof is based on the observability matrix[11] for dynamic systems. By Theorem 10 we need only to consider the unconstrained case for which the observability matrix reduces to $[\mathbf{H}^T \vdots \ldots \vdots \mathbf{H}^T]$. The QSS system is observable if and only if rank $[\mathbf{H}^T \vdots \ldots \vdots \mathbf{H}^T] = n$, which is true if and only if rank $[\mathbf{H}] = n$, which is true if and only if the steady state system is observable by Theorem 4. ■

It should be stressed that this theorem is the only link given in this paper between steady state observability and classical dynamic system observability. It only applied to the very restricted class of QSS systems derived from linear steady state systems with no inequality or other constraints. The dynamics are extremely restricted and are not easily generalized using the technique in Theorem 11.

Note that if the steady state system is unobservable, the measurements cannot distinguish between some feasible vectors, say $\mathbf{x}^1$ and $\mathbf{x}^2$. Then, if either the steady state or the QSS system undergoes a step change from $\mathbf{x}^1$ to $\mathbf{x}^2$, the measurements will not detect the change, and neither will a Kalman filter for the QSS system. This analysis assumes that the step change is physically possible. We now study the implications of steady state observability on the performance of the QSS filter regardless of the actual physical phenomena.

Theorem 12. If a steady state system with linear constraints and measurements is observable, and if $\mathbf{Q} > 0$, $\mathbf{R} > 0$, then the Kalman filter for the corresponding QSS system is stable. Furthermore, the matrices $\mathbf{P}_k(+)$ in the Riccati equation converge to a unique positive definite matrix $\mathbf{P}(+)$ as $k \rightarrow \infty$.

Proof. The Kalman filter is applied to the system, eqns (48)-(53) where $\mathbf{H}$ has full rank n by the Theorem 10. Consider the "reduced" system obtained by setting $\mathbf{v}(k) = \mathbf{0}$ for all k. The resulting "reduced" system is observable in the classical (dynamic systems) sense by Theorem 11. Furthermore, the "reduced" system is controllable (with $\mathbf{w}(k)$ as the "control variable") because the usual controllability matrix[2] reduces to $[\mathbf{I} \vdots \mathbf{I} \vdots \ldots \vdots \mathbf{I}]$ which has rank n. Since $\mathbf{Q} > 0$, $\mathbf{R} > 0$, and the "reduced" model is controllable and observable, it follows that $\mathbf{P}_k(+)$ in the Riccati equation converges to a unique positive definite matrix, and the Kalman filter is stable [22]. ■

The converse to this theorem concerning the Riccati equation is treated in Theorem 13.

Since the remaining theorems on filtering in a QSS system depend on the coordinate transformation,

$$\mathbf{x} = \mathbf{T}\,\tilde{\mathbf{x}} \tag{54}$$

where $\mathbf{T}$ is an $n \times n$ nonsingular matrix, it is convenient to summarize the results and introduce the notation. The general effect of this transformation is to replace $\mathbf{H}$, $\mathbf{x}$, $\mathbf{w}$, $\mathbf{Q}$, $\mathbf{P}_0$ by $\tilde{\mathbf{H}}$, $\tilde{\mathbf{x}}$, $\tilde{\mathbf{w}}$, $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{P}}_0$ in eqns (48)-(52), where

$$\tilde{\mathbf{H}} = \mathbf{H}\mathbf{T} \tag{55}$$

$$\tilde{\mathbf{w}} = \mathbf{T}^{-1}\mathbf{w} \tag{56}$$

$$\tilde{\mathbf{Q}} = \mathbf{T}^{-1}\mathbf{Q}(\mathbf{T}^{-1})^T \tag{57}$$

$$\tilde{\mathbf{P}}_0 = \mathbf{T}^{-1}\mathbf{P_0}(\mathbf{T}^{-1})^T. \tag{58}$$

The Riccati equation in the Kalman filter for the original QSS system is given by

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1}(-) = \mathbf{P}_k + \mathbf{Q} - \mathbf{P}_k\mathbf{A}_k\mathbf{P}_k \tag{59}$$

where

$$\mathbf{A}_k = \mathbf{H}^T(\mathbf{H}\mathbf{P}_k\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}. \tag{60}$$

After the transformation it becomes

$$\tilde{\mathbf{P}}_{k+1} = \tilde{\mathbf{P}}_k + \tilde{\mathbf{Q}} - \tilde{\mathbf{P}}_k\,\tilde{\mathbf{A}}_k\,\tilde{\mathbf{P}}_k \tag{61}$$

where

$$\tilde{\mathbf{A}}_k = \tilde{\mathbf{H}}^T(\tilde{\mathbf{H}}\,\tilde{\mathbf{P}}_k\,\tilde{\mathbf{H}}^T + \mathbf{R})^{-1}\tilde{\mathbf{H}}. \tag{62}$$

It follows that

$$\tilde{\mathbf{P}}_k = \mathbf{T}^{-1}\mathbf{P}_k(\mathbf{T}^{-1})^T \quad \text{for all } k. \tag{63}$$

Theorem 13. If a steady state system in standard form with linear constraints is unobservable, then in the Kalman filter for the corresponding QSS system, the matrices $\mathbf{P}_k(+)$ in the Riccati equation cannot converge.

Proof. Applying the coordinate transformation $\mathbf{x} = \mathbf{T}\,\tilde{\mathbf{x}}$ prescribed by Theorem 9 we have $\mathbf{H} = [\mathbf{H}_1 \vdots \mathbf{0}]$ where $\mathbf{H}_1$ is of full rank j and the subsystem $\mathbf{z} = \mathbf{H}_1\tilde{\mathbf{x}}^1 + \mathbf{v}$ is observable. If we let $\mathbf{P}_k$ be partitioned such that

$$\tilde{\mathbf{P}}_k = \begin{array}{c} \\ j \\ n-j \end{array} \begin{array}{c} \begin{array}{cc} j & n-j \end{array} \\ \left[\begin{array}{c|c} \mathbf{P}_{1k} & \mathbf{P}_{3k} \\ \hline \mathbf{P}_{3k} & \mathbf{P}_{2k} \end{array}\right] \end{array} \tag{64}$$

then eqn (62) may be rewritten as

$$\tilde{\mathbf{A}}_k = \begin{array}{c} \\ j \\ n-j \end{array} \begin{array}{c} \begin{array}{cc} j & n-j \end{array} \\ \left[\begin{array}{c|c} \mathbf{A}_{1k} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \end{array} \tag{65}$$

where

$$\mathbf{A}_{1k} = \mathbf{H}_1{}^T(\mathbf{H}_1\mathbf{P}_{1k}\mathbf{H}_1{}^T + \mathbf{R})^{-1}\mathbf{H}_1. \tag{66}$$

Hence $\tilde{\mathbf{P}}_k\,\tilde{\mathbf{A}}_k\,\tilde{\mathbf{P}}_k$ is only positive semidefinite.

Now suppose $\tilde{\mathbf{P}}_k \to \mathbf{P}$. Then taking the limits on eqn (61), $\mathbf{P} = \mathbf{P} + \mathbf{Q} - \mathbf{P}\mathbf{A}\mathbf{P}$ where $\mathbf{A} = \tilde{\mathbf{H}}^T(\tilde{\mathbf{H}}\,\mathbf{P}\,\tilde{\mathbf{H}}^T + \mathbf{R})^{-1}\tilde{\mathbf{H}}$. Hence $\mathbf{Q} = \mathbf{P}\mathbf{A}\mathbf{P}$. But for each k there exists an $\mathbf{x}(k)$ such that $\|\mathbf{x}(k)\| = 1$ and $\mathbf{x}(k)^T\mathbf{P}_k\mathbf{A}_k\mathbf{P}_k\mathbf{x}(k) = 0$. Since the sequence $\{\mathbf{x}(k)\}$ is bounded, it must contain a convergent sub-sequence which, for convenience, shall again be labelled $\{\mathbf{x}(k)\}$. Taking the limit as $k \to \infty$, $\mathbf{x}(k) \to \mathbf{x}$, $\|\mathbf{x}\| = 1$, and $\mathbf{x}^T\mathbf{P}\mathbf{A}\mathbf{P}\mathbf{x} = 0$. But $\mathbf{Q}$ is positive definite and $\mathbf{Q} = \mathbf{P}\mathbf{A}\mathbf{P}$. Hence $\tilde{\mathbf{P}}_k$ cannot converge for the transformed equation, and by eqn (63), $\mathbf{P}_k$ cannot converge in the original Riccati equation. ■

The result of Theorem 13 is confirmed by the flow networks studied earlier [2]: The Riccati equation failed to converge for every unobservable system examined. Moreover, the divergence was always to $+\infty$, that is, some diagonal elements increased without bound' toward $+\infty$. We shall now show that for certain classes of systems this type of divergence always takes place.

Theorem 14. If a linear QSS system can be transformed so that

$$\mathbf{H} = [\mathbf{H}_1 \vdots \mathbf{0}] \tag{67}$$

$$\mathbf{P}_0(-) = \mathbf{P}_0 = \left[\begin{array}{c|c} \mathbf{P}_{10} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}_{20} \end{array}\right] \tag{68}$$

$$\mathbf{Q} = \left[\begin{array}{c|c} \mathbf{Q}_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{Q}_2 \end{array}\right] \tag{69}$$

then the diagonal elements in $\mathbf{P}_{2k}$ increase without bound as $k \to \infty$ In addition, if $\mathbf{H}_1$ is of full rank, the sequence $\{\mathbf{P}_{1k}\}$ converges.

Proof. If we substitute eqns (67)-(69) into eqns (59) and (60), we obtain the following decomposition:

$$\mathbf{P}_{1,k+1} = \mathbf{P}_{1k} + \mathbf{Q}_1 - \mathbf{P}_{1k}\mathbf{A}_{1k}\,\mathbf{P}_{1k} \tag{70}$$

$$\mathbf{P}_{2,k+1} = \mathbf{P}_{2k} + \mathbf{Q}_2 \tag{71}$$

where $\mathbf{A}_{1k}$ is given by eqn (66).

If $\mathbf{H}_1$ is of full rank and $\mathbf{Q}_1 > 0$, then by Theorem 12 $\{\mathbf{P}_{1k}\}$ will converge since it corresponds to the Riccati equation for the observable QSS system with variables $\mathbf{x}^1$ and measurements $\mathbf{z}(k) = \mathbf{H}_1\mathbf{x}^1(k) + \mathbf{v}(k)$.

$\{\mathbf{P}_{2k}\}$, on the other hand, will increase without bound since $\mathbf{Q}_2$ is added at each time step according to eqn (71). ■

As an application of the above theorem, it can be shown[17] that the Kalman filter for a linear QSS system satisfying eqns (67)-(69) yields the estimates,

$$\hat{\mathbf{x}}(k) = \begin{bmatrix} \hat{\mathbf{x}}^1(k-1) \\ \hat{\mathbf{x}}^2(k-1) \end{bmatrix} + \begin{bmatrix} \mathbf{P}_{1k}(+)\mathbf{H}_1{}^T\mathbf{R}^{-1} \\ \mathbf{0} \end{bmatrix}\left[\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}^1(k-1)\right] \tag{72}$$

Equation (72) shows that $\hat{\mathbf{x}}^2(k) = \hat{\mathbf{x}}^2(0)$ for all k, unaffected by the measurements and decoupled from the estimates of $\hat{\mathbf{x}}^1(k)$.

REDUNDANCY

Closely related to observability is the concept of redundancy. A measurement is redundant if its removal causes no loss of observability. Definitions of local

redundancy, etc. follow directly from the type of observability being considered. Also, we say that an unmeasured variable is "barely observable" if it is observable but a nonredundant measurement must be used to calculate that variable. Thus, if an unmeasured variable is barely observable, failure of some instrument will render that variable unobservable. Since redundancy properties are defined in terms of observability properties, it is clear that in linear systems, all properties are global.

Redundancy is also closely related to estimator performance, but the effects may depend a great deal on the form of the estimator. If a measurement is nonredundant, a typical unbiased statistical estimator will take the measurement as the estimate of the variable, and use it directly to calculate other variables depending on it.

When measurements are redundant, constrained least-squares estimation can be used, and thus redundancy can be used to reduce the effects of measurement error. Also, redundancy is useful as a safety feature. When an instrument fails, redundancy may be utilized to fill in missing measurement values.

We shall now extend the results of Theorem 9 to the decomposition of a linear steady state system with redundant measurements.

Theorem 15. If a linear steady state system in modified standard form, eqns (12) and (27), is observable and $p + l > n$, and furthermore, if the rows of $\mathbf{H}$ are permuted so that the first $(l - j)$ rows correspond to the redundant measurements and the last j rows correspond to the nonredundant measurements, i.e. $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}$, and $j > 0$, then there exists a nonsingular $n \times n$ matrix $\mathbf{T}$ such that

$$\tilde{\mathbf{G}} = \mathbf{GT} = \begin{matrix} p \\ \end{matrix} \overset{\begin{matrix} n-j & j \end{matrix}}{\left[\begin{array}{c|c} \mathbf{G}_1 & \mathbf{0} \end{array}\right]} \tag{73}$$

$$\tilde{\mathbf{H}} = \mathbf{HT} = \begin{matrix} l-j \\ j \end{matrix} \overset{\begin{matrix} n-j & j \end{matrix}}{\left[\begin{array}{c|c} \mathbf{H}_{11} & \mathbf{0} \\ \hline \mathbf{H}_{21} & \mathbf{H}_{22} \end{array}\right]} \tag{74}$$

Rank $[\mathbf{H}_{22}] = j$, rank $\begin{bmatrix} \mathbf{G}_1 \\ \mathbf{H}_{11} \end{bmatrix} = n - j$ and every measurement corresponding to the subsystem $\begin{bmatrix} \mathbf{G}_1 \\ \mathbf{H}_{11} \end{bmatrix}$ is redundant.

Proof. Let $\mathbf{u}_i$, $i = 1,2, \ldots, n - j$ be a basis for $[\mathcal{N}(\mathbf{G}) \cap \mathcal{N}(\mathbf{H}_{1)})]^{\perp}$ and $\mathbf{u}_i$, $i = n-j+1, \ldots, n$ be a basis for $\mathcal{N}(\mathbf{G}) \cap \mathcal{N}(\mathbf{H}_{1)})$. Clearly $\mathcal{N}(\mathbf{G}) \cap \mathcal{N}(\mathbf{H}_{1)})$ must be nonempty, for it is the set of feasible perturbations undetectable by redundant measurements $\mathbf{z}^1$. If there were no such perturbations, the system would be observable just using measurements $\mathbf{z}^1$ (and the constraints), which contradicts the hypothesis $j > 0$.

Let $\mathbf{T} = [\mathbf{u}_1 \vdots \mathbf{u}_2 \vdots \ \ldots \ \vdots \mathbf{u}_n]$. Then since the system is observable, by Theorem 4,

$$\operatorname{rank} \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix} = \mathrm{n} \tag{75}$$

Hence,

$$n = \operatorname{rank} \begin{bmatrix} \mathbf{G} & \mathbf{T} \\ \mathbf{H} & \mathbf{T} \end{bmatrix} = \operatorname{rank} \begin{matrix} p \\ l-j \\ j \end{matrix} \overset{\begin{matrix} n-j & j \end{matrix}}{\begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{H}_{11} & \mathbf{0} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}}. \tag{76}$$

Since the last matrix in eqn (76) is of full rank, all its columns must be linearly independent, and hence, rank$[\mathbf{H}_{22}] = j$. Since rank $[\mathbf{H}_{21} \vdots \mathbf{H}_{22}] = j$, rank $\begin{bmatrix} \mathbf{G}_1 \\ \mathbf{H}_{11} \end{bmatrix} = n - j$.

Finally, we must show that each measurement $\mathbf{z}^1$ is redundant in the system $\mathbf{G}_1 \tilde{\mathbf{x}}^1 + \mathbf{a} = \mathbf{0}$, $\mathbf{z}^1 = \mathbf{H}^{11} \tilde{\mathbf{x}}^1 + \mathbf{c}^1$. Since the measurements $\mathbf{z}^1$ were redundant in the original system, each row of $\mathbf{H}_1$ was linearly dependent on some other rows of $\begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix}$. However, any row in $\mathbf{H}_1$ must be linearly independent of any row in $\mathbf{H}_2$, for otherwise measurement $\mathbf{z}^2$ would be redundant. Hence, any row of $\mathbf{H}_1$ is dependent on other rows in $\begin{bmatrix} \mathbf{G} \\ \mathbf{H}_1 \end{bmatrix}$. This dependency is unchanged by the transformation $\mathbf{T}$, and hence any row in $\mathbf{H}_{11}$ is dependent on other rows of $\mathbf{G}_1$ and/or $\mathbf{H}_{11}$ and hence is redundant in this system also. ■

Theorem 15 clearly shows that there is a constrained redundant subsystem of dimension $n - j$ and an unconstrained nonredundant subsystem of dimension j. Furthermore, the decomposition permits $\tilde{\mathbf{x}}^1$ to be estimated first, and then $\tilde{\mathbf{x}}^2$ can be calculated using our knowledge of $\mathbf{z}^2$ and $\tilde{\mathbf{x}}^1$. A special case of Theorem 15 involving material flow networks was derived previously [1] using graph-theoretic techniques.

CLOSING REMARKS

In this paper we developed the concepts of observability and redundancy for constrained steady state systems. We showed how local observability is directly related to local uniqueness of the solution to the measurement equation and how local unobservability leads to estimator failure for both steady state and quasi-steady state estimation. For linear constraints and measurements the conditions for local observability also hold for global observability and the system is decomposable into subsystems which are redundant, (barely) observable or unobservable, respectively.

With these results the stage is now set for the development of algorithms for observability and redundancy classification which will be the subject of the sequel to this paper.

Acknowledgement—This work was supported by the National Science Foundation Grants GK-43430 and ENG 76-18852.

NOTATION

- $\mathbf{A}_k$ a symmetric matrix defined by eqn (60)
- $\mathbf{A}$ a general vector of constants in eqn (12)
- $\{a_k\}$ a sequence of scalars
- $\mathbf{B}_j$ Hessian matrix for constraint g_j
- $\mathbf{b}$ an arbitrary vector
- $\mathbf{c}$ a general vector of constants in eqn (10)

D Matrix defined in the paragraph immediately preceding Theorem 4
d a direction vector
$E(.)$ expected value
F mass flow rate
G a constraint matrix
g a vector of constraint functions
$\mathbf{g}'(\mathbf{x}^0)$ the Jacobian matrix for the constraints defined in eqn (11) evaluated at $\mathbf{x}^0$
H enthalpy
H measurement matrix
H* a reduced measurement matrix defined by eqn (47)
h a vector of measurement functions
$\mathbf{h}'(\mathbf{x}^0)$ the Jacobian matrix of measurements calculated at $\mathbf{x}^0$
I identity matrix
K a convex constrained set defined by eqn (9)
K Kalman filter gain matrix
$\mathbf{K}_k$ gain matrix for Kalman filter at time t_k
k an integer variable
l number of measurements
$N(\bar{\mathbf{x}}, \mathbf{P}_0)$ normal distribution with expected value $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{P}_0$
n number of state variables in a steady state system or in a Kalman filter
P error covariance matrix
$\mathbf{P}_k(+)$ error covariance matrix immediately after a discrete measurement at time t_k
$\mathbf{P}_k(-)$ error covariance matrix immediately before a discrete measurement at time t_k
p number of steady state equations
Q process noise covariance matrix
R measurement noise covariance matrix
R^n n-dimensional real coordinate space
S a feasible set
$\bar{S}$ closure of set S, consisting of S and all its limit points
s a scalar multiplier in eqn (22)
T temperature
T in eqn (8) denotes a matrix referred by Yoshikawa and Bhattacharyya[10], but in Theorems 9 and 15 it represents a general nonsingular matrix
t time in eqn (16) and dummy variable in eqn (22)
t_k time at which the kth measurement is taken
$\mathbf{u}_i$ basis vector spanning the column space of T in Theorem 15
V set from which a particular value of measurement noise may be obtained
$\mathbf{v}(k)$ measurement noise vector at time t_k
$\bar{\mathbf{v}}$ expected value of $\mathbf{v}$
W linear estimator matrix in eqn (28)-(31) and $\mathbf{W}^{-1}$ is an approximation for the covariance matrix of $\mathbf{x}_0$ in eqn (38) and (40)
$\boldsymbol{w}(k)$ process noise vector at time t_k
x vector of state variables
$\mathbf{x}^0$ limiting point of the sequence $\{\mathbf{x}^k\}$
$\hat{\mathbf{x}}$ estimate of **x**
$\mathbf{x}_i$ ith element of vector **x**
$\mathbf{x}_I$ vector with elements x_i, $i \in I$, where I is an index set
$\{\mathbf{x}^k\}$ the sequence of vector $\mathbf{x}^k$
x_a slack variable defined in eqn (17)
Y estimator in eqn (24)
y augmented state variables defined in eqn (13) and a general vector in the proof of Theorem 5
z measurement vector

Greek symbols

$\boldsymbol{\Gamma}$ n x (n - p) matrix defined in eqn (46)
ω objective function in the least square estimation defined in eqn (32)

Subscripts

$_0$ prior estimate

Superscripts

$\sim$ corresponding variables and matrices after the coordinate transformation given in eqn(54)
$\hat{}$ estimate of
^{T} Transpose of a matrix
$'$ first derivative with respect to x

Operator

$\circ$ composite function operator
$^\perp$ orthogonal complement
$\mathcal{N}(\,)$ null space
$>$ $\mathbf{P} > 0$ denotes a positive definite matrix **P**

REFERENCES

[1] Mah R. S. H., Stanley G. M. and Downing D. M., *Ind. Engng Chem. Proc. Des. Dev.* 1976 **15** 1975.
[2] Stanley G. M. and Mah R. S. H., *A.I.Ch.E.J.* 1977 **23** 642.
[3] Kalman R. E., *Proc. 1st Int. Cong. of IFAC, Moscow* 1960 **1** 481, Butterworth, London 1961.
[4] Kalman R. E., *SIAM J. Contr.* 1963 **1** 152.
[5] Kostyukovskii Yu. M. L., *Auto. Rem. Contr.* 1968 **29** 1384.
[6] Kostyukovskii Yu. M. L., *Auto. Rem. Contr.* 1968 **29** 1575.
[7] Griffith E. W. and Kumar K. S. P., *J. Math. Anal. Appl.* 1971 **35** 135.
[8] Kou S. R., Elliott D. L. and Tarn T. J., *Inf. Contr.* 1973 **22** 89.
[9] Singh S. N., *Int. J. Syst. Sci.* 1975 **6** 723.
[10] Yoshikawa T. and Bhattacharyya S. P., *IEEE Trans. Automat. Contr.* 1975 **AC-20** 713.
[11] Gelb A. (Ed.), *Applied Optimal Estimation.* M.I.T. Press, Cambridge, Mass. 1974.
[12] Lin C-T., *IEEE Trans. Automat. Contr.* 1974 **AC-19** 201.
[13] Glover K. and Silverman L. M., *IEEE Trans. Automat. Contr.* 1976 **AC-21** 534.
[14] Shields R. W. and Pearson J. B., *IEEE Trans. Automat. Contr.* 1976 **AC-21** 203.
[15] Seber G. A. F., *The Linear Hypothesis: A General Theory.* Griffin, London 1966.
[16] Lang S., *Analysis I.* Addison-Wesley, Reading, Mass. 1968.
[17] Deutsch R., *Estimation Theory.* Prentice-Hall, Englewood Cliffs, New Jersey 1973.
[18] Rhodes I. B., *IEEE Trans. Automat. Contr.*1971 **AC-16** 688.
[19] Kuehn D. R. and Davidson H., *Chem. Engng Prog.* 1961 **44** 57.
[20] Stanley G. M., Ph. D. Thesis, Northwestern University, Evanston 1977.
[21] Wang S-H., Davison E. J. and Dorato P., *IEEE Trans. Automat. Contr.* 1975 **AC-20** 716.
[22] Schweppe F. C., *Uncertain Dynamic Systems.* Prentice-Hall, Englewood Cliffs, New Jersey 1973.

APPENDIX

Simple examples of process systems

As a simple example to illustrate the notation (S, **h,** V) let us consider the blender in Fig. 5(a) with two inlet streams 1 and 2 and an outlet stream 3, all streams being single-phase. Let the temperatures T_1, T_2 and T_3 and the outlet flow rate F_3, be measured, that is, $h_1 = F_3$, $h_2 = T_1$, $h_3 = T_2$, $h_4 = T_3$. Then the state variables may be $\mathbf{x} = (F_1, F_2, F_3, H_1, H_2, H_3)$ where H_i is the enthalpy of stream i. The measurement vector $\mathbf{z}$ is given by

$$z_1 = F_3 + v_1,$$
$$z_2 = T_1(H_1) + v_2,$$
$$z_3 = T_2(H_2) + v_3,$$
$$z_4 = T_3(H_3) + v_4$$

where $v_i \sim N(0,\sigma_i)$ if the noise is Gaussian. The feasible set S is defined by the constraints:

$$-x_1 - x_2 + x_3 = 0,$$
$$-x_1x_4 - x_2x_5 + x_3x_6 = 0$$
$$x_{i\min} \le x_i \le x_{imax}, \quad i = 1,2, \ldots, 6.$$

In such a system one may want to control the temperature or the concentration of some component in stream 3, using estimates of flows F_1 and F_2, in a cascade control scheme, and the question is whether F_1 and F_2 may be determined using any estimator.

Similarly, Fig. 5(b) may represent a two-stream heat exchanger for which all inlet and outlet temperatures, T_1, T_2, T_3, T_4, and one stream flow rate F_3 are measured. The state variables are again the flow rate and enthalpy of all streams, and the feasible set is delineated by the material and energy conservation and the upper and lower bounds on each variable. If the heat exchanger is a feed preheater to a distillation column, we may want to monitor the feed flow rate, and the question is whether the flow rate F_1 or F_2 may be determined using any estimator.

For both of these two problems the answers turn out to be "no" even for perfect measurements if $H_1 = H_2$; and by our definition, the flows are locally unobservable under those conditions. But it must not be supposed that local unobservability occurs only on restricted sets such as $H_1 = H_2$. Figure 5(c) depicts a common situation for which a wide range of flows might be expected and a single meter cannot cover the full range accurately. In this case $\mathbf{x} = (F_1, F_2)$, $x_1 - x_2 = 0$ and, say,

$$h_1 = \begin{cases} 0, & x_1 < 0 \\ x_1, & 0 \le x_1 \le 1 \\ 1, & x_1 > 1 \end{cases}$$

$$h_2 = \begin{cases} 0, & x_1 < 0 \\ x_2, & 0 \le x_2 \le 3 \\ 3, & x_2 > 3 \end{cases}$$

Then x_1 and x_2 are locally observable for $0 < x_2 < 3$, but locally unobservable for $x_2 < 0$ or $x_2 > 3$. z_2 is locally redundant for $0 < x_2 < 1$ but locally non-redundant for $1 < x_2 < 3$. z_1 is globally redundant—its deletion will not cause any loss of observability. Note that in this example the observability "structure" changes as the numerical values change. Therefore the numerical aspect should not be considered "incidental" or "fortuitous" in characterizing the observability and redundancy of a system. Note also that the examples cited above are by no means isolated or pathological. Flow meters such as rotameters have non-zero lower limits as well as upper limits of scale, and configurations such as Fig. 5(a)-(c) occur widely as components of more complex systems such as the crude preheat exchanger networks discussed in our earlier paper [2].

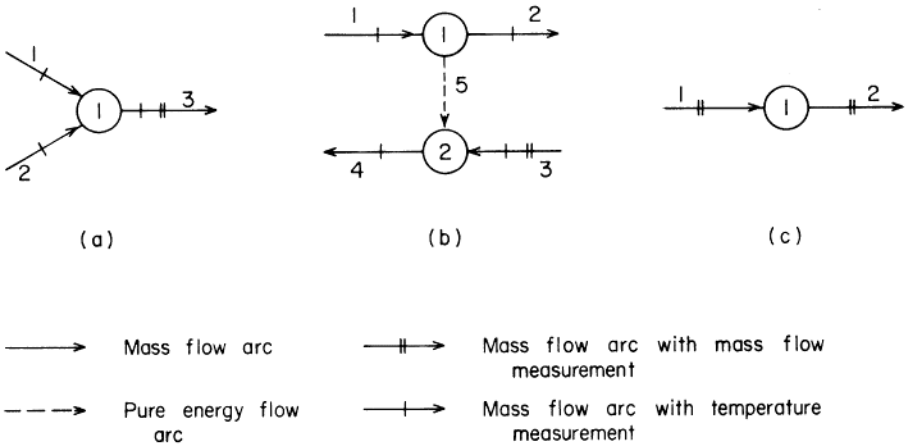

Fig. 5. Simple process flow networks.

* **To contact author**: see http://gregstanleyandassociates.com/contactinfo/contactinfo.htm