Observability of Linear Systems with Saturated Outputs

Renée Koplon[*]
Dept. of Mathematics
SYCON Center
Rutgers University
New Brunswick, NJ 08903

Eduardo D. Sontag[†]
Dept. of Mathematics
SYCON Center
Rutgers University
New Brunswick, NJ 08903

M.L.J. Hautus
Dept. of Mathematics
University of Technology
Eindhoven
The Netherlands

Abstract

In this paper, we present necessary and sufficient conditions for observability of the class of output-saturated systems. These are linear systems whose output passes through a saturation function before it can be measured.

1 Introduction

The question of observability for time-invariant linear systems is certainly a well understood problem. But what happens when the output is not fully available? That is, instead of measuring Cx, we can only measure $\boldsymbol{\sigma}(Cx)$, where $\boldsymbol{\sigma}$ is some nonlinear function. If the nonlinearity $\boldsymbol{\sigma}$ is not injective, it is no longer obvious from the observability matrix $[C' \, A'C' \, \cdots \, (A^{n-1})'C']'$ (prime indicates transpose), whether or not the state can be "observed" from the output.

In [3], we answered this question in the case in which $\boldsymbol{\sigma}$ provided the sign of the output of the linear system. That model was motivated by quantization and pattern recognition. In this paper, we will look at continuous-time sytems in which the function $\boldsymbol{\sigma}$ is the identity near the origin, but saturates the output values away from 0. (A preliminary version of this was presented at the American Automatic Control Conference, June 1992 [5].)

By an *output-saturated system*, we mean a continuous-time system

$$\begin{aligned} \Sigma : \ \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= \boldsymbol{\sigma}(Cx(t)) \end{aligned}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $\boldsymbol{\sigma}$ defined by applying the function

$$\sigma(x) = \begin{cases} -1 & x < -1 \\ x & -1 \le x \le 1 \\ 1 & 1 < x \end{cases}$$

[*] Research supported by AFOSR under grant AFOSR-91-0343

[†] Research supported by AFOSR under grant AFOSR-91-0343

to each component of the output vector. We denote the above output-saturated system by $\Sigma = (A, B, C)_{\mathrm{os}}$. This is an effective model for sensor saturation or overflow in the measurement device. A system is *observable* if given any two initial states, there is a control which provides distinguishing outputs for those two initial states. For linear systems, this definition of observability coincides with saying that for any two initial states the outputs are different, using no controls. However, for output-saturated systems, if the outputs remain large for two distinct states, they may look the same. So it may be necessary to use the control to move the output into the "linear window". It is easy to see that the boundaries of this window are not really relevant for the question of observability. We merely choose $[-1, 1]$ for convenience. In fact, $\boldsymbol{\sigma}$ could be a completely different function in each coordinate, as long as it is one-to-one in a neighborhood of 0 and saturated away from zero.

After introducing some technical definitions and lemmas in Section 2, we proceed in Section 3 to prove a characterization for observability of continuous-time output-saturated systems. Since the conditions are not necessarily easy to check, Section 4 presents some necessary and some sufficient conditions for observability depending only on certain eigenvalues of A. The two sections 5 and 6 focus separately on the cases of one, two, and more than two outputs. Finally, in Section 7 we study output-saturated systems with the added restriction of bounded inputs. We provide a characterization of observability for the class of bounded-input output-saturated systems for which the pair (A, B) is stabilizable.

2 Preliminaries

First we give Bohr's definition of almost periodicity (see e.g. [1]) then prove some technical lemmas.

Definition 2.1 A function $f : \mathbb{R} \to \mathbb{R}$ is *almost periodic* if for any $\varepsilon > 0$, there exists a number $\ell > 0$ such that for every open interval of length ℓ there exists a number τ contained in that interval such that

$$|f(t+\tau) - f(t)| < \varepsilon, \ \forall t \in \mathbb{R}.$$

This number τ, which depends on ε, is called an *ε-almost period.*

Lemma 2.2 If $f : \mathbb{R} \to \mathbb{R}$ is almost periodic, and $t_0 \in \mathbb{R}$ is arbitrary,

$$\begin{aligned} \limsup_{t\to\infty} f(t) &= \sup_{t>t_0} f(t), \\ \liminf_{t\to\infty} f(t) &= \inf_{t>t_0} f(t). \end{aligned}$$

Proof. Pick an arbitrary $T_0 > t_0$ and $\varepsilon > 0$. The function f is almost periodic, so there exists an ℓ such that in every interval of length ℓ, there is an ε-almost period, τ. Suppose we are given a $T > 0$. Then there exists a τ in the interval $(T - T_0, T - T_0 + \ell)$ so that $|f(T_0 + \tau) - f(T_0)| < \varepsilon$. Let $t_1 := T_0 + \tau > T$. As

$$f(T_0) - f(t_1) < \varepsilon,$$

we have proved that for all $T > 0$ there exists a number $t_1 > T$ such that $f(t_1) > f(T_0) - \varepsilon$ and ε was arbitrary. This implies that $\limsup f(t) \geq f(T_0)$. This is true for all $T_0 > t_0$, so $\limsup_{\ t\to\infty} f(t) \geq \sup_{t>t_0} f(t)$. Also $\sup_{t>t_0} f(t) \geq \limsup_{\ t\to\infty} f(t)$, so we conclude that

$$\limsup_{t\to\infty} f(t) = \sup_{t>t_0} f(t).$$

The second statement follows analogously. ∎

In particular, $\limsup_{t\to\infty} f(t) = \sup_{t\in\mathbb{R}} f(t)$ and $\liminf_{t\to\infty} f(t) = \inf_{t\in\mathbb{R}} f(t)$ for almost periodic functions, so:

Lemma 2.3 If $f : \mathbb{R} \to \mathbb{R}$ is almost periodic, and $\lim_{\ t\to\infty} f(t) = 0$, then $f(t) \equiv 0$. ∎

A function $f(t)$ is called a *Bohl function* if it is a finite linear combination of functions of the form $t^\ell e^{\lambda t}$, with $\ell \in \mathbb{N}$ and $\lambda \in \mathbb{C}$, or equivalently, if the Laplace transform $\hat{f}(s)$ of $f(t)$ is rational. The poles of this rational function $\hat{f}(s)$ are called the *exponents* of the Bohl function $f(t)$ and the order of a pole is called the *index* of the corresponding exponent.

A pair $(\lambda, i) \in S \subset \mathbb{C} \times \mathbb{N}$ is *maximal* for the set S if $\text{Re}(\lambda) \geq \text{Re}(\mu)$ for all $(\mu, j) \in S$, and $i \geq j$ for all other (μ, j) with $\text{Re}(\lambda) = \text{Re}(\mu)$. We say that (λ, i) is an *(exponent, index) pair* for f if λ is an exponent of f of index i. We define $\mathcal{E}(f)$ as the set of exponents, λ, of f for which (λ, i_λ) is maximal among all (exponent, index) pairs. We will call $\mathcal{E}(f)$ the set of *dominating* exponents of f.

Recall that an eigenvalue λ of a constant matrix A has *index* k if k is the size of the largest Jordan block of A corresponding to λ, i.e., k is the multiplicity of λ as a root of the minimal polynomial of A. We define the dominating eigenvalues of A, $\mathcal{E}(A)$, to be the set of eigenvalues λ of A for which the (eigenvalue, index) pair (λ, i_λ) is maximal among all (eigenvalue, index) pairs associated to A. Notice that if $p = 1$ and (A, C) is an observable pair, then A must be cyclic, so in this case, the index of λ is equal to the multiplicity of λ.

For a matrix of Bohl functions, $W(t)$, we find a minimal realization $(A_{\min}, B_{\min}, C_{\min})$ associated to $W(t)$, and then we define the dominating exponents for $W(t)$ to be

$$\mathcal{E}(W(t)) := \mathcal{E}(A_{\min}).$$

This definition does not depend on the particular realization used.

Remark 2.4 Let $W(t) = (w_1(t), \ldots, w_m(t))$ be a row of Bohl functions. We say that the pair (λ, i) is an (exponent, index) pair for $W(t)$ if there is a j so that (λ, i) is an (exponent, index) pair for $w_j(t)$. Then it follows that $\mathcal{E}(W(t))$ is the set of λ so that (λ, i) is maximal among all (exponent, index) pairs of $W(t)$.

Lemma 2.5 Let $f(t)$ be a Bohl function. Denote $\mathbb{C}_+ := \{s \in \mathbb{C} : \text{Re}(s) \geq 0\}$. If $\mathcal{E}(f) \subseteq \mathbb{C}_+ \setminus \mathbb{R}_+$, then there exists a sequence $\{t_k\}$ with $t_k \to \infty$ so that $f(t_k) = 0$.

Proof. Let α be the common real part of the dominating exponents, $\mathcal{E}(f)$. We can write

$$\begin{aligned} f(t) &= \sum_j a_j t^{\ell_j} e^{\alpha_j t}\{p_j \sin(\omega_j t) + q_j \cos(\omega_j t)\} \\ &= t^\ell e^{\alpha t} \sum_j a_j t^{\ell_j - \ell} e^{(\alpha_j - \alpha)t}\{p_j \sin(\omega_j t) + q_j \cos(\omega_j t)\}, \end{aligned}$$

where $\ell_j - \ell \leq 0$ and $\alpha_j - \alpha \leq 0$ for all j. Now let $g(t)$ be the sum of all the terms in which $\ell_j - \ell = 0$ and $\alpha_j - \alpha = 0$. Let $h(t)$ be the sum of the rest of the terms. That is, we may write $f(t) = t^\ell e^{\alpha t}\,(g(t) + h(t))$ where

$$\begin{aligned} g(t) &= \sum_j a_j\{p_j \sin(\omega_j t) + q_j \cos(\omega_j t)\} && (1) \\ h(t) &= \sum_j a_j t^{-m_j} e^{-n_j t}\{p_j \sin(\omega_j t) + q_j \cos(\omega_j t)\}. && (2) \end{aligned}$$

We are assuming that f has a complex exponent with nonnegative real part equal to α. Thus, g is not identically zero and the ω_j's are all nonzero. Any real exponents equal to α do not have maximal index so they will correspond to terms in h of the form $a_j q_j t^{-m_j}$. Note also that $m_j, n_j \geq 0$ and for each j, at least one of m_j, n_j is nonzero, so $\lim_{t\to\infty} h(t) = 0$. Since g is a continuous function of t, if we prove that $\limsup_{t\to\infty} g(t) > 0$ and $\liminf_{t\to\infty} g(t) < 0$, then together with $h(t) \to 0$, it follows that there must exist a sequence t_k with $t_k \to \infty$ as $k \to \infty$, such that $g(t_k) = -h(t_k)$. Thus $g + h$ has infinitely many zeros, and so f does too.

Since g is a linear combination of periodic functions, it is itself an almost periodic function. (See [1], Paragraph 48.) Thus we may apply Lemma 2.2 to the function g. Now suppose that $\liminf g(t) = \inf g(t) \geq 0$. Then the function G defined by $G(t) := \int_0^t g(\tau)d\tau$ is nondecreasing. By term by term integration it is easily seen that G is almost periodic and bounded ($\omega_j \neq 0$ for all j, so $G(t)$ looks again like the formula (1), but with different constants). This would imply tha G is convergent. In that case, G must be constant (Lemma 2.3), from which it follows that g is identically zero, a contradiction. So $\liminf_{\,t\to\infty} g(t) < 0$. Similarly, we may obtain that $\limsup_{\,t\to\infty} g(t) > 0$. ∎

Let $(\tilde{A}, \tilde{C})$ be the observable pair of submatrices in the Kalman observability decomposition for the pair (A, C) (see [7], Section 5.2):

$$\begin{pmatrix} \tilde{A} & 0 \\ A_1 & A_2 \end{pmatrix} \begin{pmatrix} \tilde{C} & 0 \end{pmatrix}. \qquad (3)$$

Let $T \in \mathrm{Gl}(n)$ be the matrix providing the coordinate transformation.

Proposition 2.6 Let $A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{1\times n}$. The following statements are equivalent:

1. For all $x \in \mathbb{R}^n$, $\inf_{t>0} |Ce^{tA}x| = 0$.

2. The function $\mathbb{C} \to \mathbb{C}^n : s \mapsto C(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$.

3. The matrix $\tilde{A}$, in the Kalman observability decomposition of (A, C), has no eigenvalues on $\mathbb{R}_+$.

Proof. Note that since

$$\begin{aligned} Ce^{tA}x &= \begin{pmatrix} \tilde{C}e^{t\tilde{A}} & 0 \end{pmatrix} T^{-1}x, \text{ and} \\ C(sI-A)^{-1}x &= \begin{pmatrix} \tilde{C}(sI-\tilde{A})^{-1} & 0 \end{pmatrix} T^{-1}x \end{aligned}$$

for each $x \in \mathbb{R}^n$, it suffices to prove the result for observable pairs (A, C). Thus, for the remainder of this proof, we will assume (A, C) is an observable pair.

We first prove that $C(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$ if and only if A has no eigenvalues on $\mathbb{R}_+$. Consider $C(sI - A)^{-1} = C(sI - A)^{-1}I$ as a $1 \times n$ transfer matrix. As the triple (A, I, C) is canonical (controllable and observable), the eigenvalues of A are precisely the poles of this transfer matrix (see Corollary 5.7.2 in [7]). Thus $C(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$ if and only if A has no eigenvalues on $\mathbb{R}_+$, and conditions 2 and 3 are equivalent.

To prove that condition 1 implies condition 3, suppose A has an eigenvalue $\lambda \in \mathbb{R}_+$. Let v be a corresponding eigenvector. Then

$$\inf_{t>0} |Ce^{tA}v| = \inf_{t>0} e^{\lambda t}|Cv| = |Cv| > 0,$$

because (A, C) is observable, contradicting statement 1.

Next we prove that the third condition implies the first. If A has no eigenvalues on $\mathbb{R}_+$, then $f(t) = Ce^{tA}x$ is a Bohl function with no exponents on $\mathbb{R}_+$, so, in particular, $\mathcal{E}(f) \bigcap \mathbb{R}_+ = \emptyset$. If $\mathcal{E}(f) \subseteq \mathbb{C} \setminus \mathbb{C}_+$, we have $f(t) \to 0$ as $t \to \infty$ and so $\inf_{t>0} |f(t)| = 0$. If $\mathcal{E}(f) \subseteq \mathbb{C}_+ \setminus \mathbb{R}_+$, apply Lemma 2.5. Then there is a $t > 0$ so that $f(t) = 0$, so clearly $\inf_{t>0} |f(t)| = 0$. ∎

3 Property Q

In this section we introduce a property which, together with observability of the pair (A, C), will serve to characterize output-saturated observability. This property will be used repeatedly in later sections.

Let K be any subset of $\{1, \ldots, p\}$ and define N_K by

$$N_K := \bigcap_{i \notin K} \mathcal{O}(A, C_i)$$

where $\mathcal{O}(A, C_i) = \bigcap_{k=0,\ldots,n-1} \ker(C_i A^k)$. If $K = \{1, \ldots, p\}$, then $N_K = \mathbb{R}^n$, the whole state space. If $K = \emptyset$, then $N_K = \mathcal{O}(A, C)$, thus $N_\emptyset = \{0\}$ if the pair (A, C) is observable. The states in N_K are those that cannot be distinguished from 0 for the linear system (A, B, C) using outputs not in K. Note that N_K is an A-invariant subspace since each $\mathcal{O}(A, C_i)$ is. So we may define A_{N_K} to be the operator A restricted to the subspace N_K.

For an output-saturated system $\Sigma = (A, B, C)_{\mathrm{os}}$, the sequence of $p \times m$ matrices

$$\mathcal{A} := \{CB, CAB, CA^2B, \ldots\}$$

is called the *Markov parameter sequence*. Let $I := I(\mathcal{A}) \subset \{1, \ldots, p\}$ be the indices of the nonzero rows of $\mathcal{A}$, $J := J(\mathcal{A}) \subset \{1, \ldots, p\}$ be the indices of the zero rows of $\mathcal{A}$ and define $N := N_J$.

For any index set $K \subseteq \{1, \ldots, p\}$, we let $Q(K)$ be the following property:

$$\begin{array}{c}\text{For all } \xi, \text{ and for all nonzero } v \in N_K, \text{ there exists a } j = j(\xi, v) \in K \text{ so that} \\ \sigma(C_j e^{tA}\xi) \not\equiv \sigma(C_j e^{tA}\xi + C_j e^{tA}v).\end{array} \tag{$Q(K)$}$$

(If $N_K = \{0\}$, $Q(K)$ automatically holds since there are no nonzero $v \in N_K$.) Using this notation, we let $\boldsymbol{Q} = Q(J)$ for $K = J$ as defined above, and provide a necessary and sufficient characterization for observability of continuous-time output-saturated systems.

Lemma 3.1 If $\Sigma = (A, B, C)_{\mathrm{os}}$ is a continuous-time output-saturated system, then Σ is observable if and only if

1. (A, C) is an observable pair, and
2. Property $\boldsymbol{Q}$ holds.

Proof. Necessity: Clearly (A, C) is an observable pair. For the second condition, assuming Σ is observable, we must show that property $\boldsymbol{Q}$ holds. If $N_J = \{0\}$, then there is nothing to prove. So assume $N_J \neq \{0\}$. Pick any ξ and any nonzero $v \in N_J$. Let $\eta := \xi + v$. Then $\eta - \xi \in N_J$. By definition of N_J,

$$C_i e^{tA}(\eta - \xi) \equiv 0 \quad \forall i \notin J. \tag{4}$$

Since Σ is observable, there must exist some $j \in \{1, \ldots, p\}$ and a control u so that, for some $t \geq 0$,

$$\sigma\left(C_j e^{tA}\xi + \int_0^t C_j e^{(s-t)A}Bu(s)ds\right) \neq \sigma\left(C_j e^{tA}\eta + \int_0^t C_j e^{(s-t)A}Bu(s)ds\right). \tag{5}$$

By (4), it cannot happen that such a $j \notin J$. Thus, there is a $j \in J$ which satisfies (5). Since $\mathcal{A}^j$ (the jth row of $\mathcal{A}$) $\equiv 0$ for $j \in J$, this implies that

$$\begin{aligned} \sigma(C_j e^{tA}\xi) &\neq \sigma(C_j e^{tA}\eta), \\ \text{i.e. } \sigma(C_j e^{tA}\xi) &\neq \sigma(C_j e^{tA}\xi + C_j e^{tA}v). \end{aligned}$$

Sufficiency: Given $\xi \neq \eta$, we must show that they can be distinguished. Let

$$\begin{aligned} \tilde{J} &:= \{j : C_j e^{tA}\xi \not\equiv C_j e^{tA}\eta\} \\ &= \{j : \xi - \eta \notin \mathcal{O}(A, C_j)\}. \end{aligned}$$

Note that the set $\tilde{J}$ is nonempty, by condition 1. If there is any $j \in \tilde{J}$ so that $\mathcal{A}^j \not\equiv 0$, then ξ, η can be distinguished with an argument as that used for sign-linear systems (see [4]). Otherwise, for all

$j \in \tilde{J}$, $\mathcal{A}^j \equiv 0$, so $\tilde{J} \subseteq J$. For each $i \notin \tilde{J}$, $v = \eta - \xi \in \mathcal{O}(A, C_i)$, so $v \in N_{\tilde{J}} \subseteq N_J$, $v \neq 0$ and $Q(J)$ can be applied. So there is a $j \in J$ and a t so that

$$\sigma(C_j e^{tA}\xi) \neq \sigma(C_j e^{tA}\eta),$$

and indeed ξ and η can be distinguished. (Note that when $J = \{1, \ldots, p\}$ then $N_J = \mathbb{R}^n$ in the above argument.) ∎

Of course property $\boldsymbol{Q}$ is not always an easy property to check, but we can use this property to simplify the proofs of the observability theorems of the next section. First we state some results about property $\boldsymbol{Q}$.

Remark 3.2 For any $K \subset \{1, \ldots, p\}$, if (A, C) is an observable pair and $v \in N_K$ is nonzero, then $C_j e^{tA} v \not\equiv 0$ for some $j \in K$. This is because, otherwise, $v \in \bigcap_{j \in K} \mathcal{O}(A, C_j)$. Together with $v \in N_K = \bigcap_{i \notin K} \mathcal{O}(A, C_i)$, this implies that $v \in \mathcal{O}(A, C) = \{0\}$, a contradiction.

Lemma 3.3 If (A, C) is an observable pair and for every $j \in J$, the row vector of functions $C_j(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$, then $\boldsymbol{Q}$ holds.

Proof. By Proposition 2.6, if $C_j(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$, then

$$\inf_{t>0} |C_j e^{tA} x| = 0, \text{ for all } x \in \mathbb{R}^n. \tag{6}$$

The assumptions of this Lemma then imply that for every $j \in J$, (6) holds. By Remark 3.2, for each nonzero $v \in N_J$, there is a $j \in J$ so that $C_j e^{tA} v \not\equiv 0$. Then for all $\xi \in \mathbb{R}^n$ and all nonzero $v \in N_J$, choose a $j \in J$ so that $C_j e^{tA} v \not\equiv 0$. For that j, pick a t so that $|C_j e^{tA}\xi| < 1/2$. Without loss of generality, we may assume $C_j e^{tA} v \neq 0$ for this t. Then

$$\sigma(C_j e^{tA}\xi) \neq \sigma(C_j e^{tA}\xi + C_j e^{tA} v),$$

and $\boldsymbol{Q}$ holds. ∎

Lemma 3.4 If $\boldsymbol{Q}$ holds, then for all nonzero v in N_J, there is a j in J so that $C_j e^{tA} v \not\equiv 0$ and $\inf_{t>0} |C_j e^{tA} v| = 0$.

Proof. Suppose not. Then there would exist a $v \in N_J$, $v \neq 0$ so that for all $j \in J$ either $C_j e^{tA} v \equiv 0$ or $\inf_{t>0} |C_j e^{tA} v| \neq 0$. Multiplying v by a scalar, we may assume that either $C_j e^{tA} v \equiv 0$ or $C_j e^{tA} v \geq 1$ for all $t > 0$. Let $\xi = v$. Then for all $j \in J$,

$$\sigma(C_j e^{tA}\xi) \equiv \sigma(C_j e^{tA}\xi + C_j e^{tA} v),$$

contradicting $\boldsymbol{Q}$. ∎

The next example shows that these necessary conditions are not sufficient.

Example 3.5 Let Σ be the output-saturated system defined by the following observable triple.

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \; C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Then $J = \{1\}$, $N_J = \{(x_1, x_2, x_3)' : x_1 = x_3 = 0\}$, and

$$Ce^{tA} = \begin{pmatrix} e^{2t} & e^{-t} & 0 \\ e^{2t} & 0 & e^{t} \end{pmatrix}.$$

For any $v = (0, a, 0)' \in N$, $a \neq 0$,

$$\begin{aligned} & C_1 e^{tA} v = ae^{-t} \not\equiv 0 \\ & \inf_{t>0} |ae^{-t}| = 0. \end{aligned}$$

So this system satisfies the conditions of Lemma 3.4. But Σ is not observable. For example, the states $\xi = (1, 0, 0)'$ and $\eta = (1, 1, 0)'$ are indistinguishable. Since $K_1(t) \equiv 0$, the first component of the output saturates at 1 for both initial states ξ and η. The second component of the output is exactly the same for both initial states.

4 Observability: Necessity and Sufficiency

Property $\boldsymbol{Q}$ is not an easy property to check. In this section we will give some conditions which can be checked by looking at the eigenvalues of the matrix A. For the general multiple output case, we have the following necessary conditions for observability. Recall that $\mathbb{R}_+$ denotes the nonnegative real axis $\{s \in \mathbb{R} : s \geq 0\}$.

Theorem 1 *If $\Sigma = (A, B, C)_{\text{os}}$ is a continuous-time observable system, then*

1. *(A, C) is an observable pair, and*

2. *A_N has no eigenvalues on $\mathbb{R}_+$.*

Proof. Let x be an eigenvector of A_N corresponding to an eigenvalue $\lambda \in \mathbb{R}_+$. Then $x \in N$, so $C_i e^{tA} x \equiv 0$ for all $i \notin J$. Let

$$J_0 := \{j \in J : C_j e^{tA} x \not\equiv 0\}.$$

Then for all $j \in J_0$, $\xi = x$, $\eta = 2x$, satisfy

$$\begin{aligned} C_j e^{tA} \xi &= e^{\lambda t} C_j x, \text{ and} && (7) \\ C_j e^{tA} \eta &= 2e^{\lambda t} C_j x. && (8) \end{aligned}$$

Clearly the signs of (7) and (8) are always the same, and x can be scaled so that both functions are outside the linear window $[-1, 1]$ for all t, contradicting observability. ∎

In the particular case in which $J = \{1, \ldots, p\}$, $N = \mathbb{R}^n$, so $A = A_N$. Thus:

Corollary 4.1 If $\Sigma = (A, B, C)_{\text{os}}$ is a continuous-time observable system and $\mathcal{A} \equiv 0$, then A has no eigenvalues on $\mathbb{R}_+$.

The above conditions are not in general sufficient for observability in the multiple output case.

Example 4.2 Let $\Sigma = (A, B, C)_{\text{os}}$ where

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

$y_1(t) = \sigma(x_1)$, and $y_2(t) = \sigma(x_2)$. We next check that Σ satisfies the necessary conditions of Theorem 1. The pair (A, C) is an observable pair. The second row of $\mathcal{A}$ is 0 so $J = \{2\}$ and $N = \{(x_1, x_2, x_3)' \in \mathbb{R}^3 : x_1 = 0, x_2 = -x_3\}$. The restriction of A to N is $A_N = -I$, which has no eigenvalues on $\mathbb{R}_+$. But this system is not observable. Indeed, the states $\xi = (1, 2, 1)'$ and $\eta = (1, 1, 2)'$ are indistinguishable. To see this, note that $K_2(t) \equiv 0$. The first component of the output is

$$y_1(t) = \sigma\left(4e^{2t} - 3e^t + \int_0^t K_1(t-s)u(s)ds\right),$$

for both ξ and η. The second component is $y_2(t) = \sigma(3e^t + e^{-t})$ for the initial state ξ and $y_2(t) = \sigma(3e^t + 2e^{-t})$ for η. In both cases, $y_2(t) \equiv 1$.

Remark 4.3 Let $M = \bigcap_{i \in J} \mathcal{O}(A, C_i)$ and let A^M be the operator induced by A on the quotient space $\mathbb{R}^n/M$. Then Lemma 3.3 is equivalent to saying that if (A, C) is an observable pair, and A^M has no eigenvalues on $\mathbb{R}_+$, then property $\boldsymbol{Q}$ holds.

To see this let

$$\begin{pmatrix} \tilde{A}_J & 0 \\ A_1 & A_2 \end{pmatrix} \begin{pmatrix} \tilde{C}_J & 0 \end{pmatrix}$$

be the Kalman decomposition for the pair (A, C_J) where C_J consists of just the rows of C indexed by J. Let r be the dimension of the observable component $\tilde{A}_J$. In this new basis, M is the subspace of states whose first r components are zero. Then $A_2 = A_M$ (the restriction of A to M) and $\tilde{A}_J$ is a matrix representation for A^M. Note that statements 2 and 3 in Proposition 2.6 are equivalent even in the case of arbitrary p. That is, A^M has no eigenvalues on $\mathbb{R}_+$ if and only if $C_J(sI - A)^{-1}$ has no poles on $\mathbb{R}_+$. □

The next Theorem, giving sufficient conditions for observability, follows directly from the preceeding Remark and Lemma 3.3.

Theorem 2 *The system $\Sigma = (A, B, C)_{\text{os}}$ is observable if*

1. *(A, C) is an observable pair, and*

2. *A^M has no eigenvalues on $\mathbb{R}_+$.*

Notice the subtle, but real difference between the conditions of Theorem 2 and those of Theorem 1. Recall $N = \bigcap_{i \in I} \mathcal{O}(A, C_i)$, so if we assume (A, C) is an observable pair, then $N \cap M = \{0\}$. Also, N and M are both A-invariant subspaces. Thus N can be naturally identified with $(N+M)/M$, so the eigenvalues of A_N are included among those of A^M. Equivalently, in matrix theoretic terms, there is a basis for $\mathbb{R}^n$ in which A has the form

$$\begin{pmatrix} A_M & 0 & * \\ 0 & A_N & * \\ 0 & 0 & * \end{pmatrix}.$$

Then

$$\begin{pmatrix} A_N & * \\ 0 & * \end{pmatrix}$$

is a matrix representation for A^M, from which it is obvious that

$$\sigma(A^M) \supseteq \sigma(A_N). \tag{9}$$

Observe that for the special case $p = 1$, the inclusion (9) is trivial. In fact, for $p = 1$, $A^M = A_N$. Indeed, if $I = \{1\}$ ($\mathcal{A} \not\equiv 0$), then $M = \mathbb{R}^n$ and $N = \{0\}$ so $A^M = A_N = 0$. If instead $J = \{1\}$ ($\mathcal{A} \equiv 0$), then $M = \{0\}$ and $N = \mathbb{R}^n$ so $A^M = A_N = A$.

Thus, this sufficient condition is stronger than the necessary conditions in Theorem 1 which stated that if Σ is observable, then (A, C) is an observable pair and $\sigma(A_N) \cap \mathbb{R}_+ = \emptyset$. The conditions are the same when $p = 1$. The following is an example of an observable output-saturated system which does not satisfy the stated sufficient conditions.

Example 4.4 Let Σ be an output-saturated system with associated triple

$$\begin{aligned} A &= \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & -1 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

(This triple is observable.) Then $J = \{1\}$, and the poles of $C_1(sI - A)^{-1}$ are: $2,\ 3+i,\ 3-i$. Thus, $C_1(sI-A)^{-1}$ has a pole on $\mathbb{R}_+$, so the sufficient conditions of Lemma 3.3 (and hence of Theorem 2) are not satisfied. However, property $\boldsymbol{Q}$ still holds. Note that $N_J = \{(x_1, x_2, x_3, x_4)' : x_1 = x_4 = 0\}$ and

$$Ce^{tA} = \begin{pmatrix} e^{2t} & e^{3t}\cos t & -e^{3t}\sin t & 0 \\ e^{2t} & 0 & 0 & e^t \end{pmatrix}.$$

Now take any $\xi = (a, b, c, d)' \in \mathbb{R}^4$, $v = (0, v_1, v_2, 0)' \in N$, $v \neq 0$. If $b \neq 0$ or $c \neq 0$, then

$$\begin{aligned} C_1 e^{tA}\xi &= ae^{2t} + be^{3t}\cos t - ce^{3t}\sin t \\ &= e^{3t}(ae^{-t} + b\cos t - c\sin t). \end{aligned}$$

In this case, applying Lemma 2.5, $\inf_{t>0}|C_1e^{tA}\xi| = 0$. Otherwise, if $b = c = 0$,

$$\begin{aligned} C_1e^{tA}(\xi + v) &= ae^{2t} + v_1e^{3t}\cos t - v_2e^{3t}\sin t \\ &= e^{3t}(ae^{-t} + v_1\cos t - v_2\sin t). \end{aligned}$$

Since $v \neq 0$, at least one of v_1, v_2 is not zero, so again applying Lemma 2.5, $\inf_{t>0}|C_1e^{tA}(\xi+v)| = 0$.

5 Corollaries: One and Two Outputs

5.1 The single output case

In the case of a single output system, $A^M = A_N$. In fact, if $\mathcal{A} \not\equiv 0$, then $A^M = A_N = 0$, and if $\mathcal{A} \equiv 0$, then $A^M = A_N = A$. Thus, Theorems 1 and 2 may be combined into the following result. Note that under the assumption that (A, C) is an observable pair, $\mathcal{A} \not\equiv 0$ and $B \neq 0$ are equivalent, so condition 1 is equivalent to $\mathcal{A} \not\equiv 0$.

Theorem 3 *Let $\Sigma = (A, B, C)_{\text{os}}$ be a single output continuous-time output-saturated system. Then Σ is observable if and only if (A, C) is an observable pair, and either*

1. *$B \neq 0$, or*

2. *A has no eigenvalues on $\mathbb{R}_+ := \{\lambda \in \mathbb{R}, \lambda \geq 0\}$.*

An easy corollary is the following.

Corollary 5.1 The single output system $\Sigma = (A, B, C)_{\text{os}}$ is observable if (A, C) is an observable pair and $\operatorname{rank}[sI - A, B] = n$ for all $s \in \mathbb{R}_+$.

Since stabilizability of the pair (A, B) is a particular case of the rank condition in the corollary, we see that observability plus stabilizability of the triple (A, B, C) is sufficient for observability of the output-saturated system $(A, B, C)_{\text{os}}$.

5.2 The case of two outputs

If J consists of only one element (that is, only one row of the Markov sequence is zero), we have necessary and sufficient conditions for $Q(J)$ to hold which depend only on certain eigenvalues of A. This theorem suffices to characterize observability for output-saturated systems with two outputs. Assuming (A, C) is an observable pair, which is a necessary condition for observability, there are three cases. Either both rows of the Markov sequence are nonzero, so that the system is automatically observable, both rows are zero, so that $\boldsymbol{Q}$ is equivalent to A having no eigenvalues on $\mathbb{R}_+$, or exactly one row of the Markov sequence is zero, in which case the following proposition applies.

We first give some definitions, as the results are more general than what we actually need. Let $N \subset \mathbb{R}^n$ be *any* A-invariant subspace, $N \neq \{0\}$, and $N \bigcap \mathcal{O}(A, C) = \{0\}$. That is, N is some subspace of states which are distinguishable from 0 for the linear system (A, B, C). Let $\tilde{A}$ be defined as in the Kalman observability decomposition for the pair (A, C) (see equation (3)). We denote the set of eigenvalues of the matrix $\tilde{A}$ by $\sigma(\tilde{A})$. For the purposes of the next proposition, we will understand $\boldsymbol{Q}$ to be the property that, with respect to such a space N,

$$\begin{gathered}\text{For all } \xi, \text{ and for all nonzero } v \in N, \\ \sigma(Ce^{tA}\xi) \not\equiv \sigma(Ce^{tA}\xi + Ce^{tA}v).\end{gathered}$$

Proposition 5.2 Assume $p = 1$. Let A_N be the restriction of A to N, and let

$$\tilde{\lambda} := \max\{\lambda \in \mathbb{R}_+, \lambda \in \sigma(\tilde{A})\}$$

($\tilde{\lambda} := -\infty$ if this set is empty). Property $\boldsymbol{Q}$ holds if and only if

1. $\sigma(A_N) \bigcap \mathbb{R}_+ = \emptyset$, and
2. $(\tilde{\lambda} I - A_N)$ is Hurwitz.

Proof. Assume $\boldsymbol{Q}$ holds. The necessity of condition 1 follows from Theorem 1. Next we prove that condition 2 holds. Suppose otherwise that $\tilde{A}$ has an eigenvalue $\lambda \in \mathbb{R}_+$ and A_N has an eigenvalue μ with $\mathrm{Re}(\mu) = \alpha \leq \lambda$. Since $\lambda \in \mathbb{R}_+$ is an eigenvalue of $\tilde{A}$, there is an $x = (\tilde{x}', 0)'$ so that $\tilde{x}$ is an eigenvector of $\tilde{A}$. Since $(\tilde{A}, \tilde{C})$ is an observable pair, $\tilde{C}\tilde{x} \neq 0$. Without loss of generality, assume $\tilde{C}\tilde{x} > 0$. (If not, choose $-\tilde{x}$.)

Let $v \in N$ be an eigenvector of A corresponding to μ. The function $|Ce^{(\mu-\lambda)t}v|$ is bounded. Let M be an upper bound, which we may take to be greater than 1. Choose ξ as follows. Let

$$r > (2M)/(\tilde{C}\tilde{x}),$$

and $\xi = rx$. Then

$$Ce^{tA}\xi = e^{\lambda t}(\tilde{C}r\tilde{x}) > 2M > 2$$

for all t, and

$$\begin{aligned}Ce^{tA}\xi + Ce^{tA}v &= e^{\lambda t}\left(r\tilde{C}\tilde{x} + Ce^{(\mu-\lambda)t}v\right) \\ &> e^{\lambda t}(2M - M) = Me^{\lambda t} \geq M > 1\end{aligned}$$

for all t, contradicting $\boldsymbol{Q}$.

Now we assume that the two conditions hold and prove that $\boldsymbol{Q}$ must hold. Given any $x \in \mathbb{R}^n$, and $v \neq 0$, $v \in N$, it suffices to show that $Ce^{tA}v \not\equiv 0$ and either $\inf_{t>0}|Ce^{tA}x| = 0$ or $\inf_{t>0}|Ce^{tA}(x + v)| = 0$. By assumption, $N \bigcap \mathcal{O}(A, C) = \{0\}$, so $Ce^{tA}v \not\equiv 0$ for all such nonzero $v \in N$. If $\inf_{t>0}|Ce^{tA}x| = 0$ then we are done. Now suppose $\inf_{t>0}|Ce^{tA}x| \neq 0$. Then by Proposition 2.6, $\tilde{A}$ must have an eigenvalue $\lambda \in \mathbb{R}_+$. Since $N \neq \{0\}$, A_N has at least one eigenvalue. By assumption, $\mathrm{Re}(\mu) > \tilde{\lambda}$ for all such $\mu \in \sigma(A_N)$. Condition 1 states $\sigma(A_N) \bigcap \mathbb{R}_+ = \emptyset$, so any such μ has nonzero imaginary part.

Any exponent, μ, of $Ce^{tA}v$ for $v \in N$ satisfies $\mathrm{Re}(\mu) > \tilde{\lambda}$. Suppose first that μ is not an exponent of $Ce^{tA}(x + v)$. This can happen only if the terms of $Ce^{tA}v$ having μ as an exponent are exactly cancelled by similar terms in $Ce^{tA}x$. In that case, $Ce^{tA}x$ has μ as an exponent. But then $f(t) = Ce^{tA}x$ is a Bohl function with $\mathcal{E}(f) \bigcap \mathbb{R}_+ = \emptyset$. As in the proof of Proposition 2.6, either $\mathcal{E}(f) \subseteq \mathbb{C} \setminus \mathbb{C}_+$, in which case $Ce^{tA}x \to 0$ as $t \to \infty$, or $\mathcal{E}(f) \subseteq \mathbb{C}_+ \setminus \mathbb{R}_+$ and we may apply Lemma 2.5. In either case, $\inf_{t>0} |Ce^{tA}x| = 0$, a contradiction. Thus μ must be an exponent of $Ce^{tA}(x+v)$ which implies, exactly as we argued above, that $\inf_{t>0} |Ce^{tA}(x+v)| = 0$, and $\boldsymbol{Q}$ holds. ∎

The following Theorem is simply a corollary of 5.2. For each $j \in J$, let $\tilde{A}_j$ be defined by the observability decomposition for (A, C_j):

$$\begin{pmatrix} \tilde{A}_j & 0 \\ A_1^j & A_2^j \end{pmatrix} \begin{pmatrix} \tilde{C}_j & 0 \end{pmatrix}. \tag{10}$$

Theorem 4 *If $\Sigma = (A, B, C)_{\text{os}}$ is a continuous-time output-saturated system and $J = \{j\}$, then Σ is observable if and only if*

1. *(A, C) is an observable pair,*
2. *$\sigma(A_N) \cap \mathbb{R}_+ = \emptyset$, and*
3. *$(\tilde{\lambda} I - A_N)$ is Hurwitz, where $\tilde{\lambda} := \max\{\lambda \in \mathbb{R}_+, \lambda \in \sigma(\tilde{A}_j)\}$.*

Proof. Use $N = N_J$ and $C = C_j$ in 5.2. ∎

In the single output case this Theorem reduces to the following. A single output system $\Sigma = (A, B, C)_{\text{os}}$ with $\mathcal{A} \equiv 0$ is observable if and only if (A, C) is an observable pair and $\sigma(A) \cap \mathbb{R}_+ = \emptyset$. Lemmas 3.3 and 3.1 together add the fact that if $\mathcal{A} \not\equiv 0$ then observability is equivalent to observability of the pair (A, C). This yields another proof of Theorem 3.

We now use Theorem 4 to present a complete characterization of observability for the case $p = 2$.

Proposition 5.3 Suppose $\Sigma = (A, B, C)_{\text{os}}$ is a continuous-time output-saturated system with $p = 2$. Assume (A, C) is an observable pair.

1. If $|I(\mathcal{A})| = 0$, Σ is observable if and only if A has no eigenvalues on $\mathbb{R}_+$.
2. If $|I(\mathcal{A})| = 1$, Σ is observable if and only if A_N has no eigenvalues on $\mathbb{R}_+$ and $(\tilde{\lambda} I - A_N)$ is Hurwitz.
3. If $|I(\mathcal{A})| = 2$, then Σ is observable.

Proof. 1. In this case, $J = \{1, \ldots, p\}$, so $M = \{0\}$, $N = \mathbb{R}^n$ and so $A^M = A_N = A$. Thus, Theorems 1 and 2 imply that A having no eigenvalues on $\mathbb{R}_+$ is necessary and sufficient for observability.

2. This is exactly Theorem 4.

3. If both rows of the Markov sequence are nonzero, then $M = \mathbb{R}^n$, $N = \{0\}$ and so $A^M = A_N = 0$. Once again the necessary conditions of Theorem 1 are identical to the sufficient conditions of Theorem 2 and in this case the conditions are trivially satisfied. So certainly $\Sigma = (A, B, C)_{\text{os}}$ is observable. ∎

As an illustration, applying this result to the system in Example 4.2, we obtain that $(\tilde{\lambda} I - A_N)$ is not Hurwitz, and so (using statement 2) the system is not observable.

6 Observability for the general multiple output case

In this section, we will generalize Theorem 4. We will look at a rational matrix $W(s)$, defined below, which is closely related to the matrix $\tilde{\lambda}I - A_N$ which was used in the previous case. Recall J is the set of the k indices of the zero rows of $\mathcal{A}$ and for each $j \in J$, $\tilde{A}_j$ and $\tilde{C}_j$ are defined by the observability decomposition for (A, C_j). Note that if we let $v_j(s) = C_j(sI - A)^{-1}$ and $\tilde{v}_j(s) = \tilde{C}_j(sI - \tilde{A}_j)^{-1}$, then for some constant matrix T_j,

$$v_j(s) = [\ \tilde{v}_j(s) \quad 0\ ]T_j.$$

Thus, $v_j(s)$ and $\tilde{v}_j(s)$ have the same poles (seen as functions: $\mathbb{C} \to \mathbb{C}^n$ and $\mathbb{C} \to \mathbb{C}^r$ respectively, where r is the size of $\tilde{A}_j$). But the poles of $\tilde{v}_j(s)$ are exactly the eigenvalues of $\tilde{A}_j$, because the triple $(\tilde{A}_j, I, \tilde{C}_j)$ is canonical, for all j. Thus the exponents of $\tilde{C}_j e^{t\tilde{A}_j}$, coincide with the eigenvalues of $\tilde{A}_j$, which are exactly the poles of $v_j(s)$. For each $j \in J$, define

$$\begin{aligned} \tilde{\lambda}_j &:= \max\{\lambda \in \mathbb{R}_+, \lambda \in \sigma(\tilde{A}_j)\} = \max\{\lambda \in \mathbb{R}_+, \lambda \text{ pole of } v_j(s)\} \\ A_j &:= (\tilde{\lambda}_j I - A)|_N \\ C_j^N &:= C_j|_N. \end{aligned}$$

If $v_j(s)$ has no poles on $\mathbb{R}_+$, we define $\tilde{\lambda}_j$ to be $-\infty$. Then, reordering the rows of C if necessary so that $J = \{1, \ldots, k\}$, let

$$W(s) := \begin{pmatrix} C_1^N(sI - A_1)^{-1} \\ \vdots \\ C_k^N(sI - A_k)^{-1} \end{pmatrix}.$$

Theorem 5 *Let Σ be the output-saturated system $\dot{x} = Ax + Bu$, $y = \boldsymbol{\sigma}(Cx)$, with (A, C) an observable pair. Then Σ is observable if*

1. *$\sigma(A_N) \bigcap \mathbb{R}_+ = \emptyset$, and*
2. *for each nonzero $v \in N$,*

$$\text{some component of } W(s)v \text{ has a pole with negative real part.} \qquad (11)$$

Note that if there is a j so that $\tilde{\lambda}_j = -\infty$ and $W(s)_j v \not\equiv 0$ for some v, then the poles of $W(s)_j v$ all have negative real part, and (11) is satisfied for that v.

Proof. The output-saturated system is observable if and only if $\boldsymbol{Q}$ holds (Lemma 3.1). Thus, it is enough to show that the given conditions are sufficient for property $\boldsymbol{Q}$.

We basically follow the same argument as in the proof of Proposition 5.2. Recall that if $N = \{0\}$, then $\boldsymbol{Q}$ holds automatically, so we may assume there exists some nonzero $v \in N$. For any such nonzero $v \in N$ and any ξ, we must find a $j \in J$ so that

$$\sigma(C_j e^{tA}\xi) \not\equiv \sigma(C_j e^{tA}\xi + C_j e^{tA}v).$$

Choose a row of $W(s)v$ which has a nice pole. For that row j, some nonreal pole of $C_j^N(sI-A_N)^{-1}v$ has real part greater than all possible real poles of $C_j(sI-A)^{-1}$. First note that this implies $C_je^{tA}v \not\equiv 0$, since $C_j^N(sI-A_N)^{-1}v$ has some nonzero pole. So $C_je^{tA}\xi \neq C_je^{tA}(\xi+v)$. Thus, it suffices to show that either

$$\begin{aligned}\inf_t |C_je^{tA}\xi| &= 0, \text{ or} \\ \inf_t |C_je^{tA}(\xi+v)| &= 0.\end{aligned}$$

If $\inf_t |C_je^{tA}\xi| = 0$, then $\boldsymbol{Q}$ clearly holds. Otherwise, suppose $\inf_t |C_je^{tA}\xi| \neq 0$. Then $\tilde{\lambda}_j > 0$, but $C_je^{tA}v$ has a nonreal exponent with real part strictly greater than $\tilde{\lambda}_j$. Thus, just as we argued in the proof of Proposition 5.2, the dominating exponents in $C_je^{tA}(\xi+v)$ are not on $\mathbb{R}_+$, so Lemma 2.5 implies that $\inf_t |C_je^{tA}(\xi+v)| = 0$, and $\boldsymbol{Q}$ holds. ∎

This Theorem is an improvement over the sufficient conditions of Theorem 2. Example 4.4 is an observable system which was not included in the sufficient conditions of Theorem 2, but is included in the conditions of this Theorem. If $\mathbb{R}^4 = \{(x_1,x_2,x_3,x_4)'\}$, then in that example, N is just the x_2-x_3 plane, so $\sigma(A_N) = \{3 \pm i\}$, and $\sigma(A_N) \cap \mathbb{R}_+ = \emptyset$. The pair (A,C) is already in the decomposition form, so it is easy to see that $\sigma(\tilde{A}_1) = \{2, 3 \pm i\}$ and $\tilde{\lambda}_1 = 2$. Then

$$\begin{aligned}A_j &= \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} \\ C_1^N &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ W(s) &= \frac{1}{s^2+2s+2}\begin{pmatrix} s+1 \\ -1 \end{pmatrix}.\end{aligned}$$

The poles of $W(s)$ are $-1\pm i$, which both have negative real part. In this case there is no $v = (v_1, v_2)'$ (nonzero) for which

$$(s+1)v_1 - v_2$$

cancels out either of the poles of $W(s)$. That is, for every nonzero $v \in N$, $W(s)v$ has a pole with negative real part. In the case of 2 outputs, the conditions of this Theorem are also necessary, as they are equivalent to the conditions in Theorem 4.

7 Small input observability

In this section, we investigate the observability of a class of output-saturated systems for which the inputs are restricted to be bounded. Unlike the case for linear systems, observability of an output-saturated system is intimately related to controllability. Thus, it is natural to ask what additional conditions are required for observability if controllability is restricted. We first give some general characterizations of observability for this class of bounded-input output-saturated systems. In the case in which the pair (A, B) is already known to be stabilizable we will be able to provide an explicit criterion for observability.

We will use the following notations for a (single output) bounded-input output-saturated system defined by the triple (A, B, C):

$$\begin{aligned}\Sigma_{\text{ios}}: \ \dot{x} &= Ax + Bu \\ y &= Cx \\ z &= \sigma(y)\end{aligned}$$

and $\|u\|_\infty := \sup_{t\geq 0} |u(t)| \leq 1$. (This restriction could be replaced by $u(t) \in U$, where $U \subseteq \mathbb{R}^m$ is any bounded, convex set which contains the origin of $\mathbb{R}^m$ in its interior. One could even dispose of the convexity condition using the bang-bang principle.) If $u(\cdot)$ is a measurable function satisfying $\|u\|_\infty \leq 1$, we will simply say that u is a *small input* (or "small control").

The following lemma gives a characterization of observability for single output, continuous-time output-saturated systems with bounded inputs. The term *small-input "less-than-1" output controllable* means that for any initial state, the output of the linear system can be controlled to inside the interval $(-1, 1)$ using small inputs. This is the same as saying that the state of the system can be controlled to a band around $\ker C$ using small inputs.

Lemma 7.1 The bounded-input output-saturated system $\Sigma_{\text{ios}} := (A, B, C)_{\text{ios}}$ is observable if and only if (A, C) is an observable pair and (A, B, C) is small-input "less-than-1" output controllable.

Proof. For the necessity, suppose the conclusion does not hold. Then there is some x_0 so that the output, if starting from the initial state x_0, can not be controlled to $(-1, 1)$. That is, for any small control u of any length t,

$$|y_u(t, x_0)| = \left| Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}Bu(s)ds \right| \geq 1.$$

Then the same is true of αx_0 for any $\alpha > 1$. Otherwise there would exist an admissible control such that

$$\left| Ce^{At}\alpha x_0 + \int_0^t Ce^{A(t-s)}Bu(s)ds \right| < 1.$$

But then,

$$\left| Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}B\frac{u(s)}{\alpha}ds \right| < \frac{1}{\alpha} < 1,$$

and the input is still admissible, contradicting our assumption. Thus, the entire line segment from x_0 to $2x_0$ has the property that for any initial state in that line segment, there is no control which can force the output inside the "linear window" $|Cx| < 1$. It follows by continuity of $y(t, \alpha x_0)$ on t and α that $\operatorname{sign}(y(t, \alpha x_0))$ is independent of α. That is, for all initial states in that line segment, the outputs are on the same side of the hyperplane $\ker C$. (We may assume $C \neq 0$; otherwise the system is not observable.) Thus any two points x_0^1, x_0^2 on the segment are indistinguishable since

$$\sigma\left(C\phi(0, t, x_0^1, u)\right) = \sigma\left(C\phi(0, t, x_0^2, u)\right)$$

for any u, t, where $\phi(0, t, x_0, u)$ denotes the state of the system after time t starting at initial state x_0 and applying the control u.

For the sufficiency, we simply note that given any two states x, z, we can control the output corresponding to one of them (say x) to a point inside $(-1, 1)$ at some time t. By observability of the pair (A, C), at some time $t \pm \delta$, for δ small enough, the outputs for the two initial states will be different and by continuity, the output for the initial state x will still be inside the interval. ∎

The next lemma uses Lemma 7.1 to prove a characterization of observability which is independent of the control u.

Lemma 7.2 The system Σ_{ios} is observable if and only if (A, C) is an observable pair and

$$\forall x_0 \in \mathbb{R}^n,\ \exists T \geq 0,\ |Ce^{TA}x_0| < 1 + \int_0^T |Ce^{tA}B|_1 dt, \tag{12}$$

where the norm $|\cdot|_1$ is the sum of the absolute values of the components of the vector.

Proof. Using Lemma 7.1, it is enough to show that (12) is equivalent to saying that for every $x_0 \in \mathbb{R}^n$, there exists $\tau \geq 0$ and an input function u such that, $|y_u(\tau, x_0)| < 1$.

First assume that $Cx_0 > 0$. In this case, observability implies that there exists a u and a τ such that

$$y_u(\tau, x_0) = Ce^{\tau A}x_0 + \int_0^\tau Ce^{(\tau-t)A}Bu(t)dt < 1.$$

Let $\hat{u}(t) = -\mathrm{sign}\,(Ce^{(\tau-t)A}B)$, where the sign function is applied componentwise. After a change of variables in the integral, we get

$$y_{\hat{u}}(\tau, x_0) = Ce^{\tau A}x_0 - \int_0^\tau |Ce^{tA}B|_1 dt.$$

Since $\|u(t)\|_\infty < 1$, it follows that

$$-|Ce^{tA}B|_1 \leq Ce^{tA}Bu(\tau - t)$$

for all t, so

$$y_{\hat{u}}(t, x_0) \leq y_u(t, x_0),$$

for all $t > 0$. Thus, $y_{\hat{u}}(\tau, x_0) < 1$. Also, $y_{\hat{u}}(0, x_0) > 0$, so by continuity, there exists a $T \in [0, \tau]$ so that

$$0 < y_{\hat{u}}(T, x_0) < 1.$$

In particular, $Ce^{TA}x_0 > 0$ so $Ce^{TA}x_0 = |Ce^{TA}x_0|$. Thus,

$$|Ce^{TA}x_0| < 1 + \int_0^T |Ce^{tA}B|_1 dt.$$

In the case $Cx_0 < 0$ a similar argument (starting with $y_u(\tau, x_0)$ and applying now $\hat{u}(t) = \mathrm{sign}\,(Ce^{(\tau-t)A}B)$) proves the result.

Condition (12) is also sufficient. If $Cx_0 > 0$, it amounts to saying that there exists a T and a control, namely $\hat{u}(t) = -\mathrm{sign}\,(Ce^{(T-t)A}B)$, so that

$$y_{\hat{u}}(T, x_0) \leq |Ce^{tA}x_0| - \int_0^T |Ce^{tA}B|_1 dt < 1.$$

Since $y_{\hat{u}}(0, x_0) > 0$, there must be a $\tau \in [0, T]$ so that

$$0 < y_{\hat{u}}(\tau, x_0) < 1.$$

If instead $Cx_0 < 0$, we use $\hat{u}(t) = \operatorname{sign}(Ce^{(T-t)A}B)$ and the argument is similar. ∎

An explicit criterion for the observability of Σ_{ios} can be given if we assume that (A, B) is stabilizable. Let $\mathbb{R}_{++} := \{\lambda \in \mathbb{R} | \lambda > 0\}$. Recall that $\mathcal{E}(A)$ is the set of dominating eigenvalues of A as defined in Section 2.

Theorem 6 *Let (A, B) be stabilizable. Then Σ_{ios} is observable if and only if (A, C) is an observable pair and $\mathcal{E}(A) \cap \mathbb{R}_{++} = \emptyset$.*

For the proof we will need the following auxiliary results.

Lemma 7.3 If $p = 1$, (A, C) is an observable pair, and (12) holds, then

$$\mathcal{E}(A) \cap \mathbb{R}_{++} = \emptyset.$$

Proof. Assume that there exists $\lambda \in \mathcal{E}(A) \cap \mathbb{R}_{++}$. Let the index of λ be k. Since (A, C) is observable, A' is cyclic which implies that A is also cyclic. Thus there is a vector b so that (A, b) is a controllable pair. Then $\mathcal{E}(A) = \mathcal{E}(Ce^{tA}b)$ and there is a term in the function $Ce^{tA}b$ which has exponent λ with index k. If we let $x_0 := b$, then $Ce^{tA}x_0$ contains nontrivially the term $t^{k-1}e^{\lambda t}$. So

$$Ce^{TA}x_0 = T^{k-1}e^{\lambda T}(\alpha + o(1)) \quad T \to \infty$$

for some $\alpha \neq 0$. Without loss of generality we assume that $\alpha > 0$. Otherwise we replace x_0 by $-x_0$. Note that we can make α arbitrarily large by choosing $|x_0|$ large. On the other hand, the dominating term of $Ce^{tA}B$ has exponent at most equal to λ with index k, so

$$\int_0^T |Ce^{tA}B|_1 dt \leq \beta T^{k-1} e^{\lambda T}$$

for some β. Hence, by a suitable choice of x_0 we can achieve that condition (12) is not satisfied for any $T > 0$.

∎

Lemma 7.4 Suppose $p = 1$, (A, B) is stabilizable, and (A, C) is observable. If $\text{Re}(\lambda) \geq 0$ for all eigenvalues in $\mathcal{E}(A)$, then

$$\mathcal{E}(A) \subseteq \mathcal{E}(Ce^{tA}B).$$

Proof. Let λ be an eigenvalue $\in \mathcal{E}(A)$ with multiplicity k. If we perform a Kalman controllability decomposition on the triple (A, B, C), we can easily see by stabilizability that λ is an eigenvalue of A_1 (with the same multiplicity), where (A_1, B_1, C_1) is a canonical triple. The poles of the transfer function for (A_1, B_1, C_1) (which are the same as the exponents of $Ce^{tA}B$) are exactly equal to the eigenvalues of A_1, with the same multiplicities, by Remark 2.4. Hence $\lambda \in \mathcal{E}(Ce^{tA}B)$ with the same multiplicity k. ∎

Lemma 7.5 Let f be an almost periodic function which is not identically zero. Then there exist $\gamma > 0$ and $\ell > 0$ such that for all $t_0 \in \mathbb{R}$,

$$\int_{t_0}^{t_0+\ell} |f(t)|dt \geq \gamma.$$

Proof. Let $\alpha := \sup_{t\in\mathbb{R}} |f(t)|$. We will first prove that for every t_0, there exists a t_1 in the interval $[t_0, t_0 + \ell]$ so that

$$|f(t_1)| > \alpha/2.$$

There is a T_0 so that

$$|f(T_0)| > \frac{3\alpha}{4}. \tag{13}$$

Since f is almost periodic, there is an ℓ such that every interval of length ℓ contains an $\alpha/4$-almost period. In particular, for every t_0, the interval

$$[T_0 - t_0 - \ell, T_0 - t_0]$$

contains an $\alpha/4$-almost period. Call it $\tau := \tau(t_0)$. By the definition of almost periodic,

$$|f(T_0) - f(T_0 - \tau)| \leq \alpha/4. \tag{14}$$

From (13) and (14) it follows that

$$|f(t_1)| > \alpha/2,$$

where $t_1 = T_0 - \tau$, so

$$t_1 \in [t_0, t_0 + \ell].$$

Since f is almost periodic, it is uniformly continuous. So, in fact, there exists a positive number δ such that for each $t_0 \in \mathbb{R}$, there is a t_1 in $[t_0, t_0 + \ell]$ so that for all t in a neighborhood of t_1 $(t_1 - \delta < t < t_1 + \delta)$

$$|f(t)| \geq \alpha/3.$$

Without loss of generality, $\ell > \delta$ (otherwise we increase ℓ), so

$$\int_{t_0}^{t_0+\ell} |f(t)|dt > \ell\alpha/3 \geq \delta\alpha/3 =: \gamma$$

for all t_0. ∎

Lemma 7.6 Assume that $\mathcal{E}(A) \cap \mathbb{R}_{++} = \emptyset$. If $\mathcal{E}(A) \subseteq \mathcal{E}(Ce^{tA}B)$, then (12) holds.

Proof. Let the common index of the eigenvalues in $\mathcal{E}(A)$ be k. Since $\lambda \in \mathcal{E}(Ce^{tA}B)$ with index k, we can write

$$Ce^{tA}B = t^{k-1}e^{\alpha t}(f(t) + o(1)) \quad t \to \infty, \tag{15}$$

where α is the real part of the eigenvalues in $\mathcal{E}(A)$ and f is an almost periodic function, not identically equal to 0.

Let ℓ and γ be as in Lemma 7.5, pick any $T > \ell$, and denote $T_0 := T - \ell$. Then

$$\int_0^T |Ce^{tA}B|_1 dt \ \ \geq \ \ \int_{T_0}^T |Ce^{tA}B|_1 dt \tag{16}$$

Next we replace $Ce^{tA}B$ with its expression given in (15) and factor out a lower bound for the polynomial and exponential terms. Notice that the integral of the convergent term is still $o(1)$ because the length of the integration interval remains constant as $T \to \infty$. That is, (16) is

$$\begin{aligned} &\geq \int_{T_0}^T t^{k-1} e^{\alpha t} (|f(t)| + o(1)) dt \\ &\geq T_0^{k-1} e^{\alpha T_0} (\int_{T_0}^T |f(t)| dt + o(1)) \text{ as } T \to \infty. \end{aligned} \tag{17}$$

The interval (T_0, T) is of length ℓ, so we may apply Lemma 7.5 and

$$\int_{T_0}^T |f(t)| dt \geq \gamma. \tag{18}$$

Now we show that there is a constant $\beta > 0$ so that

$$T_0^{k-1} \geq \beta T^{k-1}$$

for T large enough. If $T \geq 2\ell$, then $\ell \leq T/2$ so

$$\begin{aligned} T_0 &= T - \ell \\ &\geq T - \frac{T}{2} = \frac{T}{2}. \end{aligned}$$

Thus, if $\beta := 2^{-k}$ (remember, k is fixed),

$$T_0^{k-1} \geq \beta T^{k-1}. \tag{19}$$

Also for the term $e^{\alpha T_0}$,

$$\begin{aligned} e^{\alpha T_0} &= e^{\alpha(T-\ell)} \\ &= e^{\alpha T} e^{-\alpha \ell} \\ &= e^{\alpha T} C \end{aligned} \tag{20}$$

for some constant $C > 0$.

Using (18), (19) and (20), we see that (17) is

$$\begin{aligned} &\geq \beta C T^{k-1} e^{\alpha T} (\gamma + o(1)) \\ &= T^{k-1} e^{\alpha T} (\rho + o(1)) \end{aligned}$$

for some $\rho > 0$ and T sufficiently large.

On the other hand,

$$Ce^{TA}x_0 = T^{k-1}e^{\alpha T}(g(T) + o(1))$$

where, depending on x_0, either $g(t) \equiv 0$, or $g(t)$ is almost periodic. If $g(t)$ is almost periodic, then $Ce^{tA}x_0$ is a Bohl function with dominating exponents all in $\mathbb{C}_+ \setminus \mathbb{R}_{++}$. Then Lemma 2.5 implies that $Ce^{TA}x_0$ has an infinite sequence of zeros $\{t_k\}$ with $t_k \to \infty$. In either case, there is a T as large as necessary so that

$$g(T) = 0.$$

In particular, we may choose a $T > 2\ell$ so that

$$\begin{array}{rcl} |Ce^{TA}x_0| & = & T^{k-1}e^{\alpha T}|o(1)| \\ & < & T^{k-1}e^{\alpha T}(\rho + o(1)) \\ & < & 1 + \displaystyle\int_0^T |Ce^{tA}B|_1 dt. \end{array}$$

In other words, we may choose T so that (12) holds. ∎

We may now prove Theorem 6.

Proof. (Theorem 6) First assume that there exists $\lambda \in \mathcal{E}(A) \cap \mathbb{R}_{++}$. Then Lemma 7.3 proves that (12) is not satisfied and so Σ_{ios} is not observable.

Next we prove the converse. If $\text{Re}(\lambda) \leq 0$ for all $\lambda \in \sigma(A)$, then there exists an admissible control function u such that $x(t) \to 0$ as $t \to \infty$. (See [6].) In this case, it is obvious that for any x_0 there is a control u and a time T such that $|y_u(T, x_0)| < 1$ so Lemma 7.1 implies observability.

Assume now that $\mathcal{E}(A) \cap \mathbb{R}_{++} = \emptyset$ and $\text{Re}(\lambda) > 0$ for $\lambda \in \mathcal{E}(A)$. Then Lemma 7.4 implies that the conditions of 7.6 are satisfied, and then Lemma 7.6 implies that (12) holds. Hence Σ_{ios} is observable. ∎

References

[1] Bohr, H., *Almost Periodic Functions*, Chelsea Publishing Company, New York, 1947.

[2] Delchamps, David F., *State Space and Input-Output Linear Systems*, Springer-Verlag, New York, 1988

[3] Koplon, Renée, and Eduardo D. Sontag, "Linear Systems with Sign-Observations," *SIAM Journal of Control and Optimization*, To appear September 1993.

[4] Schwarzschild (Koplon), Renée, and Eduardo D. Sontag, "Algebraic Theory of Sign-Linear Systems," in *Proc. Amer. Automatic Control Conference*, Boston, June 1991, pp. 799-804.

[5] Schwarzschild (Koplon), Renée, Eduardo D. Sontag, and M.L.J. Hautus, "Output-Saturated Systems," in *Proc. Amer. Automatic Control Conference*, Chicago, June 1992, pp. 2504-2509.

[6] Sontag, Eduardo D., "An algebraic approach to bounded controllability of linear systems," *Int. J. Control*, 1984, Vol. 39, No. 1, pp. 181-188.

[7] Sontag, Eduardo D., *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, Springer-Verlag, New York, 1990.