Noise-Enabled Observability of Nonlinear Dynamic Systems Using the Empirical Observability Gramian

Nathan D. Powel

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Kristi Morgansen, Chair

Juris Vagners

Mehran Mesbahi

Program Authorized to Offer Degree:
Aeronautics & Astronautics

University of Washington

Abstract

Noise-Enabled Observability of Nonlinear Dynamic Systems Using the Empirical Observability Gramian

Nathan D. Powel

Chair of the Supervisory Committee:
Professor Kristi Morgansen
Aeronautics & Astronautics

While control actuation is well understood to influence the observability of nonlinear dynamical systems, actuation of nonlinear stochastic systems by process noise has received comparatively little attention in terms of the effects on observability. As noise is present in essentially all physically instantiated systems, complete analysis of observability must account for process noise. We approach the problem of process-noise-induced observability through the use of a tool called the empirical observability Gramian. We demonstrate that the empirical observability Gramian can provide a unified approach to observability analysis, by providing sufficient conditions for weak observability of continuous-time nonlinear systems, local weak observability of discrete-time nonlinear systems, and stochastic observability of continuous-time stochastic linear systems with multiplicative noise. The empirical observability Gramian can be used to extend notions of stochastic observability that depend explicitly on linear systems structure to nonlinear stochastic systems. We use Monte Carlo methods to analyze the observability of nonlinear stochastic systems with noise and control actuation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number | Page

ACKNOWLEDGMENTS

This dissertation would not have been possible without the assistance and support of my advisor, Professor Kristi Morgansen. Her guidance set me on the path to this accomplishment and helped me to become a more proficient and independent researcher, and without the robotic fish to catch my eye, I might not have ended up in control theory at all.

My thanks also go out to Jake, Brian, Natalie, Atiye, and everyone else in the Nonlinear Dynamics and Controls Lab for assistance during the production of this dissertation. I appreciated your help with edits, advice, and acting as a sounding-board when working out particularly tricky problems.

I also owe many thanks to the patience and support of my wife, Ashley. Her relentless support and unfailing understanding throughout this (long) process is deeply appreciated. Without her assistance I would not be where, or who, I am today.

Last, but not least, I would like to thank the Air Force Office of Scientific Research (AFOSR grant FA-9550-07-1-0528), Office of Naval Research (ONR MURI N000141010952), and Allan and Inger Osberg (Allan and Inger Osberg Endowed Presidential Fellowship in Engineering) for providing the funding for this research.

DEDICATION

To my friend, my partner, my love: my wife Ashley.

Chapter 1

INTRODUCTION

Sensing and estimation are pervasive in everyday life, in both engineered systems, such as thermostats, cell phones, and automobiles, and in biological systems, such as fruit fly halteres, cat whiskers, and human perception. For stability and control, engineered systems, in particular, rely heavily on the ability to sense and estimate otherwise hidden variables, and may be constrained in their design and performance by what they can estimate from a fixed set of sensors. The aim of this dissertation is to provide new tools for the analysis of the limits of sensing in nonlinear systems and to analyze the effects of noise in system dynamics on the sensing of hidden variables. Given that noise is ubiquitous in natural and engineered systems, we believe that a better understanding of noise in nonlinear systems could lead to substantial improvements in sensing and estimation.

The property of a system that denotes whether all states of the system can be estimated from the output of that system (measurements or sensor readings) is called observability. Observability is important because modern control relies on knowledge of system states for controllers to react on. In theory, if a system is observable, then an estimator can be designed to determine its states. However, if a system is not observable, then some states cannot be determined from the output of the system by *any* estimator.

For deterministic linear systems with known dynamics, observability can be determined quickly and easily, and observability and control of the system are decoupled. Checking observability for deterministic *nonlinear* systems, however, can be significantly more difficult, and choice of control input is well known to influence observability. For example, during its seminal crossing of the Atlantic ocean, the Aerosonde autonomous unmanned vehicle was required to periodically perform S-turn maneuvers in order to obtain full observability of its air-relative velocity because the vehicle was equipped with a Pitot wind-speed sensor and GPS but no compass [1].

To clarify this property of nonlinear systems, consider an example system which we will build upon as we progress through the results in this dissertation: a planar unicycle vehicle, i.e., a vehicle constrained to pivot about its center and move forward or backwards along its current heading angle. Let us assume that we have control over the vehicle's linear acceleration (along the heading direction) and rotational rate and are able to measure the position of the vehicle, but we cannot directly measure its heading. When the vehicle is stationary, the heading is not observable unless an acceleration input is applied, demonstrating that observability can be a state and/or control dependent property in nonlinear systems.

Adding stochastic elements to our analysis can complicate the question of observability further. Noise is often treated as an undesirable property to be suppressed or worked around when possible. However, noise may have beneficial effects on many kinds of systems. White Gaussian noise is intentionally injected into systems (through the control inputs) to provide persistent excitation for system identification [2], and the phenomenon of stochastic resonance, in which the addition of process noise can amplify or enhance some deterministic behavior of a system, has been observed in systems ranging from climate models [3, 4] to neuron signal transmission models [5–9]. As we show below, noise driven actuation can improve observability as well.

Furthermore, noise is ubiquitous in physically instantiated systems; no physical system is truly deterministic if we examine it closely enough. For example, aerodynamic turbulence, vibration, electrical noise, thermal fluctuations, quantum dynamical effects, and other unmodeled dynamics or incompletely understood physics can appear in our models as process noise [10]. We also note that noise can actuate dynamical systems in much the same way as control (which, as we showed above, can be important in nonlinear systems) but, in many systems, noise may be present in states that we cannot actuate directly by the control input. This ubiquity and potential for system actuation beyond the reach of control justify the additional complexity of including noise in observability analysis.

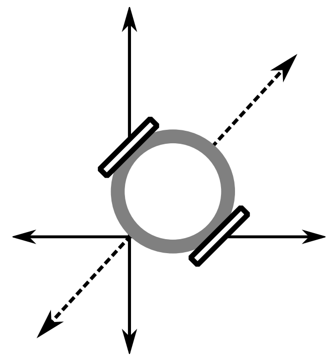


Figure 1.1: The unicycle actuated by noise in the acceleration term will make a random walk along the dashed line, which passes through the vehicle and the origin along the heading angle of the unicycle. Steering noise would add perturbations to either side of the line.

Returning to the unicycle example, we examine the case that the vehicle is stationary at the origin, and our system includes process noise in the acceleration (perhaps due to electrical noise in the drive motors). Driven by noise, the vehicle will move in a random walk along the line through the origin with the vehicle's heading angle, as shown in Figure 1.1. Through this motion, we can determine the heading angle of the vehicle (modulo 180°). Clearly, therefore, situations exist in which process noise can increase the observability of a nonlinear system (at least in the sense that it may allow us to determine the initial state of a system that we could not determine when noise was not considered).

Remark 1.1. While process noise is typically modeled as white and Gaussian for convenience (as we shall do as well), in principle not all process noise need have these properties. As we can see from the unicycle example, all we need is that noise provide actuation to the system, not that it have a particular distribution.

While the example above clearly demonstrates the potential benefits of an approach to observability that embraces noise, existing tools for deterministic nonlinear observability analysis do not apply to stochastic systems, and the existing literature on the observability of stochastic nonlinear systems does not directly address the benefits of noise in observability and is often prohibitively difficult to apply in practice. Therefore, the aim of the research presented here has been to derive common tools for nonlinear deterministic and stochastic systems and to address the following questions: *a*) When can noise actuate a system to make it observable? *b*) How do noise and control actuation interact to influence the observability of nonlinear systems? *c*) How does noise magnitude influence observability quantitatively?

In approaching these questions, we have utilized a tool from deterministic observability analysis called the empirical observability Gramian, which we generalize to include control input. We first derive a formal connection between the singular values of this Gramian and weak observability of nonlinear deterministic systems in discrete and continuous-time. We then extend the empirical Gramian to nonlinear stochastic systems, and use the singular values and condition number of the Gramian as quantitative metrics of observability to approach the analysis of stochastic systems. This approach has allowed us to measure the quantitative impact of process noise on the observability of particular nonlinear systems, and to compare that contribution to the impact of control. For linear stochastic systems with multiplicative noise, we can also show that the rank of the expected value of the empirical Gramian and stochastic observability are connected.

Note that, in this effort, we distinguish process noise from measurement noise. We base our investigation into noise-induced observability upon the known ability of control to influence observability in nonlinear systems (i.e., the lack of a separation principle for nonlinear systems). Process noise can directly actuate system dynamics that may not be actuated (or even actuate-able) by control in some nonlinear systems, which is not possible for measurement noise. Therefore, we do not expect measurement noise to directly influence observability.

1.1 Literature Review

We draw from two primary fields of research in this dissertation, and, as a result, our literature review is broken into two primary sections followed by two shorter sections discussing other relevant research and our own prior work. First, we discuss the literature surrounding the empirical observability Gramian, all of which has been conducted with deterministic continuous-time systems. Second, we discuss the literature surrounding various definitions and results in observability of stochastic systems. Following that, we will describe the publication history of the author and place those works into the context of the work in this dissertation.

1.1.1 Empirical observability Gramian

The empirical observability Gramian was first introduced by Lall, Marsden, and Glavaški [11] in the context of state-space reduction for nonlinear systems [11–13]. The Gramian (we may drop the modifiers "empirical" and "observability" for ease of discussion when there is no chance of confusion with the controllability Gramian or the linear observability Gramian) itself was an extension of the idea of the output energy function described in [14] for model balancing to make the output energy easier to compute numerically, as the output energy function is not generally possible to compute analytically for arbitrary nonlinear systems. In linear systems, the observation energy function can be computed with a weighted quadratic function of the initial state, with the Gramian as the weighting matrix.

The use of the empirical observability Gramian to provide quantitative metrics for observability in nonlinear systems appears to have first surfaced in Singh and Hahn [15], who used the minimum eigenvalue and condition number of the Gramian as useful measures of observability information. However, widespread adoption of observability metrics from the empirical Gramian did not occur until they appeared a second time and were given names in [16]. Krener and Ide [16] named the reciprocal of the minimum eigenvalue of the Gramian the *local unobservability index*, and the condition number of the Gramian the *estimation condition number*. These metrics have seen rapid adoption in applications including underwater navigation [17–19], planetary landers [20], blood glucose modeling [21], power network monitoring [22], and flow-field estimation [23, 24]. The empirical observability Gramian has also been used to give measures of partial observability in systems of partial differential equations [25, 26].

The empirical observability Gramian, as defined in [11–13], explicitly set the control input of the system to zero to capture only the drift dynamics of the system, while the definition used in [16] does not include control explicitly in the system dynamics. However, control is essential in determining the observability of nonlinear systems, so, in order to be useful in a complete observability analysis, the definition of the Gramian must accommodate any valid control input. In the works of Hinson, et al. [17, 23], control was incorporated into the Gramian for the purpose of choosing the control that maximized observability by using the Gramian of the system linearized about a nominal trajectory which was set by the control choice. In [17] the observability Gramian of the linearized system was computed analytically, while in [23] the empirical observability Gramian was used to approximate the Gramian of the linearized system. In [18, 19], the empirical Gramian was computed for a nonlinear system over a finite set of choices of control in order to choose a course that would improve observability for an underwater vehicle. And in [27], the linear observability Gramian was calculated with control for a nonlinear system with linear measurements and dynamics of the form

$$\dot{x}(t) = A(t, u(t), y(t))x(t) + B(t)u(t) \tag{1.1}$$

$$y(t) = C(t)x(t). \tag{1.2}$$

For that case, Batista, Silvestre, and Oliveira showed that the linear observability Gramian having full rank was sufficient for the system to be observable in the sense that the output measurements were unique for that control and initial condition (see Chapter 2 for more general definitions of observability). However, no existing research has addressed *formal* connections between the observability Gramian and nonlinear observability for arbitrary nonlinear systems. We are also unaware of any use of the empirical observability Gramian in discrete-time nonlinear systems.

1.1.2 Stochastic observability

While the concept of observability for linear systems has (relatively) straightforward extensions in deterministic nonlinear systems [28] that are widely accepted, extension of observability to stochastic nonlinear systems is not so straightforward. Deterministic definitions of observability derive from the concept of *indistinguishability* of outputs from different initial conditions, a concept that does not map well to stochastic trajectories, for which the outputs of the system from a single initial condition and control input generally will not be the same from run to run (points are effectively distinguishable from themselves). Some authors attempt to rectify this problem by using indistinguishability of probability distributions [29] (in stochastic systems, the system state can be formulated as a probability distribution), but this approach contains other ambiguities.[1]

Instead, several competing notions of observability exist for stochastic systems, though many have been defined and examined only in the case of linear stochastic systems [29–41], and none have seen predominant adoption. Many of these definitions of stochastic observability follow common themes, including definitions based on attempts to extend indistinguishability to probability distributions of the state [29], definitions based on convergence of state estimates or their covariance below a particular threshold [30, 31, 42–44], stochastic controllability of the dual linear system [33], mutual information between the state and the output [34, 41, 45–52], and generalizations of deterministic exact observability/detectability [37–40, 53, 54]. Some of these definitions are joint properties of a system and a chosen estimator rather than being purely system properties. Related work has also examined the observability of uncertain linear systems [55] with noise.

One of the more common observability definitions for stochastic systems is *exact observability*. A system is exactly observable if the output is zero almost surely (that is, with probability one) for all time only if the initial condition is zero [37, 38, 40]. In [37] a matrix rank condition is given for linear stochastic systems with multiplicative noise, and in [38] sufficient conditions for exact observability are extended to systems with measurement noise. Exact observability has also been examined for discrete-time linear stochastic systems with

[1]A system might be unobservable because we cannot distinguish, using the output, two distributions with the same mean, but different variances, even if we can always distinguish distributions with different means, which would potentially be enough for non-robust, full-state feedback control.

Markov jumps in [39, 40]. A similar condition is often given for *exact detectability*, which requires that the expected value of the square of the measurement becomes zero after some finite time for all initial conditions only if the expected value of the square of the state also goes to zero [53]. Because these properties do not generalize well to nonlinear observability even for deterministic systems, we do not utilize them in this work.

Gramian-based rank conditions for stochastic observability are given by [32] and [36] for linear stochastic systems with additive noise and multiplicative noise respectively. In [36], an expected Gramian is used for the definition, while in [32] the definition is based upon the convergence of the state covariance to below a threshold and a Gramian-based condition is given for this definition to hold. Definitions for stochastic observability are given by [40, 53] for discrete-time linear stochastic systems with multiplicative noise using a scalar output energy metric are roughly equivalent (accounting for the different time-base) conditions to those of [32, 36]. The definitions of the Gramians in [32, 36] depend explicitly on the linear system matrices, making them difficult to extend to nonlinear stochastic systems. While the energy-based definitions can be extended to nonlinear systems, it is not clear that they, as formulated, capture the non-global nature of observability in nonlinear systems, i.e., they may rely too much on non-local system information to apply well to nonlinear systems.

Another common observability definition for stochastic systems is *estimability* [34,47–51]. Estimability requires that the posterior state estimate error covariance be strictly less than (under a positive definite ordering) the prior covariance for an optimal estimator. For linear discrete-time systems this condition is equivalent to requiring that the mutual information between the state and the measurements is strictly positive for the set containing all measurements after n time steps. A matrix rank condition is given for the estimability of discrete-time linear stochastic systems with additive process noise in both discrete and continuous-time [34] and [51] extends this result to discrete-space discrete-time linear systems. Liu and Bitmead [47–50] examine estimability (and a specialization, *complete reconstructability*, which reduces to complete reconstructability of deterministic linear systems, or the ability to determine the state from measurements taken prior to the current time) for linear and nonlinear discrete-time stochastic systems. However, the estimability of nonlinear systems requires that every measurable function of the initial state with non-zero entropy must have strictly positive mutual information with the measurement, a property which is nearly impossible to check for most nonlinear systems. Also, note that the reconstructibility conditions given do not obviously reduce to traditional nonlinear observability conditions as the noise is removed from the system. Estimability has been applied in the analysis of GPS/INS equipped systems [56, 57] and bearing-only target tracking for underwater vehicles [46].

Beyond the work of Liu and Bitmead, comparatively little effort has been expended examining the observability of nonlinear stochastic systems. In the mid-1970s, Sunahara et al. defined stochastic observability for nonlinear systems (in continuous and discrete-time) based on a probability threshold of a linear feedback estimator converging to within specified error [31,42,43]. A metric of stochastic observability defined by Subasi and Demirekler applies to nonlinear systems, but all analysis using that metric was restricted to LTI systems [41]. And while some of the definitions of stochastic observability above could potentially be extended to nonlinear stochastic systems, definitions that depend explicitly on linear system matrices [35, 36, 58] or that assume a particular form of estimator [31, 42–44] do not extend naturally to general nonlinear systems.

In summary, while there has been significant work done in examining observability for stochastic systems, nearly all of it has been done on linear stochastic systems, and the work that has been done on nonlinear stochastic systems has failed to produce tractable tests that can be applied to even relatively simple nonlinear systems. Furthermore, to the best of our knowledge, no work has examined the influence of process noise on the observability of nonlinear systems.

1.1.3 Filter information bounds

Compared to stochastic observability, information lower bounds and covariance upper bounds for nonlinear filters have been a relatively lightly studied topic. Although the Cramer-Rao lower bound on the error covariance (upper bound on the information) of any unbiased estimator by the Fisher information matrix has been known since 1946 [59], upper bounds are harder to come by. The Fisher information matrix itself has been shown to be equal to the time-derivative of the empirical observability Gramian when the measurement noise covariance is the identity matrix times a constant [24]. Sugathadasa et al. provided a bound on the mean-square error of an extended Kalman filter for a system with linear dynamics, nonlinear measurement equation, and uncertain, but bounded, measurement noise covariance [60], but not for nonlinear dynamics. Zakai and Ziv also provided tight upper bounds on filter mean-square-error for very specific linear diffusion processes with white noise, but the derivation does not generalize well to nonlinear systems [61]. Upper and lower bounds on the density function of the Kalman filter error covariance are provided for the case of randomly dropped measurements from a linear system in [62]. However, a lower bound on the information for a general nonlinear system in the extended Kalman filter or unscented Kalman filter does not seem to be available.

In this work, we place lower bounds on the information matrix for both the extended Kalman filter and on the unscented Kalman filter for nonlinear systems with particular classes of measurement functions. We also generalize the relation of [24] into an upper and lower bound on the Fisher information matrix for arbitrary measurement covariance noise.

1.1.4 Previous work

My work relevant to the topic of this dissertation began with assisting in the implementation and testing of a single-beacon based observability optimization algorithm on the robotic fish [63]. Work from Chapter 3 on the connection between the observability Gramian and weak observability of deterministic nonlinear systems, along with Theorem 4.1, was first published in [64]. The work in Chapter 6 and Corollary 4.2 will be submitted to IEEE Transactions on Automatic Control [65], along with some of the material in Chapter 7. A further journal article with the material of Chapter 5 and the remaining material from Chapter 7 is in preparation.

1.2 Contributions

Our research addresses the gaps in the existing literature that were identified in the previous sections. We explicitly incorporate control input into the empirical observability Gramian, and rigorously tie the empirical observability Gramian to weak observability of nonlinear systems. We also demonstrate that the empirical Gramian observability guarantees apply to discrete-time systems as well. In doing so, we provide a more tractable tool for the observability analysis of nonlinear systems than the existing Lie algebraic approach, and provide a rigorous justification for the existing, widespread use of observability metrics derived from the empirical observability Gramian.

To address the lack of tractable tools for nonlinear stochastic observability analysis, we extend the empirical observability Gramian to stochastic systems. We show that the Gramian can provide an equivalent test for stochastic observability of linear stochastic systems to that of Dragan and Morozan [36] and, unlike the method of Dragan and Morozan, can be applied to nonlinear stochastic systems as well. The empirical observability Gramian has the advantage of being relatively straightforward to evaluate, even for continuous-time stochastic systems and analytically complicated nonlinear systems, and it provides a unified approach to observability for stochastic and deterministic, and for linear and nonlinear systems.

1.3 Organization

The material in this dissertation is separated into progress made on nonlinear observability in deterministic nonlinear systems (both continuous-time and discrete-time) and results pertaining to stochastic nonlinear systems. Before entering into new material, Chapter 2 contains an overview of the concepts and definitions in nonlinear observability required for the research presented here. Chapter 3 contains work demonstrating a formal connection between weak observability of deterministic nonlinear continuous-time systems and the minimum singular value of the empirical observability Gramian. A connection between the empirical observability Gramian and the Fisher information matrix appears in Chapter 4, providing a link between the Gramian and optimal estimator performance via the Cramer-Rao bound along with a brief detour into bounds on filter error covariance. The observability of deterministic discrete-time nonlinear systems and its relation to a discrete-time variant of the empirical observability Gramian is discussed in Chapter 5. Chapter 6 marks the transition to work on the observability of stochastic nonlinear systems. The extension of the empirical observability Gramian to stochastic systems is described, and the expected value of the now stochastic Gramian is derived for general nonlinear stochastic systems. We demonstrate how the inclusion of process noise can enhance the positive-definiteness of the Gramian and show that the first moment of the Gramian can be connected to the Dragan and Morozan definition of stochastic observability. Chapter 7 includes numerical results illustrating the quantitative effects on observability of process noise, and compares the effects of process noise on observability with the effects of control input. Finally, we conclude the dissertation in Chapter 8, and discuss possible future extensions of this work and remaining open questions.

Chapter 2

OBSERVABILITY BACKGROUND

Before we proceed to our examination of the empirical observability Gramian and observability of stochastic systems, we first need to place ourselves on a sound footing by introducing some mathematical and control concepts that we will need throughout the rest of this dissertation. In this chapter, we provide definitions of observability of linear and nonlinear deterministic systems in continuous and discrete-time and review how observability is determined from Lie algebraic methods in continuous-time nonlinear systems.

Ultimately with observability analysis, we are interested in taking the the states, dynamics, and measurements, and inverting the relationship between what we know (the measurements, controls, and dynamics) and what we do not know (the states). Observability addresses the question of whether such an inversion exists, and the methods for determining it are all either linear of based upon local linearization (e.g., the inverse function theorem). The process of actually performing the inversion is called estimation, and can utilize either linear or nonlinear methods. We do not address the estimation problem in detail in this work.

2.1 Linear Observability

We begin with a brief review of observability of linear systems in continuous and discrete-time. In particular we focus on the observability Gramian, as the empirical observability Gramian that plays such a prominent role in this dissertation is based upon it. More information on the material in this section can be found in [66].

For deterministic continuous-time linear systems, which we shall write as

$$\Sigma_l : \quad \begin{aligned} \dot{x} &= A(t)x + B(t)u \\ y &= C(t)x, \end{aligned} \tag{2.1}$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$ and $A(t)$, $B(t)$, and $C(t)$ are real-valued matrices of appropriate dimension, we define observability as the condition that after some finite time, $\tau \geq 0$, knowledge of $y(t)$ and $u(t)$ for all $t \in [0, \tau]$ is sufficient to uniquely determine $x_0 = x(0)$. Let $\Psi(t_1, t_0)$ represent the state-transition matrix of (2.1). One test of observability in linear systems is based upon the rank of the observability Gramian

$$W_O(\tau) = \int_0^\tau \Psi(t, 0)^T C(t)^T C(t) \Psi(t, 0) dt. \tag{2.2}$$

The system Σ_l is observable if and only if $\text{rank}(W_O(\tau)) = n$ for some $\tau > 0$. Equivalently, we can examine the observability matrix,

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}, \tag{2.3}$$

of the system (written here for the autonomous case; see [66] for the time-varying equivalent). The system Σ_l is observable if and only if $\text{rank}(\mathcal{O}) = n$.

An analogous situation holds for discrete-time linear systems

$$\Sigma_{ld} : \quad \begin{aligned} x_{k+1} &= A_k x_k + B_k u_k \\ y_k &= C_k x_k, \end{aligned} \tag{2.4}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^p$, $u_k \in \mathbb{R}^m$ and A_k, B_k, and C_k are real-valued matrices of appropriate dimension. Here, we say that the system is observable when there is a finite time index, $\kappa \geq 0$, such that knowledge of y_k and u_k for all $k \in [0, \kappa]$ is sufficient to determine x_0. For this system, let $\Psi[k_1, k_0] = A_{k_1-1}A_{k_1-2}\cdots A_{k_0}$ be the state-transition matrix. The observability Gramian for Σ_{ld} is

$$W_O[\kappa] = \sum_{k=0}^{\kappa} \Psi[k,0]^T C_k^T C_k \Psi[k,0]. \tag{2.5}$$

As in the continuous-time case, the system Σ_{ld} is observable if and only if $\text{rank}(W_O[\kappa]) = n$ for some $\kappa \geq 0$.

We note that for either discrete or continuous-time linear systems, observability is completely independent of control. Specifically, the linear system Gramian matrices are only functions of the unforced system response and the measurement function; the control terms do not appear in the linear Gramian. We also note that observability in linear systems is a global property, not dependent on the initial condition of the system. As we will see in the next section, neither of these properties holds for nonlinear systems.

2.2 Nonlinear Continuous-Time Observability

While observability is relatively straightforward for linear systems, with nonlinear systems more nuance is necessary when defining exactly what we mean by observability. We therefore include a brief review of nonlinear observability definitions and the differential geometric approach for determining the observability of nonlinear systems. The definitions of the various classes of observability in this section are drawn from [28]. Details on the Lie algebraic approach to observability can be found in [28] and [67].

Consider the autonomous nonlinear system

$$\Sigma_{nl}: \quad \begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x), \end{aligned} \tag{2.6}$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$, $u \in \mathcal{U}$, where $\mathcal{U}$ is the set of permissible controls. We will write the solution to the initial value problem (2.6), for $x(0) = x_0$, with the control input $u(t)$, as $x(t, x_0, u)$ with the corresponding output $y(t, x_0, u) = h(x(t, x_0, u))$. For the purpose of nonlinear observability analysis, f and h are typically assumed to be smooth functions [28], which, in turn, guarantees Lipschitz continuity. When f is Lipschitz continuous, the solutions, $x(t, x_0, u)$, of Σ_{nl} are unique and depend continuously upon the initial conditions [68]. If, additionally, h is continuous, then these properties extend to $y(t, x_0, u)$ as well.

Remark 2.1. We will restrict our analysis in this dissertation to autonomous systems (those systems with no explicit time dependence) for simplicity, but, in general, our results can be extended to time-varying systems as well.

We say that points $x_0, x_1 \in \mathbb{R}^n$ are *indistinguishable* if for every control, $u \in \mathcal{U}$, $y(t, x_0, u) = y(t, x_1, u)$ for all t. We say that Σ_{nl} is *weakly observable at* x_0 if there exists an open neighborhood S of x_0 such that if $x_1 \in S$ and x_0 and x_1 are indistinguishable, then $x_0 = x_1$. We say that Σ_{nl} is *weakly observable* if Σ_{nl} is weakly observable at all x.

We say that points x_0 and x_1 are *V-indistinguishable* if for every control, $u \in \mathcal{U}$, with trajectories $x(t, x_0, u)$ and $x(t, x_1, u)$ that remain in the set $V \subseteq \mathbb{R}^n$ for $t \in [0, T]$, we have $y(t, x_0, u) = y(t, x_1, u)$ for all $t \in [0, T]$. We say that Σ_{nl} is *locally weakly observable at* x_0 if there exists an open neighborhood S of x_0 such that for every open neighborhood $V \subseteq S$ of x_0, x_0 and x_1 V-indistinguishable implies that $x_0 = x_1$, and we say that Σ_{nl} is *locally weakly observable* if Σ_{nl} is locally weakly observable at all x.

Note that local weak observability implies weak observability. Intuitively, weak observability at x_0 implies that x_0 can eventually be distinguished from any neighboring points for some control, while local weak observability implies that x_0 can be instantly distinguished from any neighboring points for some control [28].

We also note that observability of nonlinear systems differs from the observability of linear systems in two important ways. One is that observability of nonlinear systems is potentially a local property in the state-space. The system may be observable at some states and not at others. The other is that, unlike the case of a linear system, some nonlinear systems require control to be applied in order to distinguish neighboring initial conditions near a given state (the "there exists $u \in \mathcal{U}$" part of the definition), or conversely that some nearby initial conditions cannot be distinguished when control input is restricted to being uniformly zero.

The usual approach to testing the observability of a nonlinear system comes from differential geometry and provides a rank condition for local weak observability. We define the *Lie derivative* of the function $h(x)$ with respect to a vector field $f_i(x)$ as

$$\mathcal{L}_{f_i(x)} h(x) = f_i(x)^T \frac{\partial h}{\partial x}. \tag{2.7}$$

Because the result of a Lie derivative operation is another vector field mapping between the same two spaces, Lie derivatives can be applied sequentially. The *observability Lie algebra*, $\mathcal{O}$, of a system Σ_{nl}, is the span of the Lie derivatives of the output function, h:

$$\mathcal{O}(x) = \text{span}\{\mathcal{L}_{X_1} \mathcal{L}_{X_2} \ldots \mathcal{L}_{X_k} h(x)\}, \tag{2.8}$$

where $X_i \in \{f(x, u_0) \mid u_0 \in \mathcal{U}\}$ for $i \in \{1, 2, \ldots, k\}$ and all $k \in \mathbb{N}$. If the Jacobian of any set of vectors in the observability Lie algebra, $d\mathcal{O}(x_0)$, is full rank at some state x_0, then the system is locally weakly observable at x_0 [28]. If the system is control affine

$$\Sigma_{\text{aff}} : \quad \begin{aligned} \dot{x} &= f_0(x) + \sum_{i=1}^{m} f_i(x) u_i \\ y &= h(x), \end{aligned} \tag{2.9}$$

where $u_i(t) \in \mathbb{R}$ are the elements of $u(t)$, then $X_i \in \{f_0, f_1, \ldots, f_m\}$ for $i \in \{1, 2, \ldots, k\}$ [67]. The Jacobian of the Lie observability algebra plays a role in nonlinear observability roughly equivalent to that of the observability matrix in linear systems.

Relative to the tests for linear observability, computing the rank of the Jacobian of the Lie observability algebra is considerably more difficult, and for highly nonlinear systems it can be extremely challenging to compute analytically. Furthermore, the rank of $d\mathcal{O}(x)$ provides only a sufficient condition for local weak observability, not a necessary one (recall that, for linear systems, the observability matrix rank condition is both sufficient and necessary). If the Jacobian is rank deficient after examining a particular set of Lie derivatives, we do not always know if we should attempt to add more Lie derivative iterations, $\mathcal{L}_{X_i}$, to our observability algebra, or if the system is simply unobservable. However, when our system is analytic, $f, h \in C^\omega$, it can be shown that we do not need to consider derivatives beyond n-th order [69]; when our system is control affine and analytic, this theorem gives us a finite, but combinatorial, number of derivatives to add to our set.

The Jacobian of the Lie observability algebra gives us some information about the controls that must be permitted for the system to be observable. For example, in a control affine system, Σ_{aff}, if $d\mathcal{L}_{f_i}$ is required to make the Jacobian full rank, then the system would not be observable with $u_i(t) = 0, \forall t \geq 0$. However, the specific structure of the control needed to distinguish all initial conditions is not always clear, particularly for non-control-affine systems.

Remark 2.2. In this dissertation we may, from time to time, describe a system as being observable for certain choices of control. As we can see from the definitions above, such statements are not strictly correct; observability is an existence property, and as long as some control exists that allows the system to satisfy the relevant rank condition, then the system is observable. When we say that the system is only observable for certain choices of control, we mean that the system would not be observable if those choices of control were not in the set of admissible controls, $u \in \mathcal{U}$.

2.3 *Nonlinear Discrete-Time Observability*

Consider the discrete-time nonlinear deterministic system

$$\Sigma_{nd}: \quad \begin{aligned} x_{k+1} &= f(x_k, u_k) \\ y_k &= h(x_k), \end{aligned} \tag{2.10}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^p$, $u_k \in \mathbb{R}^m$, $u \in \mathcal{U}$, where $\mathcal{U}$ is the set of permissible controls. Existence and uniqueness of solutions are guaranteed for (2.10) provided that the domain of f contains its own image for any permissible u_k. For the purposes of the observability analysis here, f and h are assumed to be smooth.

As with continuous-time systems, we would like to be able to perform full-state feedback control, which means that we are interested in whether we can determine the state, x_k, from a finite number of measurements, y_k. To formalize this idea, we utilize the observability definitions of [70], which are, in turn, based upon the definitions of Hermann and Krener [28], and extend the concepts of indistinguishability to discrete-time systems.

We say that states x_0 and x_1 of Σ_{nd} are *indistinguishable* if, for each $k \geq 0$ and for each sequence of controls, $\{u_0, \ldots, u_k\} \in \mathcal{U}$, we have

$$h(f_{u_k} \circ \cdots \circ f_{u_0}(x_1)) = h(f_{u_k} \circ \cdots \circ f_{u_0}(x_2)). \tag{2.11}$$

where $f \circ g(x) = f(g(x))$. We say that Σ_{nd} is *observable at* x_0 if, for any $x_1 \in \mathbb{R}^n$, x_1 is indistinguishable from x_0 only if $x_0 = x_1$. We say that Σ_{nd} is *locally weakly observable at* x_0 if there exists an open neighborhood U of x_0 such that if $x_1 \in U$ and x_0 and x_1 are indistinguishable, then $x_0 = x_1$. Finally, we say that Σ_{nd} is *locally strongly observable at* x_0 if there exists an open neighborhood U of x_0 such that if $x_1, x_2 \in U$ and x_1 and x_2 are indistinguishable, then $x_1 = x_2$. If a system has one of the three observability properties above at each state in its domain, the system itself is said to have that observability property (e.g., if Σ_{nd} is locally weakly observable at x for all $x \in \mathbb{R}^n$, then we say that Σ_{nd} is locally weakly observable).

Note that local weak observability of discrete-time systems is analogous to weak observability of continuous-time systems. Indeed, the "local" property in these definitions corresponds to the "weak" property in continuous-time observability. The continuous-time observability "local" property has no direct analog in discrete-time, and the strong/weak distinction from these definitions has no analog in the definitions from [28].

Determining observability of a nonlinear discrete-time system is typically done analogously to the Lie algebraic approach to observability of nonlinear continuous-time systems. A set of observability co-distributions is defined by

$$\mathcal{O}_1 = \{h(\cdot)\} \tag{2.12}$$

$$\mathcal{O}_k = \{h(f_{u_j} \circ \cdots \circ f_{u_0}(\cdot)) \mid \{u_0, \ldots u_j\} \in \mathcal{U}, 1 \leq j \leq k-1\}, \tag{2.13}$$

and $\mathcal{O}$ is defined as the union of all $\mathcal{O}_k$, $\mathcal{O} = \cup_k \mathcal{O}_k$. Then if $\text{rank}(d\mathcal{O}(x)) = n$, the system Σ_{nd} is locally strongly observable at x [70]. As with continuous-time systems, however, this condition can be difficult to evaluate in closed-form for general nonlinear systems.

Chapter 3

CONTINUOUS-TIME OBSERVABILITY

In this chapter we develop a connection between the empirical observability Gramian and weak observability for deterministic nonlinear systems. First, we define a version of the empirical observability Gramian with explicit control dependence. Next, we show that for the limit case of the Gramian with small initial condition perturbations, the rank of the Gramian gives sufficient conditions for weak observability. We then use that result to extend to the case of arbitrary perturbations, resulting in an upper bound on the minimum singular value of the Gramian to guarantee weak observability. The material in this chapter was originally published in [64].

3.1 Control Explicit Empirical Observability Gramian

In this chapter we consider systems of the form,

$$\Sigma_{nl}: \quad \begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x). \end{aligned} \tag{3.1}$$

We define the empirical observability Gramian for this system similarly to [16], as

$$W_o^\varepsilon(\tau, x_0, u) = \frac{1}{4\varepsilon^2}\int_0^\tau \Phi^\varepsilon(t, x_0, u)^T \Phi^\varepsilon(t, x_0, u)dt, \tag{3.2}$$

where

$$\Phi^\varepsilon(t, x_0, u) = \begin{bmatrix} y^{+1} - y^{-1} & \cdots & y^{+n} - y^{-n} \end{bmatrix} \tag{3.3}$$

and

$$y^{\pm i} = y(t, x_0 \pm \varepsilon e_i, u). \tag{3.4}$$

The vectors e_i denote the elements of the standard basis in $\mathbb{R}^n$.[1] Note that W_o^ε (for convenience we may drop the explicit dependence on t, x_0, and u when we can do so without ambiguity) can be written element-wise as

$$(W_o^\varepsilon)_{ij} = \frac{1}{4\varepsilon^2}\int_0^\tau (y^{+i} - y^{-i})^T(y^{+j} - y^{-j})dt. \tag{3.5}$$

In order for the Gramian to be well-defined, Σ_{nl} must have trajectories that are bounded on the interval $t \in [0, \tau]$. Rather than attempt to place conditions in the theorems that follow to guarantee bounded trajectories, any time that we use the Gramian we will implicitly assume that the system has bounded trajectories up to time τ.

For linear systems,

$$\Sigma_l : \quad \begin{aligned} \dot{x} &= A(t)x + B(t)u \\ y &= C(t)x, \end{aligned} \tag{3.6}$$

this definition reduces to the standard observability observability Gramian

$$W_O = \int_0^\tau \Psi(t,0)^T C(t)^T C(t)\Psi(t,0)dt \tag{3.7}$$

for any control, initial condition, and choice of ε, where, as in Chapter 2, $\Psi(t, t_0)$ is the state-transition matrix of Σ_l. This follows from noting that

$$y^{\pm i} = C(t)\Psi(t,0)(x_0 \pm \varepsilon e_i) + C(t)\int_0^t \Psi(t,t_1)B(t_1)u(t_1)dt_1, \tag{3.8}$$

so that

$$y^{+i} - y^{-i} = 2\varepsilon C(t)\Psi(t,0)e_i \tag{3.9}$$

for any choice of control, u, or initial condition, x_0. Substituting (3.9) into (3.3) gives

$$\begin{aligned} \Phi^\varepsilon &= 2\varepsilon C(t)\Psi(t,0)\begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} \\ &= 2\varepsilon C(t)\Psi(t,0), \end{aligned} \tag{3.10}$$

[1]This definition reduces to the definition in [16] in the case of $u(t) = 0$, and reduces to the definition in [11] if $u(t) = 0\ \forall t \geq 0$, $\mathcal{T}^n = \{(1/\sqrt{2})I\}$, and $\mathcal{M} = \{\sqrt{2}\varepsilon\}$ where I is the $n \times n$ identity matrix. $\mathcal{T}^n$ and $\mathcal{M}$ appear in the definition of the empirical observability Gramian of [11], but we omit that definition here for brevity.

which, when substituted into (3.2) gives W_O. Thus, the empirical observability Gramian appears to be a natural extension of the traditional observability Gramian. One question to ask, therefore, is whether the rank of the empirical Gramian has significance in observability, as the rank of the traditional Gramian does? The answer, as we show below, is yes.

As discussed in Chapter 1, previous uses of the empirical observability Gramian generally neglected the control input, setting it explicitly to $u(t) = 0$ or dropping it from the dynamics entirely. In this work, we specifically include control input in our formulation of the empirical observability Gramian, because, as we showed in Chapter 2, control input is an essential part of the definition of observability in nonlinear systems. We cannot hope to use the Gramian to address a condition that depends on the existence of any control that can allow us to distinguish initial conditions if the Gramian does not reflect that control input. While control input cancels out in the case of a linear system, it generally does not cancel in the nonlinear case.

3.2 Limit $\varepsilon \to 0$

We now examine the case of the empirical observability Gramian in the limit as $\varepsilon \to 0$. While we generally cannot compute the empirical Gramian numerically for this case (Φ^ε goes to zero, and $1/\varepsilon^2$ to infinity, making the problem numerically ill-conditioned, even when the limit exists), the limit case provides insight about the connection between the rank of empirical Gramian and weak observability.

The following lemma will be needed in order to prove our rank condition theorem for weak observability:

Lemma 3.1. *Let f, h be C^1 functions of x and u. If, for some $\tau > 0$, x_0, and $u \in \mathcal{U}$,*

$$\operatorname{rank}\left(\lim_{\varepsilon \to 0} W_o^\varepsilon(\tau, x_0, u)\right) = n, \tag{3.11}$$

then for every non-zero $\delta \in \mathbb{R}^n$, there exists $t \in [0, \tau]$ such that

$$\frac{\partial y}{\partial x_0}\delta \neq 0. \tag{3.12}$$

Proof. First, note that

$$\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \Phi^{\varepsilon}(t, x_0, u) = \frac{\partial y}{\partial x_0}(t, x_0, u) \tag{3.13}$$

by the central difference definition of the derivative. Because f is locally Lipschitz, we know that our system has a unique solution [68], and because the Gramian is assumed to be well-defined, we know that the system trajectories are bounded on any finite interval up to τ. Therefore, we can use the Lebesgue Dominated Convergence Theorem to justify commuting the integral and the limit in (3.11) giving us

$$\lim_{\varepsilon \to 0} W_o^{\varepsilon}(\tau, x_0, u) = \int_0^{\tau} \frac{\partial y}{\partial x_0}^T \frac{\partial y}{\partial x_0} dt. \tag{3.14}$$

We now proceed by contraposition, and assume that for some $\delta \in \mathbb{R}^n$, $\delta \neq 0$,

$$\frac{\partial y}{\partial x_0} \delta = 0 \tag{3.15}$$

for all $t \in [0, \tau]$. For such a δ, we have

$$\lim_{\varepsilon \to 0} W_o^{\varepsilon}(\tau, x_0, u) \delta = 0, \tag{3.16}$$

which implies that $\text{rank}(\lim_{\varepsilon \to 0} W_o^{\varepsilon}) < n$. Thus, we conclude that when $\text{rank}(\lim_{\varepsilon \to 0} W_o^{\varepsilon}) = n$, for every $\delta \in \mathbb{R}^n$, $\delta \neq 0$, there exists $t \in [0, \tau]$ such that

$$\frac{\partial y}{\partial x_0} \delta \neq 0. \tag{3.17}$$

□

In essence, the integrand is positive semidefinite at every point $t \in [0, \tau]$. Therefore, if the integrand becomes strictly positive definite at t, then the integral must also be strictly positive definite and *vice versa*.

Remark 3.1. Note that for $t > 0$,

$$\frac{\partial y}{\partial x_0} \tag{3.18}$$

is the derivative of $y(t, x_0, u)$, evaluated at time t, with respect to the state of the system at time $t = 0$. Therefore, in general,

$$\frac{\partial y}{\partial x_0} \neq \frac{\partial h}{\partial x} \tag{3.19}$$

and is generally quite difficult to compute analytically.

Armed with Lemma 3.1, we can show that, in the limit case, the empirical observability Gramian having full rank is a sufficient condition for the system to be weakly observable.

Theorem 3.2. *Let f, h be C^2 functions of x and u. If there exists $u \in \mathcal{U}$ such that*

$$\operatorname{rank}\left(\lim_{\varepsilon \to 0} W_o^{\varepsilon}(\tau, x_0, u)\right) = n \tag{3.20}$$

for some $\tau > 0$, then the system Σ_{nl} is weakly observable at x_0.

Proof. Suppose that for some $u \in \mathcal{U}$ and $\tau > 0$, $\operatorname{rank}(\lim_{\varepsilon \to 0} W_o^{\varepsilon}(\tau, x_0, u)) = n$. Consider some $x_1 \neq x_0$ and let $\delta = x_1 - x_0$. From Lemma 3.1, we know that for some $t \in [0, \tau]$

$$\frac{\partial y}{\partial x_0}\delta \neq 0. \tag{3.21}$$

Now we can expand $y(t, x_1, u)$ with a Taylor series,

$$y(t, x_1, u) = y(t, x_0, u) + \frac{\partial y}{\partial x_0}\delta + o(\|\delta\|^2) \tag{3.22}$$

where $o(\|\delta\|^2)$ is a continuous map from a neighborhood of $0 \in \mathbb{R}^n$ to $\mathbb{R}^p$ satisfying

$$\lim_{\|\delta\| \to 0} \frac{\|o(\|\delta\|^2)\|}{\|\delta\|^2} = 0 \tag{3.23}$$

and $\|\cdot\|$ is the vector 2-norm (see [71] for more detail on this notation). From the ϵ-δ definition of the limit, we know we can always choose a $0 < \beta < 1$ such that if $\|\delta\| < \beta$ then

$$\frac{\|o(\|\delta\|^2)\|}{\|\delta\|^2} \leq \left\| \frac{1}{2}\frac{\partial y}{\partial x_0}\frac{\delta}{\|\delta\|} \right\|. \tag{3.24}$$

Note that $\frac{\delta}{\|\delta\|}$, the unit vector in the δ direction, is a constant with respect to the mapping $o(\|\delta\|^2)$ and the limit as $\delta \to 0$, which permits us to use it on the right hand side of (3.24). Because we chose $\beta < 1$, we can rewrite (3.24) as

$$\|o(\|\delta\|^2)\| \leq \left\| \frac{1}{2}\frac{\partial y}{\partial x_0}\frac{\delta}{\|\delta\|} \right\| \|\delta\|^2 < \left\| \frac{\partial y}{\partial x_0}\delta \right\|. \tag{3.25}$$

Let $U = \{x \in \mathbb{R}^n \mid 0 < \|x - x_0\| < \beta\}$ and let $x_1 \in U$. We can now say that for some $t \in [0, \tau]$,

$$y(t, x_1, u) - y(t, x_0, u) = \frac{\partial y}{\partial x_0}\delta + o(\|\delta\|^2) \neq 0. \tag{3.26}$$

Thus, there exists an open neighborhood, U, of x_0 such that for all $x_1 \in U$, $x_1 \neq x_0$ there exists a control $u \in \mathcal{U}$ that distinguishes x_0 and x_1 after time t, so the system is weakly observable. □

This theorem is analogous to the Gramian rank condition for observability of a linear system. However, the restriction to the case of $\varepsilon \to 0$ limits its usefulness to systems for which we can analytically compute the empirical observability Gramian, negating one of the primary advantages of the empirical Gramian, that we need only be able to numerically evaluate our system to apply it (we will lift this restriction, at least partially, in §3.3).

We also note that, just as with the Lie derivative algebra approach, our theorem gives sufficient, but not necessary, conditions for observability. We can demonstrate that the converse of Theorem 3.2 does not hold with a simple example. Consider the system

$$\begin{aligned} \dot{x} &= ax \\ y &= x^3. \end{aligned} \tag{3.27}$$

Solving the system analytically, we get $y(t, x_0, u) = e^{3at} x_0^3$ and

$$W_0^\varepsilon(\tau, x_0, u) = \frac{1}{6a} \left(\varepsilon^2 + 3x_0^2\right)^2 \left(e^{6a\tau} - 1\right). \tag{3.28}$$

At the origin, $\operatorname{rank}(\lim_{\varepsilon \to 0} W_o^\varepsilon(\tau, 0, u)) = 0$, however, because the system has only a single state and the measurement function, $h(x)$, is invertible the system is clearly (locally) weakly observable from any state x_0. Note that this example also works for the Lie derivative approach: $\mathcal{O}(x) = \{(3a)^n x^3 \mid n \in \mathbb{N}\}$, so at the origin $\operatorname{rank}(d\mathcal{O}(0)) = 0$.

Finally, we note that the proof presented here for Theorem 3.2 cannot be extended to *local* weak observability, because we do not necessarily know at which time $t \in [0, \tau]$ (3.21) is satisfied. In particular, if the requisite t is not very small, we cannot guarantee that the system trajectories would remain in every open neighborhood $V \subseteq U$ until t. In other words, the system trajectories might have to move a least a finite distance away from the origin before they become distinguishable.

3.3 Finite ε

We now turn our attention to the case of the empirical observability Gramian with finite ε (finite being used here in the limit relative sense, meaning strictly non-zero). The finite ε Gramian has the advantage of being computable merely by simulation, but we can no longer apply Lemma 3.1. However, the fact that W_o^ε *has* a well-defined limit means that we can bound the error between the empirical Gramian with finite ε and the limit Gramian, at least for small ε. This, in turn, will allow us to extend Theorem 3.2 to the case of finite ε, though we will no longer have a pure rank condition, but a minimum singular value condition instead.

To begin, we define the error term as

$$\Delta^\varepsilon(\tau, x_0, u) = W_o^\varepsilon(\tau, x_0, u) - \lim_{\varepsilon \to 0} W_o^\varepsilon(\tau, x_0, u). \tag{3.29}$$

To prevent the equations that follow from growing too large, we will also define the following quantity, which will recur in the next several theorems:

$$\Gamma(t, x_0, u) = \max_i \sup_{\eta \in \mathcal{I}_i^\varepsilon} \left\| D^3 y(\eta)(e_i, e_i, e_i) \right\|_1, \tag{3.30}$$

where $D^k y(x)$ is the k-th Fréchet derivative of $y(t, x_0, u)$ with respect to x_0 evaluated at x, and $\mathcal{I}_i^\varepsilon = [x_0 - \varepsilon e_i, x_0 + \varepsilon e_i]$ is the closed line segment from $x_0 - \varepsilon e_i$ to $x_0 + \varepsilon e_i$. Figure 3.1 shows the set of points over which $\|D^3 y(\eta)(e_i, e_i, e_i)\|_1$ must be maximized. Fréchet derivatives are a natural extension of derivatives to functions from $\mathbb{R}^a$ to $\mathbb{R}^b$ (and, more generally, to functions on arbitrary Banach spaces) [72]. The k-th Fréchet derivative of y, $D^k y(x)$, is a k-linear operator at each point in its domain, and $D^k y(x)(e_i, e_i, e_i) \in \mathbb{R}^p$. We have already encountered the first Fréchet derivative of y,

$$Dy(x) = \frac{\partial y}{\partial x_0}. \tag{3.31}$$

In order to expand the reach of Theorem 3.2, we first bound the norm of the error between the empirical observability Gramian with finite ε and the Gramian as $\varepsilon \to 0$. Note that all matrix norms to follow are induced p-norms.

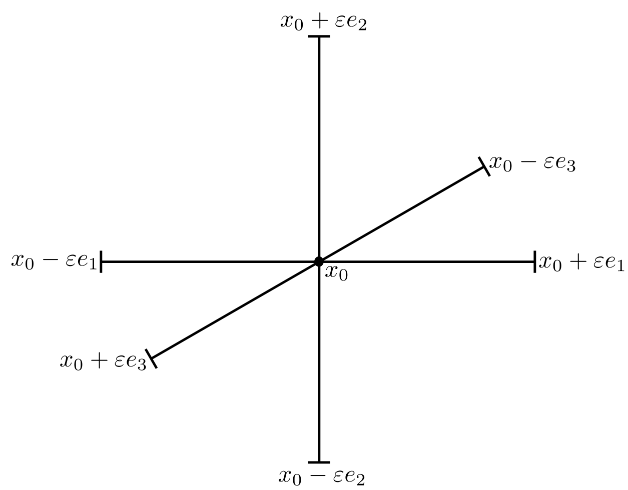


Figure 3.1: The domain for the maximization operation in the definition of Γ for a three-dimensional system.

Lemma 3.3. *Let f, h be C^3 functions of x and u. The error between the empirical observability Gramian and the limit of the Gramian as ε goes to zero is bounded above by*

$$\|\Delta^\varepsilon\|_2 \leq \sup_{t\in[0,\tau]} \left(\frac{\sqrt{n}\varepsilon^2\tau}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4\tau}{36}\Gamma^2 \right) \tag{3.32}$$

for any given ε.

Proof. Let $\varepsilon > 0$. Using Fréchet-Taylor expansion about x_0 for $y(t, x_0 \pm \varepsilon e_i, u)$ (see [72] for details) we can write

$$y^{\pm i} = y(x_0) \pm \varepsilon Dy(x_0)(e_i) + \frac{\varepsilon^2}{2!} D^2 y(x_0)(e_i, e_i) \pm \frac{\varepsilon^3}{3!} R_i^{\pm}(x_0), \tag{3.33}$$

where $R_i^{\pm}(x_0) \in \mathbb{R}^p$ is a remainder term. We used the k-linearity of $D^k y(x_0)$ to factor the $\pm\varepsilon$ terms. Using Taylor's theorem for Banach spaces (permitted by $f, h \in C^3$) we can bound the remainder by

$$\left\| R_i^{\pm}(x_0) \right\|_1 \leq \sup_{\eta\in[x_0, x_0\pm\varepsilon e_i]} \left\| D^3 y(\eta)(e_i, e_i, e_i) \right\|_1, \tag{3.34}$$

where $[x_0, x_0 \pm \varepsilon e_i]$ refers to the line segment connecting x_0 and $x_0 \pm e_i$. Subtracting y^{-i} from y^{+i}, and dividing by 2ε, we get,

$$\frac{y^{+i} - y^{-i}}{2\varepsilon} = \frac{\partial y}{\partial x_0} e_i + \frac{\varepsilon^2}{2 \cdot 3!} (R_i^+(x_0) + R_i^-(x_0)). \tag{3.35}$$

Define

$$\mathbf{R}(x_0) = \begin{bmatrix} R_1^+(x_0) + R_1^-(x_0) & \cdots & R_n^+(x_0) + R_n^-(x_0) \end{bmatrix} \tag{3.36}$$

and note that

$$\begin{aligned} \|\mathbf{R}(x_0)\|_2 &\leq \sqrt{n}\,\|\mathbf{R}(x_0)\|_1 \\ &= \sqrt{n}\max_i \left\|R_i^+(x_0) + R_i^-(x_0)\right\|_1 \\ &\leq 2\sqrt{n}\max_i \sup_{\eta\in\mathcal{I}_i^\varepsilon} \left\|D^3 y(\eta)(e_i, e_i, e_i)\right\|_1 \\ &= 2\sqrt{n}\Gamma. \end{aligned} \tag{3.37}$$

Going back to the definition of Φ^ε from (3.3), we see

$$\begin{aligned} \frac{1}{2\varepsilon}\Phi^\varepsilon &= \begin{bmatrix} \frac{\partial y}{\partial x_0}e_1 & \frac{\partial y}{\partial x_0}e_2 & \cdots & \frac{\partial y}{\partial x_0}e_n \end{bmatrix} + \frac{\varepsilon^2}{2\cdot 3!}\mathbf{R} \\ &= \frac{\partial y}{\partial x_0} + \frac{\varepsilon^2}{2\cdot 3!}\mathbf{R}. \end{aligned} \tag{3.38}$$

From here we can begin the chain of inequalities that conclude the proof. By substituting (3.38) and (3.14) into (3.29) and applying the triangle inequality, we get

$$\begin{aligned} \|\Delta^\varepsilon\|_2 &= \left\| \int_0^\tau \frac{\Phi^\varepsilon}{2\varepsilon}^T \frac{\Phi^\varepsilon}{2\varepsilon} - \frac{\partial y}{\partial x_0}^T \frac{\partial y}{\partial x_0} dt \right\|_2 \\ &\leq \int_0^\tau \frac{\varepsilon^2}{2(3!)} \left\| \frac{\partial y}{\partial x_0}^T \mathbf{R} + \mathbf{R}^T \frac{\partial y}{\partial x_0} \right\|_2 + \frac{\varepsilon^4}{4(3!)^2} \left\|\mathbf{R}^T\mathbf{R}\right\|_2 dt \\ &\leq \sup_{t\in[0,\tau]} \left(\frac{\varepsilon^2\tau}{3!} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \|\mathbf{R}\|_2 + \frac{\varepsilon^4\tau}{4(3!)^2} \|\mathbf{R}\|_2^2 \right) \\ &\leq \sup_{t\in[0,\tau]} \left(\frac{\sqrt{n}\varepsilon^2\tau}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4\tau}{36}\Gamma^2 \right). \end{aligned} \tag{3.39}$$

□

This bound on the error is potentially conservative. For the bound to be tight, it would be necessary for the error to take the same value for all $t \in [0, \tau]$ and for the error in the Fréchet-Taylor expansion to be maximal at each time as well. The right-hand side of (3.32) can be very difficult to evaluate analytically, but we do note that the error in the Gramian scales, in the worst case, with $O(\varepsilon^2\tau)$.

We can now use our bound on the error between the finite ε Gramian and the limit Gramian to state conditions on the minimum singular value of the empirical Gramian for which the limit of the Gramian as $\varepsilon \to 0$ must be full rank. We shall denote the minimum and maximum singular value of a matrix, A, as $\underline{\sigma}(A)$ and $\bar{\sigma}(A)$ respectively, and the minimum and maximum eigenvalues similarly as $\underline{\lambda}(A)$ and $\bar{\lambda}(A)$ respectively.

Theorem 3.4. *Let f, h be C^3 functions of x and u. If there exists $u \in \mathcal{U}$ such that*

$$\underline{\sigma}(W_o^\varepsilon) > \sup_{t\in[0,\tau]} \left(\frac{\sqrt{n}\varepsilon^2\tau}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4\tau}{36}\Gamma^2 \right) \tag{3.40}$$

for some $\tau > 0$, then the system is weakly observable at x_0.

Proof. If

$$\underline{\sigma}\left(W_o^\varepsilon\right) > \sup_{t\in[0,\tau]} \left(\frac{\sqrt{n}\varepsilon^2\tau}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4\tau}{36}\Gamma^2 \right), \tag{3.41}$$

then it follows from Lemma 3.3 that $\underline{\sigma}(W_o^\varepsilon) > \|\Delta^\varepsilon\|_2 = \bar{\sigma}(\Delta^\varepsilon)$. As both matrices are symmetric, this inequality further implies $\underline{\lambda}(W_o^\varepsilon) > \bar{\lambda}(\Delta^\varepsilon)$. Therefore, by Weyl's inequality,

$$\underline{\lambda}\left(\lim_{\varepsilon\to 0} W_o^\varepsilon\right) = \underline{\lambda}(W_o^\varepsilon - \Delta^\varepsilon) \geq \underline{\lambda}(W_o^\varepsilon) - \bar{\lambda}(\Delta^\varepsilon) > 0. \tag{3.42}$$

This proves that $\underline{\lambda}(\lim_{\varepsilon\to 0} W_o^\varepsilon) > 0$, therefore $\lim_{\varepsilon\to 0} W_o^\varepsilon(\tau, x_0, u) \succ 0$. By Theorem 3.2, we conclude that the system is weakly observable. □

Essentially, if the empirical observability Gramian is sufficiently positive definite (minimum singular value large enough), then the limit of the Gramian as $\varepsilon \to 0$ must also be full rank. As before, we note that this theorem gives only sufficient, not necessary conditions for weak observability. Theorem 3.4 is potentially conservative in that it assumes that the limit form of the Gramian is as close to the edge of the positive definite cone as possible, and sets the bound on the minimum singular value of the empirical Gramian accordingly. The bound on Δ^ε from Lemma 3.3 is similarly potentially conservative. For some systems, the limit form of the Gramian could be even more positive definite than the empirical Gramian, in which case the bound could be relaxed.

A useful corollary to Theorem 3.4 formalizes the local unobservability index by giving an upper bound to the unobservability index sufficient to guarantee weak observability of the system.

Corollary 3.5. *Let f, h be C^3 functions of x and u. If, for some $u \in \mathcal{U}$ and $t > 0$,* $\operatorname{rank}(W_o^\varepsilon) = n$ *and the local unobservability index is bounded above by*

$$\frac{1}{\underline{\lambda}(W_o^\varepsilon)} < \frac{1}{\sup_{\tau\in[0,t]}\left(\frac{\sqrt{n}\varepsilon^2 t}{3}\left\|\frac{\partial y}{\partial x_0}\right\|_2 \Gamma + \frac{n\varepsilon^4 t}{36}\Gamma^2\right)} \tag{3.43}$$

then the system is weakly observable at x_0.

Proof. Because W_o^ε is full rank by assumption and positive semidefinite by definition, we know that $\underline{\lambda}(W_o^\varepsilon) > 0$, so the local unobservability index is well defined. We can also say that

$$\frac{1}{\underline{\lambda}(W_o^\varepsilon)} = \frac{1}{\underline{\sigma}(W_o^\varepsilon)} \tag{3.44}$$

and the corollary follows from Theorem 3.4. □

The empirical observability Gramian, and the quantitative observability metrics derived from it, have seen rapid widespread adoption due to the relative ease of application compared the traditional nonlinear observability tools. Theorem 3.4 and Corollary 3.5 provide a strong theoretical justification for the use of the empirical Gramian and unobservability index, as well as clarifying the limits of the metrics (i.e., the local unobservability index relates to weak, not local weak observability). Theorem 4.1 and Corollary 4.2, in Chapter 4, provide a better theoretical understanding of the estimation condition number as well.

A difficulty in applying the bound of Theorem 3.4 is that we cannot analytically compute $\Gamma(t, x_0, u)$ or $\frac{\partial y}{\partial x_0}(t, x_0, u)$ for $t > 0$ without solving the system dynamics in closed-form. Provided that the system can be simulated numerically, however, then both these quantities can be computed by finite difference methods. We also point out that the empirical observability Gramian can only provide observability information at a single point. Determining the observability of the system over a region requires either repeated computation of the

Gramian over points in that region, or the ability to explicitly determine the dependence of the minimum singular value of the Gramian on the initial condition.

Finally, much as with Lyapunov stability, the onus is on the user to find a control input that satisfies the conditions of the theorem. Lack of rank of the Gramian cannot be used to show that the system is unobservable, only that the current choice of control does not actuate the system in a way that exposes all states to the sensors. Another control policy could show that the system is observable at the given initial condition. However, this information can still be used to evaluate the relative merits of different control policies on estimation quality.

Chapter 4

BOUNDS ON FILTER INFORMATION

In this chapter we make a brief side-trip to describe work placing bounds on the error covariance of nonlinear estimators. The Cramer-Rao theorem [59] places, as a lower bound on the error covariance of any unbiased nonlinear filter, the inverse of the Fisher information matrix. Utilizing the empirical observability Gramian (in the limit as $\varepsilon \to 0$), we can bound the Fisher information matrix for an arbitrary nonlinear system, a result which was originally published in [64]. For the specific case of the extended and unscented Kalman filters, we are also able to place upper bounds on the error covariance for nonlinear systems with linear or bounded measurement functions, h, and arbitrary dynamics, f.

4.1 Fisher Information Matrix

We begin by showing how the empirical observability Gramian relates to the Fisher information matrix. For a system

$$\Sigma_F : \quad \begin{aligned} \dot{x} &= f(x,u) \\ \tilde{y} &= h(x) + v, \end{aligned} \tag{4.1}$$

where $v \sim \mathcal{N}(0,R)$ i.i.d, the Fisher information matrix of $y(t,x_0,u)$ with respect to x_0 is defined component-wise as

$$F_{ij} = -\operatorname{E}\left[\frac{\partial^2}{\partial x_{0i}\partial x_{0j}} \log\left(f_{\tilde{y}}(\tilde{y};x_0)\right)\right], \tag{4.2}$$

where $\operatorname{E}[\cdot]$ denotes the expected value operation (in this case with respect to x_0) and $f_{\tilde{y}}(\tilde{y};x_0)$ is the probability density function for $\tilde{y}$ given x_0 [59]. We assume, without loss of generality, that R is strictly positive definite. To relate the empirical Gramian to the Fisher information, we note that one property of the Gramian, in the limit as $\varepsilon \to 0$, is that the integrand matches the form of the Fisher information matrix when the measurement noise of the system is Gaussian. Based on this property we can use the Gramian to derive bounds for the Fisher matrix.

Theorem 4.1. *For a nonlinear system* Σ_F*,*

$$\underline{\sigma}\left(R^{-1}\right)\frac{d}{dt}\lim_{\varepsilon\to 0}W_o^\varepsilon(t,x_0,u)\preccurlyeq F(t)\preccurlyeq\bar{\sigma}\left(R^{-1}\right)\frac{d}{dt}\lim_{\varepsilon\to 0}W_o^\varepsilon(t,x_0,u),\tag{4.3}$$

where $\bar{\sigma}(R^{-1})$ *and* $\underline{\sigma}(R^{-1})$ *denote the maximum and minimum singular values of* R^{-1} *respectively.*

Proof. By differentiating (3.14) with respect to t we get,

$$\frac{d}{dt}\lim_{\varepsilon\to 0}W_o^\varepsilon(t,x_0,u)=\frac{\partial y}{\partial x_0}^T\frac{\partial y}{\partial x_0}.\tag{4.4}$$

Because the state dynamics are deterministic and the measurement noise at any time instant is normally distributed, we can write

$$f_{\tilde{y}}(\tilde{y};x_0)=\frac{1}{\sqrt{(2\pi)^p|R|}}e^{-\frac{1}{2}(\tilde{y}-y(t,x_0,u))^TR^{-1}(\tilde{y}-y(t,x_0,u))},\tag{4.5}$$

where $|R|$ is the determinant of R, so that

$$F=\frac{\partial y}{\partial x_0}^T R^{-1}\frac{\partial y}{\partial x_0}.\tag{4.6}$$

In general, for any matrix H, and positive definite matrix Q, we can say $\underline{\sigma}(Q)H^TH\preccurlyeq H^TQH\preccurlyeq\bar{\sigma}(Q)H^TH$, so we conclude

$$\underline{\sigma}\left(R^{-1}\right)\frac{d}{dt}\lim_{\varepsilon\to 0}W_o^\varepsilon(t,x_0,u)\preccurlyeq F(t)\preccurlyeq\bar{\sigma}\left(R^{-1}\right)\frac{d}{dt}\lim_{\varepsilon\to 0}W_o^\varepsilon(t,x_0,u).\tag{4.7}$$

□

The Fisher information matrix is essentially an R^{-1} weighting of the time-derivative of the limit form of the empirical observability Gramian. Note that a special case of Theorem 4.1 was proven with equality for the case $R=\zeta I$ in [24]. As we can see above, when $R\to\zeta I$, Theorem 4.1 reduces to the result of [24] by the squeeze theorem.

Theorem 4.1 suggests that the shape of the empirical observability Gramian can be useful in determining the likely performance limits and conditioning of nonlinear estimators for our system around a particular state. Because the Fisher information is bounded above and below by scalings of the integrand of the empirical Gramian, when the estimation condition number of the system is high (and particularly when the condition number of R^{-1} is also low), the Fisher information is also likely to have a high condition number, which in turn places constraints on the numerical conditioning of unbiased estimators applied to the problem. We can formalize this in the following corollary to Theorem 4.1.

Corollary 4.2. *For the system Σ_F given by (4.1),*

$$\max\left\{1, \frac{\kappa\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)}{\kappa(R)}\right\} \leq \kappa(F) \leq \kappa(R)\kappa\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right) \tag{4.8}$$

where $\kappa(A)$ is the condition number of the matrix A.

Proof. From Theorem 4.1, we know that

$$\bar{\lambda}(F) \leq \bar{\lambda}(R^{-1})\bar{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right) \tag{4.9}$$

and

$$\underline{\lambda}(F) \geq \underline{\lambda}(R^{-1})\underline{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right), \tag{4.10}$$

using the fact the $\lambda_i(R^{-1}) = \sigma_i(R^{-1})$ because $R \succ 0$. It follows immediately that

$$\kappa(F) = \frac{\bar{\lambda}(F)}{\underline{\lambda}(F)} \leq \frac{\bar{\lambda}(R^{-1})\bar{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)}{\underline{\lambda}(R^{-1})\underline{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)} = \kappa(R)\kappa\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right), \tag{4.11}$$

because the condition number of a positive-definite matrix and its inverse are the same.

To arrive at the other part of the inequality, let us assume that $F = \bar{\lambda}(R^{-1})\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)$. This choice of F fits the bounds of Theorem 4.1 and has the largest possible minimum eigenvalue, $\underline{\lambda}(F) = \bar{\lambda}(R^{-1})\underline{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)$. Now we smoothly reduce the maximum eigenvalue of F, reducing the condition number of F, until either $\bar{\lambda}(F) = \underline{\lambda}(F)$, in which case

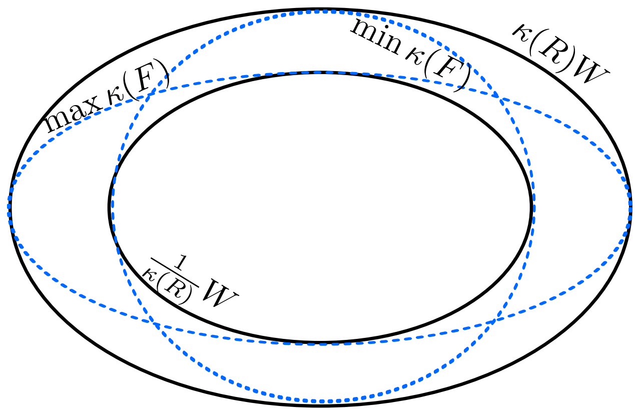


Figure 4.1: The maximum and minimum of the condition number of the Fisher information matrix are bounded by the condition number of the integrand of the empirical observability Gramian scaled by the condition number of the covariance of the measurement noise (W stands for $\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon$).

$\kappa(F) = 1$, its smallest possible value, or until we run into the lower bound from Theorem 4.1, $\bar{\lambda}(F) = \underline{\lambda}(R^{-1})\bar{\lambda}\left(\lim_{\varepsilon\to 0} W_o^\varepsilon\right)$. Thus, we have

$$\max\left\{1, \frac{\kappa\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)}{\kappa(R)}\right\} = \max\left\{1, \frac{\underline{\lambda}(R)\bar{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)}{\bar{\lambda}(R)\underline{\lambda}\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)}\right\} \le \frac{\bar{\lambda}(F)}{\underline{\lambda}(F)} = \kappa(F) \tag{4.12}$$

We could also have begun with $F = \underline{\lambda}(R^{-1})\left(\frac{d}{dt}\lim_{\varepsilon\to 0} W_o^\varepsilon\right)$ and smoothly increased the minimum eigenvalue of F to arrive at the same result. □

Figure 4.1 illustrates the intuition behind Corollary 4.2. The Fisher information matrix ellipsoid is constrained to remain between the two scalings of the ellipsoid given by the integrand of the Gramian. The larger the condition number of the measurement noise covariance, the further apart those ellipsoids will lie, and the larger the freedom there is in the condition number of the Fisher information. If the condition number of R is unity, then there is no room between the ellipsoids at all, and the inequalities of Theorem 4.1 become equality.

Note that $\kappa\left(\lim_{\varepsilon\to 0} W_o^\varepsilon\right)$ is the estimation condition number of the empirical observability Gramian in the limit as $\varepsilon \to 0$. Thus, the conditioning of the Fisher information matrix is bounded above by a scaling of the estimation condition number. The close the condition number of the measurement noise covariance is to unity, the tighter the connection between the estimation condition number and the Fisher information matrix. Thus, we can rigorously connect the numerical conditioning of the estimation problem to the other metric of observability from [16].

4.2 Filter Information Bounds

In addition to placing bounds on the Fisher information matrix, we are able to bound the estimator error covariance directly for specific filter types applied to specific nonlinear systems. These bounds are motivated by the information lower bound on the linear Kalman filter, instead of utilizing the empirical observability Gramian, so for completeness we will include a derivation of that bound here. A more complete treatment may be found in [59].

4.2.1 Kalman filter

Consider the linear system with continuous-time dynamics and discrete-time measurements

$$\Sigma_{\text{KF}} : \quad \begin{aligned} \dot{x} &= Ax + Bu \\ y_k &= H_k x(t_k) + v_k, \end{aligned} \tag{4.13}$$

where $v_k \sim \mathcal{N}(0, R_k)$ represents measurement noise. Assuming that the system is observable and that the initial estimate distribution is Gaussian, the linear Kalman filter is the optimal unbiased estimator for this system [59]. Because the system is linear, the state estimates will remain Gaussian, and can be completely represented at any time t_k as an n-dimensional mean, $\hat{x}_k$, and a $n \times n$ covariance matrix P_k.

The estimate covariance update step of the Kalman filter can be written as

$$P_k^+ = P_k^- - K_k H_k P_k^-, \tag{4.14}$$

where

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{4.15}$$

or, combining (4.14) and (4.15), as

$$P_k^+ = P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^-, \tag{4.16}$$

where P_k^- is the prior error covariance at time t_k, and P_k^+ is the posterior error covariance at time t_k.

Using the Woodbury matrix identity,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \tag{4.17}$$

(assuming that A and C are invertible), we can rewrite the covariance update step in information form as

$$\mathcal{P}_k^+ = \mathcal{P}_k^- + H_k^T R_k^{-1} H_k, \tag{4.18}$$

where $\mathcal{P}_k^\pm = (P_k^\pm)^{-1}$. As R_k is positive definite (by virtue of being a covariance matrix), we can lower bound the Kalman filter information by

$$\mathcal{P}_k^+ \succcurlyeq H_k^T R_k^{-1} H_k. \tag{4.19}$$

Note that this lower bound is essentially the information of a sequential least squares estimate. Unfortunately, the lower bound is not generally invertible, so it cannot be easily converted to an upper bound on the estimate covariance except in special cases.

4.2.2 Extended Kalman filter

We now move our attention to nonlinear systems with discrete-time measurements. Consider

$$\Sigma_{\text{EKF}} : \quad \begin{aligned} \dot{x} &= f(x, u) \\ y_k &= h(x(t_k)) + v_k, \end{aligned} \tag{4.20}$$

where the v_k are, as before, independent Gaussian random variables with zero mean and covariance R_k. Due to the nonlinear dynamics, the linear Kalman filter, and the corresponding estimate information bound, no longer apply. In this section we derive an information bound for the extended Kalman filter (EKF), an extension of the linear Kalman filter based on linearization of our dynamics.

The covariance update step for the extended Kalman filter is

$$P_k^+ = P_k^- - K_k H(\hat{x}_k^-) P_k^- \tag{4.21}$$

$$K_k = P_k^- H(\hat{x}_k^-)^T (H(\hat{x}_k^-) P_k^- H(\hat{x}_k^-)^T + R_k)^{-1} \tag{4.22}$$

$$H(\hat{x}_k^-) = \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_k^-}. \tag{4.23}$$

Note that the EKF clearly makes an implicit assumption that f and h are differentiable in x and u to permit linearization. However, unlike $\frac{\partial y}{\partial x_0}$, $\frac{\partial h}{\partial x}$ is generally straightforward to compute for known dynamics. By applying the same procedure from (4.14)-(4.19), we can arrive at an information lower bound for the EKF

$$\mathcal{P}_k^+ \succcurlyeq H(\hat{x}_k^-)^T R_k^{-1} H(\hat{x}_k^-). \tag{4.24}$$

Because this lower bound depends on the state estimate at a given time, it provides less *a priori* information for the design of the filter. However, for certain classes of output function, $h(x)$, we can remove the state dependence of the lower bound. Note that in the special case of the linear measurement function, $y_k = Hx(t_k) + v_k$, we recover the bound from (4.19).

Theorem 4.3. *For a nonlinear system Σ_{EKF}, if $\forall x$, $\frac{\partial h}{\partial x}^T \frac{\partial h}{\partial x} \succcurlyeq M$ for some $M \succcurlyeq 0$, then for the EKF with covariance given by* (4.21)-(4.23)

$$\mathcal{P}_k^+ \succcurlyeq \frac{1}{\bar{\lambda}(R_k)} M. \tag{4.25}$$

Proof. Let $H(x) = \frac{\partial h}{\partial x}$. Assuming that $H(x)^T H(x) \succcurlyeq M$ $\forall x$ and $M \succcurlyeq 0$, then we can lower bound the EKF estimate information by noting that

$$\begin{aligned} H(\hat{x}_k^-)^T R_k^{-1} H(\hat{x}_k^-) &\succcurlyeq \underline{\lambda}(R_k^{-1}) H(\hat{x}_k^-)^T H(\hat{x}_k^-) \\ &\succcurlyeq \frac{1}{\bar{\lambda}(R_k)} M. \end{aligned} \tag{4.26}$$

Applying this inequality to (4.24), we get

$$\mathcal{P}_k^+ \succcurlyeq \frac{1}{\bar{\lambda}(R_k)} M. \tag{4.27}$$

□

While Theorem 4.3 applies to a narrower class of systems than Theorem 4.1, and only to the extended Kalman filter, it provides a bound directly to the estimate information, rather than to the Fisher information matrix. When M is strictly positive definite, the bound can be inverted to give an upper bound on the error covariance of the EKF as well. The theorem bounds the worst case performance of an EKF applied to a particular system.

4.2.3 Unscented Kalman filter

While the extended Kalman filter is probably the most popular nonlinear estimator, for highly nonlinear systems it can fail to converge. In such cases, another popular filter is the unscented Kalman filter (UKF), which does not rely on linearization. The UKF is a derivative-free filter that numerically approximates the nonlinear transform of the estimate probability distribution by propagating a set of sample points (called sigma points) through the nonlinear dynamics (as opposed to the EKF, which exactly propagates the estimate distribution using linearized dynamics). There are multiple minor variations on the UKF that differ slightly on how exactly to generate the sigma points. We use [59] for the form of the UKF used in this derivation. In that work the update step is given by

$$P_k^+ = P_k^- - K_k P_k^{yy} K_k^T \tag{4.28}$$

$$K_k = P_k^{xy} (P_k^{yy})^{-1} \tag{4.29}$$

$$P_k^{xy} = \sum_{i=0}^{2L} W_c^i (\chi_k^i - \hat{x}_k^-)(\mathcal{Y}_k^i - \hat{y}_k^-)^T \tag{4.30}$$

$$P_k^{yy} = \sum_{i=0}^{2L} W_c^i (\mathcal{Y}_k^i - \hat{y}_k^-)(\mathcal{Y}_k^i - \hat{y}_k^-)^T \tag{4.31}$$

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_m^i \chi_k^i \tag{4.32}$$

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_m^i h(\chi_k^i, {}_v\chi_k^i), \tag{4.33}$$

whereχ_k^i and ${}_v\chi_k^i$ are sigma points, $\mathcal{Y}_k^i = h(\chi_k^i, {}_v\chi_k^i)$, and

$$W_m^0 = \frac{\lambda}{L+\lambda} \tag{4.34}$$

$$W_c^0 = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta \tag{4.35}$$

$$W_m^i = W_c^i = \frac{1}{2(L+\lambda)}, \qquad i > 0, \tag{4.36}$$

for some tuning parameters α, β, and λ. L is the sum of the dimensions of x, w, and v (the state, process noise, and measurement noise respectively).

In the case that the output function, $h(x, v)$, is linear, we can lower bound the information of the UKF similarly to Theorem 4.3. Thus, we consider systems of the form

$$\Sigma_{\text{UKF}} : \begin{aligned} x_{k+1} &= f(x_k, u_k, w_k) \\ y_k &= Hx_k + v_k. \end{aligned} \tag{4.37}$$

Theorem 4.4. *For a nonlinear system Σ_{UKF}, where v_k is uncorrelated with both the state, x_k, and the process noise, w_k, then for the UKF with covariance update given by* (4.28)-(4.33)

$$\mathcal{P}_k^+ \succcurlyeq H^T R^{-1} H. \tag{4.38}$$

Proof. We start by making note of several facts about the UKF in the case of linear measurements. First,

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_m^i (H\chi_k^i + {}_v\chi_k^i) \tag{4.39}$$

$$= H \sum_{i=0}^{2L} W_m^i \chi_k^i + \sum_{i=0}^{2L} W_m^i {}_v\chi_k^i \tag{4.40}$$

$$= H\hat{x}_k^-, \tag{4.41}$$

where $\hat{x}_k^-$ is the mean of the sigma points, χ_k^i are the propagated sigma points for the state, and ${}_v\chi_k^i$ are the sigma points for the measurement noise. The derivation follows because v is zero mean. Also, we note that

$$\sum_{i=0}^{2L} W_c^i (\chi_k^i - \hat{x}_k^-)({}_v\chi_k^i)^T = 0 \tag{4.42}$$

when measurement noise, state, and output noise are uncorrelated and v is zero mean. In these circumstances, ${}_v\chi_k^i$ is zero for $i \notin \mathcal{N}_v$ ($\mathcal{N}_v$ is the set of indices for the the measurement noise-based sigma points in the augmented state vector), and the state and process noise portions of the sigma points are zero for $i \in \mathcal{N}_v$ (i.e., for $i \in \mathcal{N}_v$, $\chi_k^i = \hat{x}_k^+ \pm {}_x\sigma_k^i = \hat{x}_k^+$ and ${}_w\chi_k^i = \bar{w}_k \pm {}_w\sigma_k^i = \bar{w}_k$). Thus we can rewrite (4.42) as

$$\begin{aligned}\sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)({}_v\chi_k^i)^T &= \sum_{i\in\mathcal{N}_v} W_c^i(\chi_k^i - \hat{x}_k^-)(\pm{}_v\chi_k^i)^T \\ &= 0.\end{aligned} \tag{4.43}$$

Next, we note that the update step, (4.28), can be rewritten as

$$P_k^+ = P_k^- - P_k^{xy}(P_k^{yy})^{-1}(P_k^{xy})^T. \tag{4.44}$$

From the definition of P_k^{xy} we see

$$\begin{aligned}P_k^{xy} &= \sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)(\mathcal{Y}_k^i - \hat{y}_k^-)^T \\ &= \sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)(H\chi_k^i + {}_v\chi_k^i - H\hat{x}_k^-)^T \\ &= \sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)(H\chi_k^i - H\hat{x}_k^-)^T + \sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)({}_v\chi_k^i)^T \\ &= \sum_{i=0}^{2L} W_c^i(\chi_k^i - \hat{x}_k^-)(\chi_k^i - \hat{x}_k^-)^T H^T \\ &= P_k^- H^T.\end{aligned} \tag{4.45}$$

Similarly, for P_k^{yy} we can write

$$\begin{aligned} P_k^{yy} &= \sum_{i=0}^{2L} W_c^i (\mathcal{Y}_k^i - \hat{y}_k^-)(\mathcal{Y}_k^i - \hat{y}_k^-)^T \\ &= \sum_{i=0}^{2L} W_c^i (H\chi_k^i + {}_v\chi_k^i - H\hat{x}_k^-)(H\chi_k^i + {}_v\chi_k^i - H\hat{x}_k^-)^T \\ &= \sum_{i=0}^{2L} W_c^i H(\chi_k^i - \hat{x}_k^-)(\chi_k^i - \hat{x}_k^-)^T H^T + \sum_{i=0}^{2L} W_c^i H(\chi_k^i - \hat{x}_k^-)({}_v\chi_k^i)^T \\ &\quad + \sum_{i=0}^{2L} W_c^i ({}_v\chi_k^i)(\chi_k^i - \hat{x}_k^-)^T H^T + \sum_{i=0}^{2L} W_c^i ({}_v\chi_k^i)({}_v\chi_k^i)^T \\ &= HP_k^- H^T + R_k. \end{aligned} \tag{4.46}$$

We can now rewrite the covariance update as

$$P_k^+ = P_k^- - P_k^- H^T (HP_k^- H^T + R_k)^{-1} HP_k^-, \tag{4.47}$$

and proceed as in the linear Kalman filter case to the conclusion

$$\mathcal{P}_k^+ \succcurlyeq H^T R_k^{-1} H. \tag{4.48}$$

□

Although the conditions on the measurement functions of Σ_{EKF} and Σ_{UKF} required for Theorem 4.3 and Theorem 4.4 are strict, notice that no restrictions were required on the dynamics of the system. The lower bound on the information of the filters reflects only the contribution of measurements to the information of the estimator. Any contribution of the dynamics of the system to the filter information would necessarily depend on the initial conditions of the estimator and their covariance, which are generally selected somewhat arbitrarily.

Chapter 5

DISCRETE-TIME OBSERVABILITY

Though continuous-time systems are widely used in control system design and modeling, implementation of control algorithms is often performed with the aid of digital computers and requires discretization of the systems in question. A useful task, therefore, is to examine the extension of the results of Chapter 3 to discrete-time systems. It turns out that this extension is largely straightforward, though the notions of observability used must be adjusted slightly (see Chapter 2 for the definitions we use here).

We follow the approach of Chapter 3 by giving a rank condition on the discrete-time empirical observability Gramian for discrete-time local weak observability (which is analogous to continuous-time weak observability) as $\varepsilon \to 0$. From there we can define an error function for the finite ε Gramian and use that to arrive at an lower bound on the minimum singular value of the Gramian that is sufficient for local weak observability of systems of the form

$$\Sigma_{nd} : \quad \begin{aligned} x_{k+1} &= f(x_k, u_k) \\ y_k &= h(x_k). \end{aligned} \tag{5.1}$$

5.1 Discrete-Time Empirical Observability Gramian

As with the continuous-time case, we can define a control-dependent empirical observability Gramian that we can compute so long as we can simulate the system. We define this discrete-time Gramian as

$$W_o^\varepsilon(\kappa, x_0, u) = \frac{1}{4\varepsilon^2} \sum_{k=0}^{\kappa} \Phi^\varepsilon(k, x_0, u)^T \Phi^\varepsilon(k, x_0, u) dt, \tag{5.2}$$

where

$$\Phi^\varepsilon(k, x_0, u) = \begin{bmatrix} y^{+1} - y^{-1} & \cdots & y^{+n} - y^{-n} \end{bmatrix} \tag{5.3}$$

and

$$y^{\pm i} = y(k, x_0 \pm \varepsilon e_i, u). \tag{5.4}$$

Note that this definition is largely the same definition as (3.2), with the integration replaced by summation.

As with the continuous-time case, the empirical Gramian reduces to the standard discrete-time observability Gramian

$$W_o[\kappa] = \sum_{k=0}^{\kappa} (A^T)^k C^T C A^k \tag{5.5}$$

when the system in question is linear, for any initial condition, control, or choice of ε.

We will show that this discrete-time empirical observability Gramian can be used to prove the local weak observability of a nonlinear discrete-time system in the case as $\varepsilon \to 0$ and that an upper bound on the unobservability index in the finite ε case can guarantee local weak observability. Note, to avoid cumbersome descriptors, we will assume, for the rest of this chapter, that all systems are discrete-time unless explicitly described as continuous-time. The opposite convention is used in the rest of the dissertation.

5.2 *Limit* $\varepsilon \to 0$

We start with a lemma showing that, if the rank of the limit form of the Gramian is n, then the intersection of the null-spaces of the terms of the summation is empty (except for the zero vector).

Lemma 5.1. *Let f, h be C^1 functions of x and u. If, for some $\kappa \geq 0$, x_0, and $u \in \mathcal{U}$,*

$$\operatorname{rank}\left(\lim_{\varepsilon \to 0} W_o^\varepsilon(\kappa, x_0, u)\right) = n, \tag{5.6}$$

then for every $\delta \in \mathbb{R}^n$, $\delta \neq 0$, there exists $k \in \{0, \ldots, \kappa\}$ such that

$$\frac{\partial y}{\partial x_0}\delta \neq 0. \tag{5.7}$$

Proof. First, note that

$$\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon}\Phi^\varepsilon(k, x_0, u) = \frac{\partial y}{\partial x_0}(k, x_0, u) \tag{5.8}$$

by the central difference definition of the derivative. In the interests of brevity, we will occasionally drop the explicit dependence of the derivative on k, x_0, and u going forward. Thus

$$\lim_{\varepsilon\to 0} W_o^\varepsilon(\kappa, x_0, u) = \sum_{k=0}^{\kappa} \frac{\partial y}{\partial x_0}^T \frac{\partial y}{\partial x_0}. \tag{5.9}$$

We know that $\partial y/\partial x_0$ exists for smooth f and h because $y(k, x_0, u)$ is continuous and differentiable in x_0.

We now proceed by contraposition, and assume that, for some $\delta \in \mathbb{R}^n$, $\delta \neq 0$,

$$\frac{\partial y}{\partial x_0}\delta = 0 \tag{5.10}$$

for all $k \in [0, \ldots, \kappa]$. Then

$$\lim_{\varepsilon\to 0} W_o^\varepsilon(\kappa, x_0, u)\delta = 0, \tag{5.11}$$

which violates our assumption that $\lim_{\varepsilon\to 0} W_o^\varepsilon$ is full rank. Thus, for every $\delta \in \mathbb{R}^n$, $\delta \neq 0$, there exists $k \in \{0, \ldots, \kappa\}$ such that

$$\frac{\partial y}{\partial x_0}\delta \neq 0, \tag{5.12}$$

□

As before, we note that $\frac{\partial y}{\partial x_0} \neq \frac{\partial h}{\partial x}$. Provided that our system has solutions that are well-defined and continuous with respect to initial conditions, these derivatives pose no obstacle. As we mentioned in Chapter 2, this situation is the case when f and h are continuous and the domain of f contains its own image under f.

We are now prepared to extend Theorem 3.2 to discrete-time systems.

Theorem 5.2. *Let f, h be C^2 functions of x and u. If there exists $u \in \mathcal{U}$ such that*

$$\operatorname{rank}\left(\lim_{\varepsilon\to 0} W_o^\varepsilon(\kappa, x_0, u)\right) = n \tag{5.13}$$

for some $\kappa \geq 0$, then the system Σ_{nd} is locally weakly observable at x_0.

Proof. Suppose that for some $u \in \mathcal{U}$ and $\kappa \geq 0$, $\operatorname{rank}(\lim_{\varepsilon \to 0} W_o^\varepsilon(\kappa, x_0, u)) = n$. Consider some $x_1 \neq x_0$ and let $\delta = x_1 - x_0$. From Lemma 5.1, we know that for some $k \in \{0, \ldots, \kappa\}$

$$\frac{\partial y}{\partial x_0}\delta \neq 0. \tag{5.14}$$

Now we expand $y(k, x_1, u)$ with a Taylor series about x_0,

$$y(k, x_1, u) = y(k, x_0, u) + \frac{\partial y}{\partial x_0}\delta + o(\|\delta\|^2) \tag{5.15}$$

where $o(\|\delta\|^2)$ is a continuous map from a neighborhood of $0 \in \mathbb{R}^n$ to $\mathbb{R}^p$ satisfying

$$\lim_{\|\delta\| \to 0} \frac{\left\|o(\|\delta\|^2)\right\|}{\|\delta\|^2} = 0. \tag{5.16}$$

By the definition of the limit, we can always choose a $0 < \beta < 1$ such that if $\|\delta\| < \beta$ then

$$\frac{\left\|o(\|\delta\|^2)\right\|}{\|\delta\|^2} \leq \left\|\frac{1}{2}\frac{\partial y}{\partial x_0}\frac{\delta}{\|\delta\|}\right\|, \tag{5.17}$$

or

$$\left\|o(\|\delta\|^2)\right\| \leq \left\|\frac{1}{2}\frac{\partial y}{\partial x_0}\frac{\delta}{\|\delta\|}\right\| \|\delta\|^2 < \left\|\frac{\partial y}{\partial x_0}\delta\right\|. \tag{5.18}$$

Let $U = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq \beta\}$ and let $x_1 \in U$. We can now say that for some $k \in \{0, \ldots, \kappa\}$,

$$y(k, x_1, u) - y(k, x_0, u) = \frac{\partial y}{\partial x_0}\delta + o(\|\delta\|^2) \neq 0. \tag{5.19}$$

Thus, there exists an open neighborhood, U, of x_0 such that for all $x_1 \in U$, $x_1 \neq x_0$ there exists a control $u \in \mathcal{U}$ that distinguishes x_0 and x_1, so the system is locally weakly observable. □

As before, the empirical observability Gramian being full rank is sufficient to guarantee local weak observability of Σ_{nd} but not necessary. Consider the system

$$\Sigma_{sd} : \quad \begin{aligned} x_{k+1} &= x_k \\ y_k &= x_k^2. \end{aligned} \tag{5.20}$$

Solving, we get $y_k = x_0^2$. Letting $x_0 = 0$, we find that for any neighborhood U of x_0 and any $x_1 \in U$, x_0 and x_1 are indistinguishable if and only if $x_1 = x_0$, so the system is locally weakly observable (the origin can be distinguished from any neighboring point as being the only point with strictly zero output). However, $\lim_{\varepsilon\to 0} W_o^\varepsilon = 0$ at the origin.

We cannot use the proof technique of Theorem 5.2 to show sufficiency of the rank condition for local strong observability, because we can construct systems such that for any neighborhood of x_0, we can choose $\delta_1 = x_1 - x_0$ and $\delta_2 = x_2 - x_0$ such that $\|\delta_1 - \delta_2\| \leq o(\|\delta_1\|) + o(\|\delta_2\|)$. Thus, we cannot use Taylor's theorem to guarantee that the outputs of the system are unique in some neighborhood of x_0, as opposed to simply being distinct from $y(k, x_0, u)$. Whether local strong observability can be shown using some other method remains an open question.

We also note that, as we are now summing and not integrating, we can use a non-strict inequality, $k \geq 0$, rather than a strict inequality as in the continuous-time case.

Once again, however, we encounter the problem that we must be able to solve for each term of the sum in closed-form in order to be able to compute the limit of the Gramian. For nonlinear systems, even in discrete-time, this computation is not generally possible.

5.3 Finite ε

Our path, in the discrete-time case, from Theorem 5.2 to a local weak observability proof based on the finite ε Gramian, mirrors that taken in Chapter 3. We begin by defining an error term between the limit Gramian and the finite ε Gramian as

$$\Delta^\varepsilon = W_o^\varepsilon(\kappa, x_0, u) - \lim_{\varepsilon\to 0} W_o^\varepsilon(\kappa, x_0, u), \tag{5.21}$$

and follow by defining

$$\Gamma(k, x_0, u) = \max_i \sup_{\eta\in\mathcal{I}_i^\varepsilon} \left\| D^3 y(\eta)(e_i, e_i, e_i)\right\|_1, \tag{5.22}$$

where $D^j y(x)$ is the j-th Fréchet derivative of $y(k, x_0, u)$ with respect to x_0 evaluated at x, and $\mathcal{I}_i^\varepsilon = [x_0 - \varepsilon e_i, x_0 + \varepsilon e_i]$ is the closed line segment from $x_0 - \varepsilon e_i$ to $x_0 + \varepsilon e_i$. No changes need be made to these definitions from (3.29) and (3.30), except to replace the time parameter with the discrete-time index.

Lemma 5.3. *Let f, h be C^3 functions of x and u. The error between the empirical observability Gramian and the limit of the Gramian as ε goes to zero is bounded above by*

$$\|\Delta^\varepsilon\|_2 \leq \max_{k\in\{0,\dots,\kappa\}} \left(\frac{\sqrt{n}\varepsilon^2(\kappa+1)}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4(\kappa+1)}{36}\Gamma^2 \right) \tag{5.23}$$

for any given ε.

Proof. Let $\varepsilon > 0$. Using Fréchet-Taylor expansion about x_0 for $y(k, x_0 \pm \varepsilon e_i, u)$, we can write

$$y^{\pm i} = y(x_0) \pm \varepsilon D y(x_0)(e_i) + \frac{\varepsilon^2}{2!} D^2 y(x_0)(e_i, e_i) \pm \frac{\varepsilon^3}{3!} R_i^\pm(x_0), \tag{5.24}$$

where $R_i^\pm(x_0) \in \mathbb{R}^p$ is a remainder term. Using Taylor's theorem for Banach spaces, we can bound the remainder by

$$\left\| R_i^\pm(x_0) \right\|_1 \leq \sup_{\eta\in[x_0, x_0\pm\varepsilon e_i]} \left\| D^3 y(\eta)(e_i, e_i, e_i) \right\|_1, \tag{5.25}$$

where $[x_0, x_0 \pm \varepsilon e_i]$ refers to the line segment connecting x_0 and $x_0 \pm e_i$. Subtracting y^{-i} from y^{+i}, and dividing by 2ε, we get

$$\frac{y^{+i} - y^{-i}}{2\varepsilon} = \frac{\partial y}{\partial x_0} e_i + \frac{\varepsilon^2}{2\cdot 3!}(R_i^+(x_0) + R_i^-(x_0)). \tag{5.26}$$

Define

$$\mathbf{R}(x_0) = \begin{bmatrix} R_1^+(x_0) + R_1^-(x_0) & \cdots & R_n^+(x_0) + R_n^-(x_0) \end{bmatrix}, \tag{5.27}$$

and note that

$$\begin{aligned} \|\mathbf{R}(x_0)\|_2 &\leq \sqrt{n}\, \|\mathbf{R}(x_0)\|_1 \\ &= \sqrt{n} \max_i \left\| R_i^+(x_0) + R_i^-(x_0) \right\|_1 \\ &\leq 2\sqrt{n} \max_i \sup_{\eta\in\mathcal{I}_i^\varepsilon} \left\| D^3 y(\eta)(e_i, e_i, e_i) \right\|_1 \\ &= 2\sqrt{n}\Gamma. \end{aligned} \tag{5.28}$$

Going back to the definition of Φ^ε from (5.3), we see

$$\begin{aligned} \frac{1}{2\varepsilon}\Phi^\varepsilon &= \begin{bmatrix} \frac{\partial y}{\partial x_0} e_1 & \frac{\partial y}{\partial x_0} e_2 & \cdots & \frac{\partial y}{\partial x_0} e_n \end{bmatrix} + \frac{\varepsilon^2}{2\cdot 3!}\mathbf{R} \\ &= \frac{\partial y}{\partial x_0} + \frac{\varepsilon^2}{2\cdot 3!}\mathbf{R}. \end{aligned} \tag{5.29}$$

From here we can begin the chain of inequalities that conclude the proof. By substituting (5.29) and (5.3) into (5.21) and applying the triangle inequality, we get

$$\begin{aligned}
\|\Delta^\varepsilon\|_2 &= \left\| \sum_{k=0}^{\kappa} \frac{\Phi^{\varepsilon}}{2\varepsilon}^T \frac{\Phi^\varepsilon}{2\varepsilon} - \frac{\partial y}{\partial x_0}^T \frac{\partial y}{\partial x_0} \right\|_2 \\
&\leq \sum_{k=0}^{\kappa} \frac{\varepsilon^2}{2(3!)} \left\| \frac{\partial y}{\partial x_0}^T \mathbf{R} + \mathbf{R}^T \frac{\partial y}{\partial x_0} \right\|_2 + \frac{\varepsilon^4}{4(3!)^2} \left\| \mathbf{R}^T \mathbf{R} \right\|_2 \\
&\leq \max_{k\in\{0,\ldots,\kappa\}} \left(\frac{\varepsilon^2(\kappa+1)}{3!} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \|\mathbf{R}\|_2 + \frac{\varepsilon^4(\kappa+1)}{4(3!)^2} \|\mathbf{R}\|_2^2 \right) \\
&\leq \max_{k\in\{0,\ldots,\kappa\}} \left(\frac{\sqrt{n}\varepsilon^2(\kappa+1)}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4(\kappa+1)}{36} \Gamma^2 \right).
\end{aligned} \tag{5.30}$$

□

We note a few small changes from the corresponding theorems of Chapter 3. First, the time multiplier must be made one unit larger ($\kappa+1$) to account for the summation including the zero time index. Second, because there were a finite number of time indices to consider, we were able to convert the supremum on the bound to a maximum.

Lemma 5.3 contains the bulk of the work needed to arrive at our intended result, a sufficient upper bound on the minimum singular value of the discrete-time empirical observability Gramian for local weak observability at x_0.

Theorem 5.4. *Let f, h be C^3 functions of x and u. If there exists $u \in \mathcal{U}$ such that*

$$\underline{\sigma}(W_o^\varepsilon) > \max_{k\in\{0,\ldots,\kappa\}} \left(\frac{\sqrt{n}\varepsilon^2(\kappa+1)}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4(\kappa+1)}{36} \Gamma^2 \right) \tag{5.31}$$

for some $\kappa \geq 0$, then the system Σ_{nd} is weakly observable at x_0.

Proof. If

$$\underline{\sigma}\left(W_o^\varepsilon\right) > \max_{k\in\{0,\ldots,\kappa\}} \left(\frac{\sqrt{n}\varepsilon^2(\kappa+1)}{3} \left\| \frac{\partial y}{\partial x_0} \right\|_2 \Gamma + \frac{n\varepsilon^4(\kappa+1)}{36} \Gamma^2 \right), \tag{5.32}$$

then it follows from Lemma 5.3 that $\underline{\sigma}(W_o^\varepsilon) > \|\Delta^\varepsilon\|_2 = \bar{\sigma}(\Delta^\varepsilon)$. As both matrices are symmetric, this inequality further implies $\underline{\lambda}(W_o^\varepsilon) > \bar{\lambda}(\Delta^\varepsilon)$. Therefore, by Weyl's inequality,

$$\underline{\lambda}\left(\lim_{\varepsilon\to 0} W_o^\varepsilon\right) = \underline{\lambda}(W_o^\varepsilon - \Delta^\varepsilon) \geq \underline{\lambda}(W_o^\varepsilon) - \bar{\lambda}(\Delta^\varepsilon) > 0. \tag{5.33}$$

This proves that $\underline{\lambda}(\lim_{\varepsilon\to 0} W_o^\varepsilon) > 0$, therefore $\lim_{\varepsilon\to 0} W_o^\varepsilon(\kappa, x_0, u) \succ 0$. By Theorem 5.2 we conclude that the system is locally weakly observable. □

Corollary 3.5 also holds in discrete-time with only minor modifications.

Corollary 5.5. *Let f, h be C^3 functions of x and u. If, for some $u \in \mathcal{U}$ and $k \geq 0$, $\operatorname{rank}(W_o^\varepsilon) = n$ and the local unobservability index is bounded above by*

$$\frac{1}{\underline{\lambda}(W_o^\varepsilon)} < \frac{1}{\max_{k\in\{0,\dots,\kappa\}}\left(\frac{\sqrt{n}\varepsilon^2(\kappa+1)}{3}\left\|\frac{\partial y}{\partial x_0}\right\|_2\Gamma + \frac{n\varepsilon^4(\kappa+1)}{36}\Gamma^2\right)} \tag{5.34}$$

then the system is locally weakly observable at x_0.

Proof. Because W_o^ε is full rank by assumption and positive semidefinite by definition, we know that $\underline{\lambda}(W_o^\varepsilon) > 0$, so the local unobservability index is well defined. We can also say that

$$\frac{1}{\underline{\lambda}(W_o^\varepsilon)} = \frac{1}{\underline{\sigma}(W_o^\varepsilon)} \tag{5.35}$$

and the corollary follows from Theorem 5.4. □

It would be natural to ask if we can combine the results of the preceding two chapters and arrive at similar theorems for hybrid systems. Unfortunately, for even relatively simple hybrid systems, such as continuous-time systems with discontinuous dynamics and hybrid automata, we are unable to guarantee that the solutions to our system will be continuous in the initial conditions. Theorems 3.4 and 5.4 both rely on this property, making it unlikely that the empirical Gramian will be as useful in such cases. Note, however, that systems with continuous-time smooth dynamics and discrete-time smooth measurements,

$$\Sigma_h : \quad \begin{aligned} \dot{x} &= f(x, u) \\ y_k &= h(x(t_k)), \end{aligned} \tag{5.36}$$

are covered by the proofs of Theorem 5.4 for the Gramian and Theorem 3.4 for existence and uniqueness of solutions. Observability of these systems, which can be reformulated as a type of fully discrete-time system with dynamics

$$\Sigma_h : \quad \begin{aligned} x_{k+1} &= x_k + \int_{t_k}^{t_{k+1}} f(x,u)dt \\ y_k &= h(x(t_k)), \end{aligned} \tag{5.37}$$

can be shown in the discrete-time sense using Theorem 5.4.

Chapter 6

STOCHASTIC OBSERVABILITY

We now turn our attention to the observability of stochastic nonlinear systems in continuous-time. As in Chapter 3, the empirical observability Gramian provides a useful and numerically tractable tool for observability analysis, though once again we must amend our notions of observability to accommodate stochastic dynamics. As one of the main results of this chapter, we demonstrate that the empirical Gramian can provide a means of testing the definition of stochastic observability of linear systems from [36], and that the Gramian may provide a path to extend stochastic observability to nonlinear systems. As we mentioned in Chapter 1, existing definitions of observability for nonlinear systems are generally not tractably testable. An empirical Gramian-based approach to stochastic observability provides a more tractable method of approaching the problem. Much of the material in this chapter will be submitted in an article to IEEE Transactions on Automatic Control [65].

The stochastic nonlinear systems considered in this chapter will have the form

$$\Sigma_S : \quad \begin{aligned} dX &= f(X,u)dt + \sigma(X,u)dW \\ Y &= h(X), \end{aligned} \tag{6.1}$$

where dW is a vector of independent differentials of the Itô sense. We will refer to the $\sigma(X,u)dW$ term of (6.1) as the *process noise* of the system. This form for stochastic nonlinear equations is the most straightforward extension of the continuous-time nonlinear systems addressed in Chapter 3. However, X now represents a vector of random variables over $\mathbb{R}^n$, and there is no longer a single unique state trajectory that satisfies (6.1). Deterministic dynamics for the probability distribution of $X(t)$ are given by the Fokker-Planck equations, and sample trajectories of the system satisfying Σ_S in probability can be drawn by methods such as the Euler-Maruyama method or the Milstein Method.

To simplify the analysis, we will be neglecting measurement noise in this paper. Noise in the output of the system cannot influence the state trajectory, so non-state-dependent measurement noise should be incapable of providing information about system states, which is the primary case that we are interested in. In this chapter we discuss how to apply the empirical observability Gramian to systems of the form of Σ_S, and attempt to derive general properties of the resulting Gramian. Numerical results and simulations of stochastic systems and their observability Gramians are deferred to Chapter 7.

6.1 Stochastic Empirical Observability Gramian

We first wish to define an extension of the empirical observability Gramian to stochastic systems. However, unlike the deterministic case, with stochastic systems we are not guaranteed (or are even likely) to get the same trajectory for multiple sample trajectories of the system. As a result, we have at least three reasonable ways to about computing the Gramian of a stochastic system.

We could compute ensemble averages for entries $y^{\pm i}(t, x_0, u)$, and then compute the Gramian as defined in (3.2) and (3.3). However, in this case, the Gramian would not always reflect the contribution of process noise to the system output. Returning to the example of the unicycle with noise actuated acceleration from Chapter 1, the ensemble average of the output would be 0, not reflecting that, for any particular vehicle trajectory, the output would be non-zero almost certainly. Furthermore, computing the Gramian would require significant computational effort, because it would require simulating the system many times for each perturbed initial condition in order to compute the ensemble average. A potential advantage of this approach would be that the resulting Gramian would be largely deterministic.

Another approach would be to compute a sample trajectory for $y^{\pm i}(t, x_0, u)$ for each of the two times that a given perturbed trajectory is needed in the Gramian (the left and right $\Phi^{\varepsilon}(t, x_0, u)$ terms of (3.2)). While this procedure would allow the Gramian to reflect the influence of process noise, it would have two primary disadvantages. First, it would require $4n$ sample trajectories of the system to compute the Gramian, compared to the $2n$

trajectories required for the deterministic Gramian from Chapter 3. Second, the resulting Gramian would almost certainly be asymmetric, violating our expectation that a Gramian should be symmetric and positive semidefinite.

Finally, we could compute a single sample trajectory for $y^{\pm i}(t, x_0, u)$ for each perturbed initial condition. This option would provide the benefits of the previous option (reflecting process noise), and result in a symmetric positive semidefinite Gramian with only $2n$ sample trajectories required. Note that both this option, and the previous, result in an empirical observability Gramian that is a random variable and that reduces to the deterministic Gramian as $\sigma(X, u)$ goes uniformly to zero.

In this paper we will compute the empirical observability Gramian for nonlinear stochastic systems according to the last algorithm: for each $i \in [0, \dots, n]$ we compute a pair of sample trajectories

$$y^{+i}(t, x_0, u) = y(t, x_0 + \varepsilon e_i, u) \tag{6.2}$$

$$y^{-i}(t, x_0, u) = y(t, x_0 - \varepsilon e_i, u) \tag{6.3}$$

and then compute the Gramian from

$$\Phi^\varepsilon(t, x_0, u) = \begin{bmatrix} y^{+1} - y^{-1} & \cdots & y^{+n} - y^{-n} \end{bmatrix} \tag{6.4}$$

$$W_o^\varepsilon(\tau, x_0, u) = \frac{1}{4\varepsilon^2} \int_0^\tau \Phi^\varepsilon(t, x_0, u)^T \Phi^\varepsilon(t, x_0, u) dt, \tag{6.5}$$

as before. We can compute ensembles of the Gramian to find the mean and variance of the stochastic Gramian if necessary, or apply Monte Carlo methods to find the distribution of the local unobservability index and estimation condition number of the stochastic nonlinear system at a point.

6.2 Expected Value of the Empirical Observability Gramian

Now that the empirical observability Gramian is a random variable, we are interested in finding the moments of the Gramian, particularly the first moment, in order to further characterize the properties of the stochastic Gramian. While the moments of the Gramian can be computed from Monte Carlo ensembles, we often learn more by deriving a result analytically as far as possible. For an arbitrary nonlinear stochastic system we cannot completely determine the mean of the empirical observability Gramian analytically, but we can partially derive a solution, as well as derive a complete solution for simple cases.

Before we proceed with our expected value derivation, we will need the following result.

Lemma 6.1. *For a random vector X, of length n,*

$$\mathrm{E}\left[X^T X\right] = \mathrm{E}[X]^T \mathrm{E}[X] + \mathrm{tr}(\mathrm{Cov}[X]) \tag{6.6}$$

Proof.

$$\begin{aligned} \mathrm{E}\left[X^T X\right] &= \mathrm{E}\left[\sum_{i=0}^{n} X_i X_i\right] \\ &= \sum_{i=0}^{n} \mathrm{E}[X_i X_i] \\ &= \sum_{i=0}^{n} \mathrm{E}[X_i]^2 + \mathrm{Var}[X_i] \\ &= \mathrm{E}[X]^T \mathrm{E}[X] + \mathrm{tr}(\mathrm{Cov}[X]) \end{aligned} \tag{6.7}$$

□

We can prove a similar result for square matrices, provided that the columns are independently distributed. Before we proceed, we must define the $\mathrm{diag}(\cdot)_i$ operator, which is defined as mapping an n-dimensional vector, v specified component-wise as v_i, to a diagonal $n \times n$ matrix whose diagonal elements are given by the components of v and whose off-diagonal elements are all 0. In other words

$$\mathrm{diag}(v_i)_i = \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & v_n \end{bmatrix}. \tag{6.8}$$

Lemma 6.2. *For a random matrix X, of length n, with independent columns*

$$X = \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix}, \tag{6.9}$$

we have

$$\mathrm{E}\left[X^T X\right] = \mathrm{E}[X]^T\,\mathrm{E}[X] + \mathrm{diag}(\mathrm{tr}(\mathrm{Cov}[X_i]))_i \tag{6.10}$$

Proof.

$$\begin{aligned}\mathrm{E}\left[X^T X\right] &= \mathrm{E}\begin{bmatrix} X_1^T X_1 & X_1^T X_2 & \dots & X_1^T X_n^T \\ X_2^T X_1 & X_2^T X_2 & & \vdots \\ \vdots & & \ddots & \\ X_n^T X_1 & \dots & & X_n^T X_n \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}[X_1^T X_1] & \mathrm{E}[X_1^T X_2] & \dots & \mathrm{E}[X_1^T X_n] \\ \mathrm{E}[X_2^T X_1] & \mathrm{E}[X_2^T X_2] & & \vdots \\ \vdots & & \ddots & \\ \mathrm{E}[X_n^T X_1] & \dots & & \mathrm{E}[X_n^T X_n] \end{bmatrix}.\end{aligned} \tag{6.11}$$

We can now apply Lemma 6.1 to each entry of the matrix and see that

$$\begin{aligned}\mathrm{E}\left[X^T X\right] &= \begin{bmatrix} \mathrm{E}[X_1^T]\,\mathrm{E}[X_1] + \mathrm{tr}(\mathrm{Cov}[X_1]) & \mathrm{E}[X_1^T]\,\mathrm{E}[X_2] & \dots & \mathrm{E}[X_1^T]\,\mathrm{E}[X_n] \\ \mathrm{E}[X_2^T]\,\mathrm{E}[X_1] & \ddots & & \vdots \\ \vdots & & \ddots & \\ \mathrm{E}[X_n^T]\,\mathrm{E}[X_1] & \dots & & \mathrm{E}[X_n^T]\,\mathrm{E}[X_n] + \mathrm{tr}(\mathrm{Cov}[X_n]) \end{bmatrix} \\ &= \mathrm{E}[X]^T\,\mathrm{E}[X] + \mathrm{diag}(\mathrm{tr}(\mathrm{Cov}[X_i]))_i,\end{aligned} \tag{6.12}$$

where we have used the independence of the columns to drop the covariance of the off-diagonal terms. □

The expected value of the stochastic Gramian can be broken into two parts, one of which corresponds roughly to the deterministic Gramian (calculated from the mean *output* trajectory), and a second which captures most of the stochastic actuation of the system.

Theorem 6.3. *Let* $\bar{W}_o^\varepsilon(t_1, x_0, u)$ *be the matrix defined by*

$$\left(\bar{W}_o^\varepsilon\right)_{ij} = \frac{1}{4\varepsilon^2} \int_0^{t_1} \left(\mathrm{E}[y^{+i}] - \mathrm{E}[y^{-i}]\right)^T \left(\mathrm{E}[y^{+j}] - \mathrm{E}[y^{-j}]\right) dt \tag{6.13}$$

and let $\hat{W}_o^\varepsilon(t_1, x_0, u)$ *be the diagonal matrix defined by*

$$(\hat{W}_o^\varepsilon)_{ii} = \frac{1}{4\varepsilon^2} \int_0^{t_1} \mathrm{tr}\left(\mathrm{Cov}[y^{+i}] + \mathrm{Cov}[y^{-i}]\right) dt. \tag{6.14}$$

Then $\mathrm{E}[W_o^\varepsilon(t_1, x_0, u)] = \bar{W}_o^\varepsilon(t_1, x_0, u) + \hat{W}_o^\varepsilon(t_1, x_0, u)$.

Proof. By definition,

$$(W_o^\varepsilon)_{ij} = \frac{1}{4\varepsilon^2} \int_0^{t_1} \left(y^{+i} - y^{-i}\right)^T \left(y^{+j} - y^{-j}\right) dt, \tag{6.15}$$

where the t_1, x_0, and u arguments have been dropped for brevity. Taking the expectation on both sides, we get

$$\begin{aligned} \mathrm{E}\left[(W_o^\varepsilon)_{ij}\right] &= \mathrm{E}\left[\frac{1}{4\varepsilon^2} \int_0^{t_1} \left(y^{+i} - y^{-i}\right)^T \left(y^{+j} - y^{-j}\right) dt\right] \\ &= \frac{1}{4\varepsilon^2} \int_0^{t_1} \mathrm{E}\left[\left(y^{+i} - y^{-i}\right)^T \left(y^{+j} - y^{-j}\right)\right] dt. \end{aligned} \tag{6.16}$$

When $i \neq j$ the sample trajectories y^{+i} and y^{-i} are independent of y^{+j} and y^{-j}, so for off-diagonal terms of the Gramian we get

$$\mathrm{E}[(W_o^\varepsilon)_{ij}] = \frac{1}{4\varepsilon^2} \int_0^{t_1} \left(\mathrm{E}[y^{+i}] - \mathrm{E}[y^{-i}]\right)^T \left(\mathrm{E}[y^{+j}] - \mathrm{E}[y^{-j}]\right) dt. \tag{6.17}$$

Clearly, when $i = j$, independence does not hold. By Lemma 6.1, the diagonal terms of the Gramian become

$$\begin{aligned} \mathrm{E}[(W_o^\varepsilon)_{ii}] = \frac{1}{4\varepsilon^2} \int_0^{t_1} &\left(\mathrm{E}[y^{+i}] - \mathrm{E}[y^{-i}]\right)^T \left(\mathrm{E}[y^{+i}] - \mathrm{E}[y^{-i}]\right) dt \\ &+ \frac{1}{4\varepsilon^2} \int_0^{t_1} \mathrm{tr}\left(\mathrm{Cov}[y^{+i} + y^{-i}]\right) dt. \end{aligned} \tag{6.18}$$

We can break these diagonal terms into two parts: one that depends on the variance of the samples and one that does not. Note that the first term of (6.18) matches the right-hand side of (6.13). We can break down the second term of (6.18) slightly further by noting that y^{+i} and y^{-i} are independent, meaning that

$$\mathrm{Cov}[y^{+i} - y^{-i}] = \mathrm{Cov}[y^{+i}] + \mathrm{Cov}[y^{-i}]. \tag{6.19}$$

Therefore, the second term of (6.18) is just $(\hat{W}_o^{\varepsilon})_{ii}$. Thus, we have shown that

$$\mathrm{E}[W_o^{\varepsilon}(t_1, x_0, u)] = \bar{W}_o^{\varepsilon}(t_1, x_0, u) + \hat{W}_o^{\varepsilon}(t_1, x_0, u). \tag{6.20}$$

□

For arbitrary nonlinear systems we cannot generally go any further than this theorem in closed-form, because for nonlinear measurement functions, $h(X)$, we cannot move the expectation inside the function, i.e., $\mathrm{E}[Y(t, x_0, u)] \neq h(\mathrm{E}[X(t, x_0, u])$. However, when the output is linear ($Y = CX$), we can do so, and the $\bar{W}_o^{\varepsilon}$ term becomes the Gramian of the expected trajectory. For similar reasons, analytically computing the higher moments of the Gramian is not generally possible.

Note that the second term of the expected Gramian contains all of the variance between the sample trajectories. Each term of the expected Gramian is positive semidefinite, and $\hat{W}_o^{\varepsilon}$ is strictly positive definite whenever the system has no states that are decoupled from states with non-zero process noise input. As a result, the $\hat{W}_o^{\varepsilon}$ term, which captures much of the process noise influence, cannot increase the local unobservability index of the expected Gramian, though the estimation condition number can be increased or decreased by $\hat{W}_o^{\varepsilon}$.

Theorem 6.3 also provides insight into the effect of measurement noise on the empirical observability Gramian. Provided that the measurement noise is independent, zero-mean, additive noise, then the measurement noise will cancel out of the $\bar{W}_o^{\varepsilon}$ term, and contribute only to the $\hat{W}_o^{\varepsilon}$ term. Thus measurement noise can only affect the diagonal entries of the mean of the Gramian, with increasing noise covariance increasing the eigenvalues of the mean of the Gramian.

In practice, computing the expected Gramian from ensembles produced by Monte Carlo simulation may be more efficient than attempting to find the mean and covariance of the perturbed output trajectories themselves. Such an approach would also allow the calculation of higher moments of the Gramian simultaneously. However, Theorem 6.3 provides useful insight into the way in which process noise can influence and improve nonlinear observability.

6.2.1 Linear stochastic systems with additive noise

While we cannot generally proceed analytically any further than Theorem 6.3, for linear stochastic systems we can compute $\mathrm{E}[W_o^\varepsilon]$ directly. In particular, for the Ornstein-Uhlenbeck process with scalar state,

$$\Sigma_{ou}: \quad \begin{aligned} dX &= -aX\,dt + \sigma dW \\ Y &= X, \end{aligned} \tag{6.21}$$

we can compute the expected Gramian completely.

Corollary 6.4. *For the system Σ_{ou} with $a \neq 0$,*

$$\mathrm{E}[W_o^\varepsilon(t_1, x_0, u)] = \frac{1}{8\varepsilon^2 a}\left(\frac{\sigma^2 + 4a\varepsilon^2}{a} + 2\,\mathrm{Cov}[X(0)]\right)\left(1 - e^{-2at_1}\right) + \frac{\sigma^2 t_1}{4\varepsilon^2 a}. \tag{6.22}$$

Proof. For the first moment of Σ_{ou} we can write

$$\frac{d}{dt}\,\mathrm{E}[X] = -a\,\mathrm{E}[X], \tag{6.23}$$

which has the solution

$$\mathrm{E}[Y(t)] = \mathrm{E}[X(t)] = e^{-at}\,\mathrm{E}[X(0)]. \tag{6.24}$$

Similarly, the second moment of Σ_{ou} has dynamics

$$\frac{d}{dt}\,\mathrm{E}[X^2] = -2a\,\mathrm{E}[X^2] + \sigma^2, \tag{6.25}$$

which has the solution

$$\mathrm{E}[X(t)^2] = \frac{\sigma^2}{2a} - \left(\frac{\sigma^2}{2a} - \mathrm{E}[X(0)^2]\right)e^{-2at}. \tag{6.26}$$

Therefore,

$$\begin{aligned} \mathrm{Cov}[Y(t)] &= \mathrm{E}[X(t)^2] - \mathrm{E}[X(t)]^2 \\ &= \frac{\sigma^2}{2a} - \left(\frac{\sigma^2}{2a} + \mathrm{E}[X(0)]^2 - \mathrm{E}[X(0)^2]\right)e^{-2at}. \end{aligned} \tag{6.27}$$

By substituting $\mathrm{E}[X(0) \pm \varepsilon]$ and $\mathrm{E}[(X(0) \pm \varepsilon)^2]$ into (6.24) and (6.27) as appropriate, we can get

$$\bar{W}_o^{\varepsilon} = \frac{1}{2a}(1 - e^{-2at_1}) \tag{6.28}$$

and

$$\begin{aligned} \hat{W}_o^{\varepsilon} &= \frac{1}{8\varepsilon^2 a}\left(\frac{\sigma^2}{a} + 2\,\mathrm{E}[X(0)]^2 - 2\,\mathrm{E}[X(0)^2]\right)\left(1 - e^{-2at_1}\right) + \frac{\sigma^2 t_1}{4\varepsilon^2 a} \\ &= \frac{1}{8\varepsilon^2 a}\left(\frac{\sigma^2}{a} + 2\,\mathrm{Cov}[X(0)]\right)\left(1 - e^{-2at_1}\right) + \frac{\sigma^2 t_1}{4\varepsilon^2 a}. \end{aligned} \tag{6.29}$$

Applying Theorem 6.3 completes the proof. □

Note that for a deterministic system (where $\sigma = 0$ and $\mathrm{Cov}[X(0)] = 0$), this quantity is just the ordinary empirical observability Gramian for a scalar linear system, as we would expect. The stochastic aspects of the Gramian in this case improve the local unobservability index by increasing the Gramian; the estimation condition number is always unity. We also note that we cannot take $\lim_{\varepsilon \to 0} \mathrm{E}[W_o^{\varepsilon}]$ for Σ_{ou}. This result is because $y^{+i} - y^{-i}$ does not go to 0 almost surely as $\varepsilon \to 0$, as would be necessary for $\lim_{\varepsilon \to 0} \mathrm{E}[W_o^{\varepsilon}]$ to be finite. We will see this property for all the stochastic systems we examine, but it does not pose a problem in connecting the Gramian to stochastic observability.

We can also compute the expectation of the Gramian for a non-scalar extension of the Ornstein-Uhlenbeck process given by

$$\Sigma_{OU} : \quad \begin{aligned} dX &= AX dt + \Omega dW \\ Y &= CX, \end{aligned} \tag{6.30}$$

though not quite to the same degree of simplicity as for the single-state system. However, the separation of the noise and deterministic components of the Gramian is still clear.

The stochastic observability of systems of this type has been studied extensively [29, 30, 32–35, 41, 45, 47, 49, 51, 52], though in all but [33], the analysis was in discrete-time. In [29] the system being studied was a hybrid system with additional quantized states that altered the continuous state dynamics, and in [45] the dynamics jumped between different linear systems at intervals given by a Markov chain. For more discussion on the limitations of, and differences between, these works, see Chapter 1.

Corollary 6.5. *For the system Σ_{OU},*

$$\begin{aligned}\mathrm{E}[W_o^\varepsilon(t_1, x_0, u)] = W_O(t_1) &+ \frac{1}{2\varepsilon^2} I \operatorname{tr}(W_O(t_1) \operatorname{Cov}[X(0)]) \\ &+ \frac{1}{2\varepsilon^2} I \int_0^{t_1} \int_0^t \operatorname{tr}(C e^{A(t-\tau)} \Omega \Omega^T e^{A^T(t-\tau)} C^T) d\tau dt,\end{aligned} \tag{6.31}$$

where W_O is the deterministic linear observability Gramian for the system

$$\begin{aligned}\dot{x} &= Ax \\ y &= Cx,\end{aligned} \tag{6.32}$$

and I is the $n \times n$ identity matrix.

Proof. For the first moment we have

$$\frac{d}{dt} \mathrm{E}[X] = A \, \mathrm{E}[X], \tag{6.33}$$

which has the solution

$$\mathrm{E}[X(t)] = e^{At} \, \mathrm{E}[X(0)]. \tag{6.34}$$

Therefore, by linearity of the expectation,

$$\mathrm{E}[Y^{\pm i}(t)] = C e^{At} (\mathrm{E}[X(0)] \pm \varepsilon e_i). \tag{6.35}$$

To compute the variance of $Y(t)$, we first need $\mathrm{E}[X(t) X^T(t)]$, which has dynamics

$$\frac{d}{dt} \mathrm{E}[X X^T] = A \, \mathrm{E}[X X^T] + \mathrm{E}[X X^T] A^T + \Omega \Omega^T, \tag{6.36}$$

a Riccati equation. The Riccati equation has a well-known solution (see [73])

$$\mathrm{E}[X X^T] = U_2 U_1^{-1} \tag{6.37}$$

where $U_1, U_2 \in \mathbb{R}^{n \times n}$ are governed by

$$\frac{d}{dt} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} -A^T & 0 \\ \Omega \Omega^T & A \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{6.38}$$

with initial conditions

$$\mathrm{E}[X(0)X^T(0)] = U_{2,0}U_{1,0}^{-1}. \tag{6.39}$$

Note that these initial conditions give only n^2 equations for $2n^2$ unknowns, however, $U_{1,0}$ will cancel out of our final result and can be chosen arbitrarily, provided it is chosen to be invertible.

We solve straightforwardly for U_1

$$U_1 = e^{-A^T t}U_{1,0}, \tag{6.40}$$

which we can plug into the dynamics of U_2 to get

$$\begin{aligned} \dot{U}_2 &= AU_2 + \Omega\Omega^T U_1 \\ &= AU_2 + \Omega\Omega^T e^{-A^T t}U_{1,0}. \end{aligned} \tag{6.41}$$

Solving this linear system, we get

$$\begin{aligned} U_2 &= e^{At}U_{2,0} + \int_0^t e^{A(t-\tau)}\Omega\Omega^T e^{-A^T\tau}U_{1,0}d\tau \\ &= e^{At}U_{2,0} + e^{At}\int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}d\tau U_{1,0}. \end{aligned} \tag{6.42}$$

Therefore, we have

$$\begin{aligned} \mathrm{E}[XX^T] &= e^{At}U_{2,0}U_{1,0}^{-1}e^{A^T t} + e^{At}\int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}d\tau e^{A^T t} \\ &= e^{At}\left(\mathrm{E}[X(0)X(0)^T] + \int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}d\tau\right)e^{A^T t}. \end{aligned} \tag{6.43}$$

Because Y is simply a linear function of X, we can write

$$\begin{aligned} \mathrm{Cov}[Y] &= C\,\mathrm{Cov}[X]C^T \\ &= C(\mathrm{E}[XX^T] - \mathrm{E}[X]\,\mathrm{E}[X]^T)C^T. \end{aligned} \tag{6.44}$$

Substituting in (6.34) and (6.43) we find that the covariance of a sample measurement at time t is given by

$$\begin{aligned} \mathrm{Cov}[Y(t)] &= Ce^{At}\left(\mathrm{E}[X(0)X(0)^T] - \mathrm{E}[X(0)]\,\mathrm{E}[X(0)]^T + \int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}d\tau\right)e^{A^T t}C^T \\ &= Ce^{At}\left(\mathrm{Cov}[X(0)]^T + \int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}d\tau\right)e^{A^T t}C^T. \end{aligned} \tag{6.45}$$

Note that

$$\begin{aligned}\mathrm{Cov}[X(0) \pm \varepsilon e_i] &= \mathrm{E}[(X(0) \pm \varepsilon e_i)(X(0) \pm \varepsilon e_i)^T] - \mathrm{E}[(X(0) \pm \varepsilon e_i)]\,\mathrm{E}[(X(0) \pm \varepsilon e_i)]^T \\ &= \mathrm{E}[X(0)X(0)^T] - \mathrm{E}[X(0)]\,\mathrm{E}[X(0)]^T \\ &= \mathrm{Cov}[X(0)].\end{aligned} \tag{6.46}$$

In other words, the covariance of the initial state distribution is not affected by perturbation. Because the covariance of the measurement, $Y(t)$, depends linearly on the covariance of the state initial condition, we find that

$$\mathrm{Cov}[Y^{\pm i}(t)] = \mathrm{Cov}[Y(t)], \tag{6.47}$$

i.e., the covariance of the measurement is also not affected by perturbations in the initial condition.

Now we can approach $\bar{W}_o^\varepsilon$ and $\hat{W}_o^\varepsilon$. First, looking at the first term,

$$\begin{aligned}(\bar{W}_o^\varepsilon)_{ij} &= \frac{1}{4\varepsilon^2}\int_0^{t_1} (2\varepsilon C e^{At} e_i)^T (2\varepsilon C e^{At} e_j) dt \\ &= \int_0^{t_1} e_i^T e^{A^T t} C^T C e^{At} e_j dt,\end{aligned} \tag{6.48}$$

so that

$$\begin{aligned}\bar{W}_o^\varepsilon &= \int_0^{t_1} e^{A^T t} C^T C e^{At} dt \\ &= W_O(t_1).\end{aligned} \tag{6.49}$$

The second term, $\hat{W}_o^\varepsilon$, is given component-wise by,

$$\begin{aligned}(\hat{W}_o^\varepsilon)_{ii} &= \frac{1}{2\varepsilon^2}\int_0^{t_1} \mathrm{tr}\left(C e^{At}\left(\mathrm{Cov}[X(0)] + \int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau} d\tau\right) e^{A^T t} C^T\right) dt \\ &= \frac{1}{2\varepsilon^2}\int_0^{t_1} \mathrm{tr}\left(e^{A^T t} C^T C e^{At}\left(\mathrm{Cov}[X(0)] + \int_0^t e^{-A\tau}\Omega\Omega^T e^{-A^T\tau} d\tau\right)\right) dt \\ &= \frac{1}{2\varepsilon^2}\mathrm{tr}\left(W_O\,\mathrm{Cov}[X(0)]\right) + \frac{1}{2\varepsilon^2}\int_0^{t_1}\int_0^t \mathrm{tr}\left(e^{A^T t} C^T C e^{At} e^{-A\tau}\Omega\Omega^T e^{-A^T\tau}\right) d\tau dt \\ &= \frac{1}{2\varepsilon^2}\mathrm{tr}\left(W_O\,\mathrm{Cov}[X(0)]\right) + \frac{1}{2\varepsilon^2}\int_0^{t_1}\int_0^t \mathrm{tr}\left(C e^{A(t-\tau)}\Omega\Omega^T e^{A^T(t-\tau)} C^T\right) d\tau dt.\end{aligned} \tag{6.50}$$

The corollary then follows from Theorem 6.3. □

Note that (6.31) can be re-written as

$$\mathrm{E}[W_o^\varepsilon(t_1, x_0, u)] = W_O(t_1) + \frac{1}{2\varepsilon^2} I \,\mathrm{tr}(W_O(t_1)\,\mathrm{Cov}[X(0)]) + \frac{1}{2\varepsilon^2} I \int_0^{t_1} \mathrm{tr}(CW_C(t)C^T) d\tau dt, \tag{6.51}$$

where $W_C(t)$ is the controllability Gramian of the linear system

$$\Sigma_\Omega : \quad \begin{aligned} \dot{x} &= Ax + \Omega u \\ y &= Cx, \end{aligned} \tag{6.52}$$

or as

$$\mathrm{E}[W_o^\varepsilon(t_1, x_0, u)] = W_O(t_1) + \frac{1}{2\varepsilon^2} I \,\mathrm{tr}(W_O(t_1)\,\mathrm{Cov}[X(0)]) + \frac{1}{2\varepsilon^2} I \int_0^{t_1} \mathrm{tr}(W_O(t)\Omega\Omega^T) d\tau dt. \tag{6.53}$$

We can interpret $W_C(t_1)$ as noise transfer from control to state, meaning that stochastic observability is influenced by the noise-to-output power.

As in the single-state system, we note that more noise (larger Ω) never decreases the positive-definiteness of the expected Gramian, though as we show in the next chapter, the effect of noise on the estimation condition number is not necessarily monotonic. In particular, the expected Gramian for Σ_{OU} can be positive definite even when the deterministic linear component of the system is not observable. Furthermore, the effect of noise on the Gramian increases as $\varepsilon \to 0$. Increasing the initial covariance of the stochastic state, $\mathrm{Cov}[X(0)]$, also increases the positive-definiteness of the expected Gramian. Because $\hat{W}_o^\varepsilon \succcurlyeq 0$, the expected value of the Gramian for Σ_{OU} will always be strictly positive definite when the deterministic component of the system is observable. If W_O is not full rank, then W_O and $\hat{W}_o^\varepsilon$ must have non-intersecting null-spaces (except at the origin) in order for $\mathrm{E}[W_o^\varepsilon]$ to be full rank.

6.2.2 Linear stochastic systems with multiplicative noise

We now move to another stochastic variant of the classic LTI dynamics

$$\Sigma_{BS} : \quad \begin{aligned} dX &= AXdt + \sum_{j=1}^{m} \Omega_j X dw_j \\ Y &= CX, \end{aligned} \tag{6.54}$$

where the noise now depends multiplicatively on the state and dw_j are the independent components of the process noise dW. The stochastic observability of this system has been studied in continuous and discrete-time, usually with the addition of Markovian jumps between a finite set of dynamics (A_i, Ω_i, C_i) in [36–40, 53, 54, 58]. We will restrict ourselves to the simple LTI case for this dissertation.

Note that one difference in these dynamics from the additive noise case discussed previously is that, once the system reaches equilibrium, it will remain there. Because

$$\frac{d}{dt}\,\mathrm{E}[X] = A\,\mathrm{E}[X] \tag{6.55}$$

we see that the system is stable in expectation when A is Hurwitz. In such a case we expect that the system state will eventually go to zero.

We define the operator $\mathrm{vec}(\cdot)$, which maps an $n \times n$ matrix to an $n^2 \times 1$ vector by stacking the columns of the matrix, and the operator vec^{-1}, the inverse operation. Note that both operators are linear. We use $\otimes$ to denote the Kronecker product, and $\oplus$ to denote the Kronecker sum of matrices.

Corollary 6.6. *For the system Σ_{BS},*

$$\mathrm{E}[W_o^{\varepsilon}(t_1, x_0, u)] = W_O(t_1) + \frac{1}{2}\int_0^{t_1} \mathrm{diag}\left(\mathrm{tr}\left(\mathrm{vec}^{-1}\left(\omega_i\right)\right)\right)_i dt, \tag{6.56}$$

where W_O is the deterministic linear observability Gramian for the system

$$\begin{aligned} \dot{x} &= Ax \\ y &= Cx, \end{aligned} \tag{6.57}$$

$Q = A \oplus A + \sum_{j=1}^{m} \Omega_j \otimes \Omega_j$, *and*

$$\begin{aligned} \omega_i = (C \otimes C)\Bigg(&\frac{1}{\varepsilon^2} e^{Qt}\,\mathrm{vec}(\mathrm{E}[X(0)X(0)^T]) + e^{Qt}\,\mathrm{vec}(e_i e_i^T) \\ &- \frac{1}{\varepsilon^2} e^{(A\oplus A)t}\,\mathrm{vec}(\mathrm{E}[X(0)]\,\mathrm{E}[X(0)]^T) - e^{(A\oplus A)t}\,\mathrm{vec}(e_i e_i^T)\Bigg). \end{aligned} \tag{6.58}$$

Proof. As before, for the first moment we have

$$\frac{d}{dt}\,\mathrm{E}[X] = A\,\mathrm{E}[X], \tag{6.59}$$

which has the solution

$$\mathrm{E}[X(t)] = e^{At}\,\mathrm{E}[X(0)]. \tag{6.60}$$

Therefore, by linearity of the expectation,

$$\mathrm{E}[Y^{\pm i}(t)] = Ce^{At}(\mathrm{E}[X(0)] \pm \varepsilon e_i). \tag{6.61}$$

To compute $\hat{W}_o^\varepsilon$, we need the covariance of $Y(t)$, for which we first need $\mathrm{E}[X(t)X^T(t)]$, which has dynamics

$$\frac{d}{dt}\,\mathrm{E}[XX^T] = A\,\mathrm{E}[XX^T] + \mathrm{E}[XX^T]A^T + \sum_{j=1}^{m} \Omega_j\,\mathrm{E}[XX^T]\Omega_j^T. \tag{6.62}$$

We can simplify this equation by making use of the identity $\mathrm{vec}(\sum_{j=1}^m \Omega_j\,\mathrm{E}[XX^T]\Omega_j^T) = \sum_{j=1}^m (\Omega_j \otimes \Omega_j)\,\mathrm{vec}(\mathrm{E}[XX^T])$. Applying the identity, we get

$$\begin{aligned}
\frac{d}{dt}\,\mathrm{vec}(\mathrm{E}[XX^T]) &= \mathrm{vec}(A\,\mathrm{E}[XX^T]) + \mathrm{vec}(\mathrm{E}[XX^T]A^T) + \sum_{j=1}^{m} \mathrm{vec}(\Omega_j\,\mathrm{E}[XX^T]\Omega_j^T) \\
&= \left(A \otimes I + I \otimes A + \sum_{j=1}^{m} \Omega_j \otimes \Omega_j\right) \mathrm{vec}(\mathrm{E}[XX^T]) \\
&= \left(A \oplus A + \sum_{j=1}^{m} \Omega_j \otimes \Omega_j\right) \mathrm{vec}(\mathrm{E}[XX^T]) \\
&= Q\,\mathrm{vec}(\mathrm{E}[XX^T]).
\end{aligned} \tag{6.63}$$

Therefore,

$$\mathrm{vec}(\mathrm{E}[XX^T]) = e^{Qt}\,\mathrm{vec}(\mathrm{E}[X(0)X(0)^T]). \tag{6.64}$$

Using the vec identity again, we can write

$$\begin{aligned}
\mathrm{vec}(\mathrm{Cov}[Y]) &= \mathrm{vec}(C\,\mathrm{Cov}[X]C^T) \\
&= (C \otimes C)(\mathrm{vec}(\mathrm{E}[XX^T]) - \mathrm{vec}(\mathrm{E}[X]\,\mathrm{E}[X]^T)).
\end{aligned} \tag{6.65}$$

Substituting in (6.60) and (6.64) we find that the covariance of a sample measurement at time t is given by

$$\operatorname{vec}(\operatorname{Cov}[Y(t)]) = (C \otimes C)\left(e^{Qt}\operatorname{vec}(\operatorname{E}[X(0)X(0)^T]) - e^{(A\oplus A)t}\operatorname{vec}(\operatorname{E}[X(0)]\operatorname{E}[X(0)]^T)\right) \quad (6.66)$$

The covariance of the perturbed measurements is given by

$$\begin{aligned}\operatorname{vec}(\operatorname{Cov}[Y^{\pm i}(t)]) =&(C \otimes C)\left(e^{Qt}\operatorname{vec}(\operatorname{E}[X(0)X(0)^T] \pm \varepsilon e^{Qt}\operatorname{vec}(e_i\operatorname{E}[X(0)]^T)\right.\\ &\pm \varepsilon e^{Qt}\operatorname{vec}(\operatorname{E}[X(0)]e_i^T) + \varepsilon^2 e^{Qt}\operatorname{vec}(e_ie_i^T)\\ &- e^{(A\oplus A)t}\operatorname{vec}(\operatorname{E}[X(0)]\operatorname{E}[X(0)]^T)) \mp \varepsilon e^{(A\oplus A)t}\operatorname{vec}(e_i\operatorname{E}[X(0)]^T)))\\ &\left.\mp\varepsilon e^{(A\oplus A)t}\operatorname{vec}(\operatorname{E}[X(0)]e_i^T) - \varepsilon^2 e^{(A\oplus A)t}\operatorname{vec}(e_ie_i^T)\right)\end{aligned} \quad (6.67)$$

Now we can solve for $\bar{W}_o^\varepsilon$ and $\hat{W}_o^\varepsilon$. As before, the first term is given by

$$\begin{aligned}(\bar{W}_o^\varepsilon)_{ij} &= \frac{1}{4\varepsilon^2}\int_0^{t_1}(2\varepsilon Ce^{At}e_i)^T(2\varepsilon Ce^{At}e_j)dt\\ &= \int_0^{t_1} e_i^T e^{A^Tt}C^TCe^{At}e_j dt,\end{aligned} \quad (6.68)$$

giving

$$\begin{aligned}\bar{W}_o^\varepsilon &= \int_0^{t_1} e^{A^Tt}C^TCe^{At}dt\\ &= W_O(t_1).\end{aligned} \quad (6.69)$$

The $\hat{W}_o^\varepsilon$ term has entries given by

$$\begin{aligned}(\hat{W}_o^\varepsilon)_{ii} = \frac{1}{2}\int_0^{t_1}\operatorname{tr}\Bigg(\operatorname{vec}^{-1}\Bigg((C\otimes C)\Bigg(&\frac{1}{\varepsilon^2}e^{Qt}\operatorname{vec}(\operatorname{E}[X(0)X(0)^T]) + e^{Qt}\operatorname{vec}(e_ie_i^T)\\ &-\frac{1}{\varepsilon^2}e^{(A\oplus A)t}\operatorname{vec}(\operatorname{E}[X(0)]\operatorname{E}[X(0)]^T) - e^{(A\oplus A)t}\operatorname{vec}(e_ie_i^T)\Bigg)\Bigg)\Bigg)\,dt\end{aligned} \quad (6.70)$$

The corollary then follows from Theorem 6.3. □

Unlike the case of the additive noise Ornstein-Uhlenbeck model, the condition number of the $\hat{W}_o^\varepsilon$ matrix is not always unity for Σ_{BS}. When the noise depends on the state, the direction of the initial condition perturbation matters in the covariance of the output. As before, the effect on the Gramian of noise increases as $\varepsilon \to 0$, but in the case of multiplicative noise, the effect is removed when the Gramian is evaluated with a Dirac delta initial condition at the origin, for which $\mathrm{E}[X(0)X(0)^T] = \mathrm{E}[X(0)]\,\mathrm{E}[X(0)]^T = 0$.

Remark 6.1. While we felt that, in this case, the notation would be less cumbersome and the intuition more clear if we restricted ourselves to the time-invariant system, Corollary 6.6 applies to the LTV version of Σ_{BS} as well. The primary change required would be to replace each matrix exponential with the fundamental matrix of the corresponding LTV system and to replace the $e^{(A\oplus A)t}$ term with a Kronecker product and fundamental matrices.

6.3 Stochastic Observability

We are now ready to present one of the main results of this chapter: a rank condition for the expected value of the empirical observability Gramian for stochastic observability. The definition of stochastic observability that we use here originates with Dragan and Morozan [36] for linear stochastic systems. Writing that definition in our notation, we say that the system Σ_{BS} is *stochastically observable* if there exists $\beta > 0$ and $t_1 > 0$ such that

$$\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)dt\right] \succcurlyeq \beta I \tag{6.71}$$

where $\Psi(t,t_0)$ is the fundamental matrix solution of Σ_{BS}. Note that this definition is a slight simplification of the original, which applies to LTV systems with multiplicative noise and Markovian switches in the system matrices.

In our non-switching system, we can expand the left-hand side of (6.71) in more detail. The fundamental matrix solution of a stochastic linear system is itself a random variable, defined such that

$$X(t) = \Psi(t,0)X(0), \tag{6.72}$$

i.e., the random variable of the state at time t is the product of the random $\Psi(t,0)$ and the random initial state, which we will assume to be independently distributed. We can derive the following properties of the random fundamental matrix,

$$\mathrm{E}[X(t)] = \mathrm{E}[\Psi(t,0)]\,\mathrm{E}[X(0)] \tag{6.73}$$

$$\Psi(t,t) = I \quad \text{w.p. 1.} \tag{6.74}$$

We can also see that the columns of $\Psi(t,0)$ must be independent, because the i-th column is simply the solution of the system with initial condition e_i and the solutions of the system from independent initial conditions must be independent.

Therefore, expanding and applying Lemma 6.2, we get

$$\begin{aligned}
\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)ds\right] &= \int_0^{t_1} \mathrm{E}\left[\Psi^T(t,0)C(t)^T C(t)\Psi(t,0)\right] dt \\
&= \int_0^{t_1} \mathrm{E}\left[C(t)\Psi(t,0)\right]^T \mathrm{E}\left[C(t)\Psi(t,0)\right] dt \\
&\quad + \int_0^{t_1} \operatorname{diag}\left(\operatorname{tr}\left(\operatorname{Cov}\left[C(t)\Psi(t,0)e_i\right]\right)\right)_i dt \\
&= \int_0^{t_1} \mathrm{E}[\Psi^T(t,0)]C(t)^T C(t)\,\mathrm{E}[\Psi(t,0)]dt \\
&\quad + \int_0^{t_1} \operatorname{diag}\left(\operatorname{tr}\left(\operatorname{Cov}\left[C(t)\Psi(t,0)e_i\right]\right)\right)_i dt \\
&= W_O(t_1) + \int_0^{t_1} \operatorname{diag}\left(\operatorname{tr}\left(\operatorname{Cov}\left[C(t)\Psi(t,0)e_i\right]\right)\right)_i dt.
\end{aligned} \tag{6.75}$$

Taking a closer look at the second term, we get

$$\begin{aligned}
\operatorname{Cov}[C(t)\Psi(t,0)e_i] =& C(t)\operatorname{Cov}[\Psi(t,0)e_i]C(t)^T \\
=& C(t)\operatorname{Cov}[X^{+i}(t)]C(t)^T
\end{aligned} \tag{6.76}$$

or, substituting from (6.66),

$$\operatorname{vec}(\operatorname{Cov}[C(t)\Psi(t,0)e_i]) = (C(t)\otimes C(t))\left(e^{Qt}\operatorname{vec}(e_i e_i^T) - e^{(A\oplus A)t}\operatorname{vec}(e_i e_i^T)\right). \tag{6.77}$$

We are now ready to state the main result of this chapter.

Theorem 6.7. *The system Σ_{BS} is stochastically observable in the sense of [36] if and only if*

$$\operatorname{rank}(\mathrm{E}[W_o^\varepsilon(t_1, \delta(0), 0)]) = n \tag{6.78}$$

for some $t_1 > 0$, where $\delta(x)$ is the Dirac distribution centered at x.

Proof. To begin, we will look at $\mathrm{E}[W_o^\varepsilon(t_1, \delta(0), 0)]$. Substituting $X(0) = \delta(0)$ into Corollary 6.6, we get

$$\begin{aligned}\mathrm{E}[W_o^\varepsilon(t_1, \delta(0), 0)] &= W_O(t_1) \\ &\quad + \frac{1}{2}\int_0^{t_1} \operatorname{diag}\left(\operatorname{tr}\left(\operatorname{vec}^{-1}\left((C \otimes C)\left(e^{Qt}\operatorname{vec}(e_i e_i^T) - e^{(A\oplus A)t}\operatorname{vec}(e_i e_i^T)\right)\right)\right)\right) dt,\end{aligned} \tag{6.79}$$

which we note closely matches $\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)dt\right]$ except for the factor of $\frac{1}{2}$. In fact

$$\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)dt\right] = \mathrm{E}[W_o^\varepsilon] + \hat{W}_o^\varepsilon, \tag{6.80}$$

where the arguments to the empirical observability Gramian have been dropped for clarity.

Now, if $\operatorname{rank}(\mathrm{E}[W_o^\varepsilon]) = n$ for some $t_1 > 0$, then $\mathrm{E}[W_o^\varepsilon] \succ 0$ because W_o^ε is symmetric and positive semidefinite by construction.

Applying Weyl's inequality, we get

$$\underline{\lambda}\left(\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)dt\right]\right) \geq \underline{\lambda}(\mathrm{E}[W_o^\varepsilon]) + \underline{\lambda}(\hat{W}_o^\varepsilon). \tag{6.81}$$

Let $\beta = \underline{\lambda}(\mathrm{E}[W_o^\varepsilon]) + \underline{\lambda}(\hat{W}_o^\varepsilon)$. We know that $\underline{\lambda}(\mathrm{E}[W_o^\varepsilon]) > 0$, and $\underline{\lambda}(\hat{W}_o^\varepsilon) \geq 0$, so it follows that $\beta > 0$. Then

$$\mathrm{E}\left[\int_0^{t_1} \Psi^T(t,0)C(t)^T C(t)\Psi(t,0)dt\right] \succcurlyeq \beta I, \tag{6.82}$$

and the system is stochastically observable by the definition of [36].

Now we consider the reverse case, and assume that the system is stochastically observable in the sense of [36]. Assume that $\mathrm{E}[W_o^\varepsilon]$ is not strictly positive definite. It follows that neither

W_O or $\hat{W}_o^\varepsilon$ can be strictly positive definite either, and that there exists at least one vector, η, that lies in the null spaces of both W_O and $\hat{W}_o^\varepsilon$. However, in that case,

$$\begin{aligned} \eta^T \mathrm{E}\left[\int_0^{t_1} \phi^T(s,t) C^T C \phi(s,t) ds\right] \eta &= \eta^T W_O \eta + 2\eta^T \hat{W}_o^\varepsilon \eta \\ &= 0. \end{aligned} \tag{6.83}$$

However, that means that there can be no $\beta > 0$ that satisfies (6.71), which contradicts our assumption of stochastic observability. Thus, by contradiction, $\mathrm{E}[W_o^\varepsilon] \succ 0$, or $\mathrm{rank}(\mathrm{E}[W_o^\varepsilon]) = n$. □

We note that the definition of stochastic observability from [36] does not readily extend to nonlinear systems, as it depends on the fundamental matrix, which has no analog in nonlinear systems. The empirical observability Gramian extends naturally to nonlinear stochastic systems, however, meaning that the rank of the Gramian can be used for a definition of stochastic nonlinear observability that is equivalent to existing definitions of both stochastic linear observability, deterministic linear observability, and, as shown in Chapter 3, partially equivalent to definitions of weak observability of deterministic nonlinear systems.

In Theorem 6.7 the ε terms cancel out, just as they do in linear deterministic systems, but in the general nonlinear stochastic case we cannot assume that this cancellation would take place. In Chapter 3, the rank of the empirical Gramian in the limit as $\varepsilon \to 0$ was used to demonstrate weak observability of nonlinear deterministic systems. However, as we mentioned after Theorem 6.3, for nonlinear stochastic systems, the limit as $\varepsilon \to 0$ does not exist. As a result, a singular value condition, similar to Theorem 3.4, proportional to $\varepsilon^2\tau$ would be a more appropriate way of defining stochastic nonlinear observability with the expectation of the empirical observability Gramian.

6.3.1 Example

Consider an example system

$$\Sigma_1 : \quad \begin{aligned} dX &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} X dt + \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} X dW \\ Y &= \begin{bmatrix} 1 & 0 \end{bmatrix} X, \end{aligned} \tag{6.84}$$

Examining the deterministic component of this system, we see that the linear observability Gramian of the system is

$$W_O(\tau) = \frac{1}{2} \begin{bmatrix} 1 - e^{-2\tau} & 0 \\ 0 & 0 \end{bmatrix}. \tag{6.85}$$

Clearly, Σ_1 is not deterministically observable. Of course, this property is visible at a glance from the system dynamics as well, because the states are decoupled and the output includes only a single state.

Letting $\Omega_{11} = \Omega_{12} = \Omega_{22} = 1$ and $\Omega_{21} = 0$, the expected value of the empirical observability Gramian with initial distribution $\delta(0)$ is

$$\mathrm{E}[W_o^\varepsilon(\tau, \delta(0), 0)] = \begin{bmatrix} 1 - e^{-\tau} & 0 \\ 0 & 3 - ((\tau + 1) + (\tau + 1)^2)e^{-\tau} \end{bmatrix} \tag{6.86}$$

which is positive definite for all $\tau > 0$. Therefore Σ_1 is stochastically observable. Intuitively, we can measure the covariance of the noise in the x_1 state and how that changes to determine x_2, though only up to magnitude (a sign ambiguity would remain). Sign ambiguities are not necessarily a problem for observability in principle; weak observability of nonlinear deterministic systems, for example, only requires distinguishability from neighboring points, and that is allowed when the state is non-zero we can distinguish magnitude. From (6.84) we see that process noise clearly begins to influence the observability even in linear systems (though whether this system, with multiplicative noise, is truly still linear, or whether it has taken the first step into nonlinearity might, be a matter for debate).

If we consider the case that $\Omega_{12} = 0$, which completely decouples the dynamics, we find that the expected value of the Gramian is now

$$\mathrm{E}[W_o^\varepsilon(\tau,\delta(0),0)] = \begin{bmatrix} 1-e^{-\tau} & 0 \\ 0 & 0 \end{bmatrix}, \tag{6.87}$$

and that the system is not stochastically observable. Indeed, the same is true even if Ω_{21} is non-zero. This particular system is observable only when the x_2 state can influence the x_1 state, either directly through the deterministic dynamics, or less directly, through the noise.

The system remains observable even if the process noise of each component of the system is independent, with dynamics

$$\Sigma_{1a}: \quad \begin{aligned} dX &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} X dt + \begin{bmatrix} \Omega_{11} & 0 \\ 0 & 0 \end{bmatrix} X dW_1 + \begin{bmatrix} 0 & 0 \\ 0 & \Omega_{22} \end{bmatrix} X dW_2 + \begin{bmatrix} 0 & \Omega_{12} \\ 0 & 0 \end{bmatrix} X dW_3 \\ Y &= \begin{bmatrix} 1 & 0 \end{bmatrix} X. \end{aligned} \tag{6.88}$$

In this case, with $\Omega_{11} = \Omega_{12} = \Omega_{22} = 1$, the expected value of the Gramian of the system is

$$\mathrm{E}[W_o^\varepsilon(\tau,\delta(0),0)] = \begin{bmatrix} \frac{e^{-2\tau}(e^{\tau}-1)^2}{2} & 0 \\ 0 & 1-(\tau+1)e^{-\tau} \end{bmatrix}, \tag{6.89}$$

which is full rank for all $\tau > 0$.

6.4 Noise as Modeling Error

Noise enters our dynamics in a variety of ways. Thermal fluctuations, aerodynamic turbulence, ambient electrical and radiological effects can all create noise. These phenomena are also all examples of unmodeled dynamics, or modeling errors arising from simplifying approximations. To understand how noise-as-modeling-error can influence the observability of a system, we create a simplified example system based on a linearization of an nonlinear system.

Consider the nonlinear system

$$\Sigma_2 : \quad \begin{aligned} \dot{x} &= \begin{bmatrix} -x_1 + \frac{1}{2}x_2^2 \\ -x_2 \end{bmatrix} \\ y &= x_1, \end{aligned} \tag{6.90}$$

and its linearization at the equilibrium point $x = 0$

$$\Sigma_{2_0} : \quad \begin{aligned} \dot{x} &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} x \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x, \end{aligned} \tag{6.91}$$

From the Lie observability algebra of $\mathcal{O} = \{h, \mathcal{L}_f h\}$, we find that the system Σ_2 is observable when $x_2 \neq 0$, while clearly Σ_{2_0} is nowhere observable.

Now consider the stochastic system

$$\Sigma_{2s} : \quad \begin{aligned} dx &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} x dt + \begin{bmatrix} \frac{1}{2}x_2^2 \\ 0 \end{bmatrix} dW \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x, \end{aligned} \tag{6.92}$$

The noise in this case is proportional to the modeling error in the linearization. Note that this system is not linear, so the definition of stochastic observability given above does not strictly apply. Nevertheless, we can compute (numerically) the expected Gramian of the system to analyze observability. Numerically computing the empirical observability Gramian of Σ_{2s} from Monte Carlo simulation, we get an estimation condition number with a median value of 8.7 and an unobservability index with a median value of 20.5. Comparing these values to the deterministic estimation condition number of 22.5 and unobservability index of 40.2 for the system Σ_{2s}, we see that the observability of the stochastic approximation has metrics on the same order as the original system, and far more representative than the linearization Σ_{2_0}, for which the estimation condition number and local unobservability index are effectively infinite.

While this example may seem a bit contrived, it serves to illustrate how we can capture the influence on the observability of a nonlinear system of some kinds of unmodeled dynamics. In particular, we have shown that an unobservable deterministic approximation to an observable system, can be seen to be observable when the noise due to modeling error is accounted for. Of course, in practice our noise model will never be exactly proportional to modeling error because if we knew what the modeling error was, we could simply incorporate it into the model. However, in many cases we can experimentally determine approximate noise characteristics to use in an approximate model, without knowing an exact original model. This approximate noise model could then be used in observability analysis of the approximate system.

Chapter 7

SIMULATION

In general, as mentioned in Chapter 6, we cannot compute the empirical observability Gramian in closed-form for arbitrary nonlinear stochastic systems. However, the empirical Gramian does lend itself naturally to numerical computation by simulation. In the case of stochastic systems, this property means sampling the Gramian by drawing trajectory samples for each perturbation direction and using Monte Carlo methods to find Gramian statistics.

To illustrate the use of the empirical observability Gramian for observability of stochastic nonlinear systems, we have numerically computed ensembles of samples of the empirical observability Gramian for systems with process noise for three sample systems. The first is a simple approximately linear system, and the second a nonlinear unicycle model. Finally, we demonstrate the the observations that we develop for the simpler sample systems in the first two sections apply to a high fidelity model of lab-scale quadrotor vehicle. In each case we demonstrate the effect of process noise on the observability of the system.

We use the Euler-Maruyama method to integrate the stochastic differential equations and obtain sample trajectories. Euler-Maruyama is an extension of the forward Euler method for deterministic ordinary differential equations to the Itô calculus. A sample trajectory for Σ_S is given by

$$X_{t+1} = X_t + f(X_t, u_t)\Delta t + \sigma(X_t, u_t) Z_t \sqrt{\Delta t}, \tag{7.1}$$

where each $Z_t \in \mathbb{R}^q$ is independently distributed as $\mathcal{N}(0, I)$, X_0 is sampled from the initial condition distribution, and Δt is the discretization interval. The sample outputs, Y, are then given by

$$Y_t = h(X_t). \tag{7.2}$$

The sample trajectories of $y(t)$ can then be used to compute the empirical observability

Gramian, as described by the procedure of §6.1,

$$W_o^\varepsilon(t_1, x_0, u) = \frac{1}{4\varepsilon^2}\int_0^{t_1} \Phi^\varepsilon(t, x_0, u)^T \Phi^\varepsilon(t, x_0, u) dt, \tag{7.3}$$

where

$$\Phi^\varepsilon(t, x_0, u) = \begin{bmatrix} Y_t^{+1} - Y_t^{-1} & \cdots & Y_t^{+n} - Y_t^{-n} \end{bmatrix} \tag{7.4}$$

and $Y^{\pm i}(t)$ are independent sample trajectories of the system Σ_S, with control input $u(t)$, initialized from $X_0 \pm \varepsilon e_i$. Note that we compute Φ^ε only once for each Gramian, so that the $\Phi^{\varepsilon T}$ and Φ^ε terms of (7.3) are *not* independent of each other. While X_0 itself is a random variable, we will assume for the purposes of our simulations that its distribution is described by the Dirac probability density function $\delta(X_0) = x_0$, that is, a single point. This assumption can be relaxed, provided that the initial states of the sample trajectory are randomly chosen according to the desired initial condition distribution.

As discussed in Chapter 6, this modification results in a Gramian that is a random variable, and as a result, the local unobservability index and estimation condition number are also random variables. We can numerically approximate the distribution of the observability metrics by computing an ensemble of Gramians for a given initial condition. As the Gramian condition number and minimum singular value are both bounded below, we must note that the distributions of these variables will not be symmetrically distributed, even if the stochastic Gramian is.

7.1 Control and Noise Affine Dynamics

First we demonstrate a control and noise affine system system that is approximately linear. Consider the dynamics

$$\Sigma_3 : \quad \begin{aligned} dx &= \begin{bmatrix} -x_2 \\ x_1 u \end{bmatrix} dt + \begin{bmatrix} 0 \\ q x_1 \end{bmatrix} dW \\ y &= x_2. \end{aligned} \tag{7.5}$$

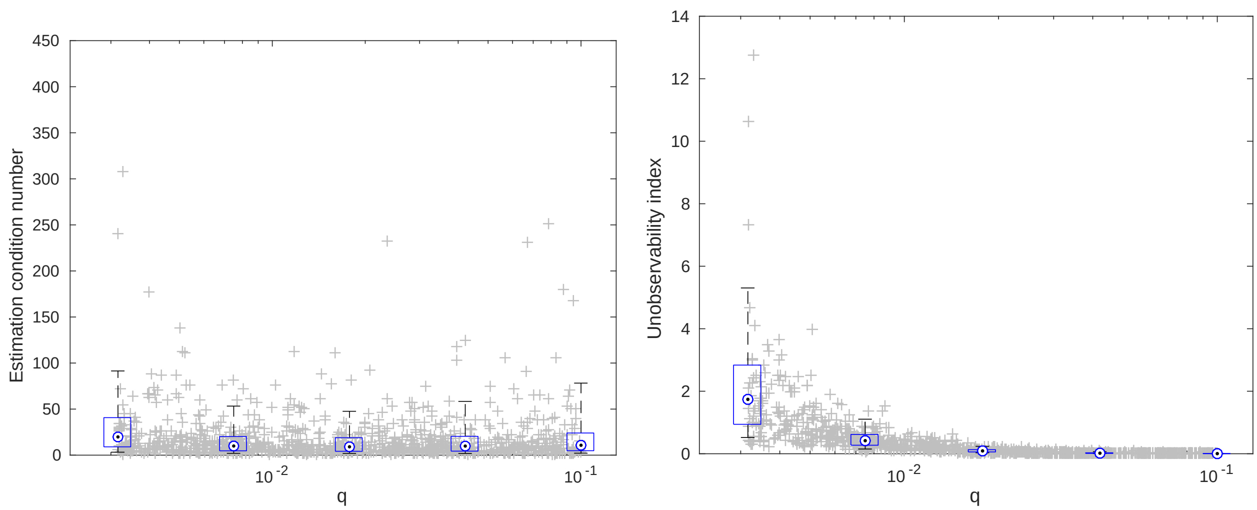
First, we note that when $q = 0$ (no process noise), the system is observable if and only if $u(t) \neq 0$. When $u(t) = 1$, we have a simple oscillator. Note also that this system does not quite meet the requirements of §6.3 because of the nonlinear x_1u term in $f(x, u)$.

When the empirical observability Gramian is computed with no noise ($q = 0$) and control $u(t) = 0.1$, we find that the local unobservability index is 2.012 and the condition number is 10.1. As expected when $q = 0$ and $u(t) = 0$, the local unobservability index is undefined (infinite) and the condition number is undefined. The system is linear for any particular constant control input, therefore these values are invariant to initial condition of the system.

If we consider the case $u(t) = 0$ and $q \neq 0$, we find that the estimation condition number and local unobservability index vary significantly depending on the sample trajectory. Figure 7.1 shows observability metrics for a range of q values. The gray points in the image are computed from Gramians with q values for Σ_3 sampled uniformly from a logarithmic q domain of $[10^{-3}, 10^{-1}]$, while the box plots are generated by an ensemble of 500 points at a fixed q. The density of the gray points represent the distribution of the observability metrics over q values and sample trajectories, while the box plots summarize the marginal distributions over the sample trajectories for a particular q. The boxes cover the second and third quartiles, and the center dot shows the median. The whiskers extend from the 5th to the 95th percentiles.

As the box plots show, the distribution of the metrics is strongly asymmetric, suggesting that mean and standard deviation are not the most pertinent descriptors of the values. Instead, we will use the median as our primary summary statistic. An important item to note from the Figure 7.1a, is that there appears to be a local minimum in the estimation condition number median as the noise variance, q, is varied. This minimum indicates that too little noise insufficiently actuates the system to produce observability (indeed, as $q \to 0$, the system becomes completely unobservable), while too much noise also impairs observability, perhaps by masking the actual dynamics of the system.

To compare the observability metrics resulting from control input and from noise, we



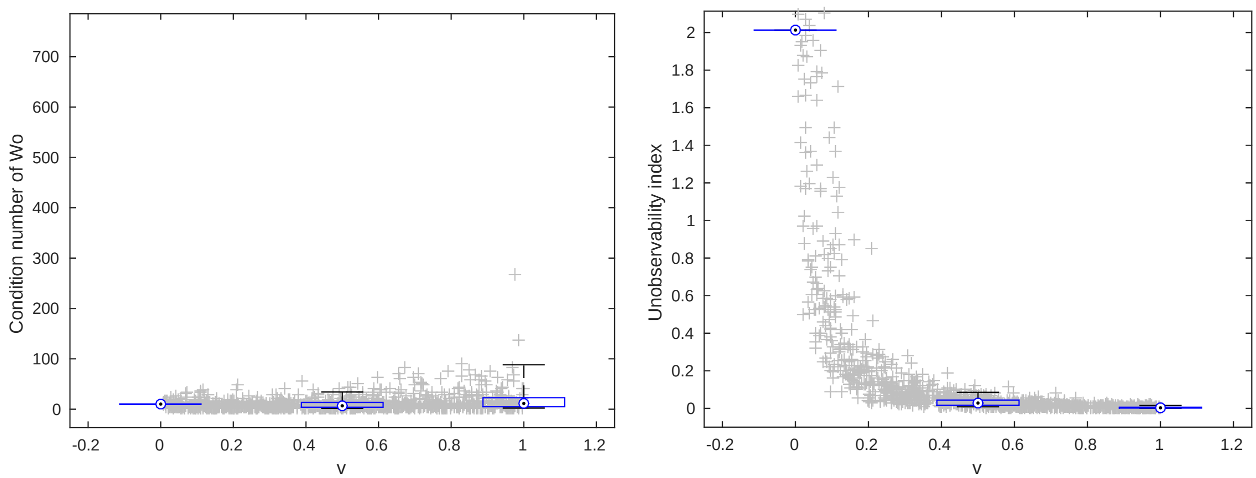
(a) (b)

Figure 7.1: Increasing noise monotonically decreases the expected unobservability index in the noise affine system. The estimation condition number for the linear system shows a minimum near a noise variance of $q = 0.02$.

computed the unobservability index and estimation condition number for the system

$$\Sigma_{3a} : \quad \begin{aligned} dx &= \begin{bmatrix} -x_2 \\ x_1(1-v)u \end{bmatrix} dt + \begin{bmatrix} 0 \\ vqx_1 \end{bmatrix} dW \\ y &= x_2. \end{aligned} \tag{7.6}$$

for $v \in [0, 1]$. The parameter v controls the trade-off between control and noise. When $v = 0$ there is no noise in the system, while for $v = 1$ there is no control input. As Figure 7.2 shows, for $u = 0.1$ and $q = 0.1$, noise produces similar levels of observability to control of similar amplitude, though with a much greater variation.



(a) (b)

Figure 7.2: Median estimation condition number improves slightly with mixed control and noise in the noise affine system, but pure control is better than pure noise. Moving the balance from control to noise provided a sharp decrease in the unobservability index overall.

7.2 Unicycle Dynamics

We can perform a similar set of computations for a nonlinear unicycle-type vehicle with position measurements. Consider the dynamics

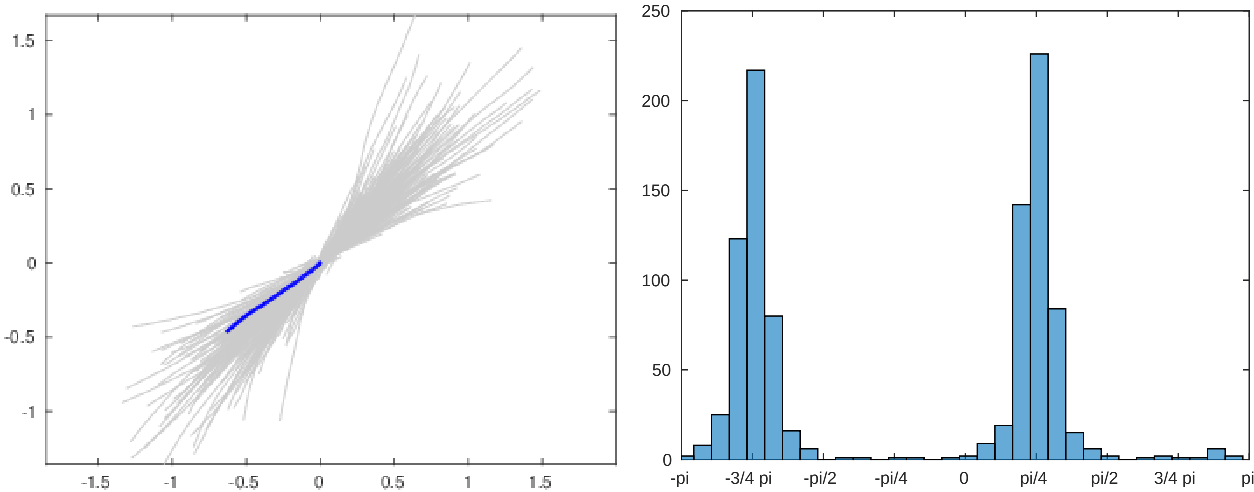
$$\Sigma_4 : \quad \begin{aligned} dx &= \begin{bmatrix} x_4 \cos(x_3) \\ x_4 \sin(x_3) \\ u_1 \\ u_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ 0 \\ 0 \\ q \end{bmatrix} dW \\ y &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \end{aligned} \tag{7.7}$$

Note that the deterministic component of the system is fully observable only when the vehicle is moving, $x_4 \neq 0$, or accelerating, $u_2 \neq 0$. If both $x_4 = 0$ and $u_2 = 0$, the heading angle, x_3, cannot be determined from the measurements, y. As before, we can compute Gramians for the system at equilibrium, $x = 0$ with no control input, $u(t) = 0$, across a range of noise values, q. We expect, in this case, that including noise in our analysis will reveal that the system is observable even at rest. We intuitively justify this expectation by noting that noise in the speed, x_4, will result in the vehicle jittering along the heading direction of the vehicle, providing information in the output about the heading that would not otherwise be available.

Figure 7.3 demonstrates this intuition by showing the measurement trajectories of a unicycle across 1000 sample runs, each starting at the origin with an initial heading of 45°. Note that for Figure 7.3 we have also added noise to the steering dynamics (x_3) in order to demonstrate that our intuition works even when the heading angle is further obscured. The highlighted trajectory from Figure 7.3a shows that for any particular sample, we can more-or-less judge the initial heading of the vehicle, which deterministic observability analysis indicates is unobservable. Figure 7.3b shows that, in ensemble, the initial heading can be seen in a maximum likelihood sense, up to a forward/backward ambiguity. Ensemble conclusions are probably not as useful in real-time estimation, as they require multiple runs from the system, but they can be useful to illustrate our intuition in this scenario. For real-time estimation, Figure 7.3 shows that stochastic observability may not be sufficient to *uniquely and exactly* identify an initial state, but that stochastic observability may allow us to estimate the state in a maximum likelihood sense.

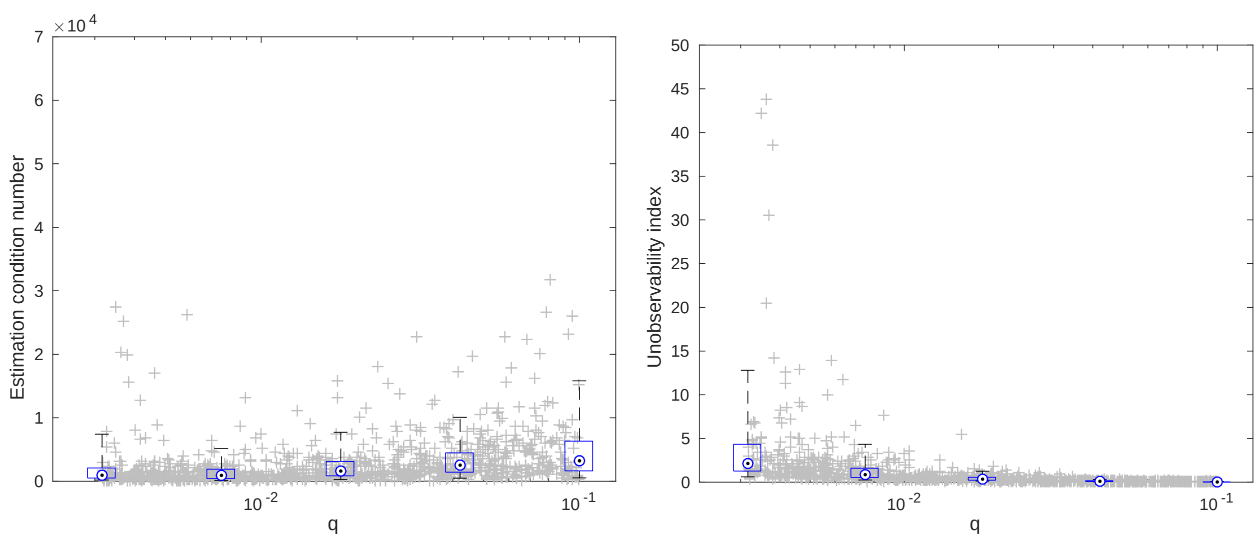
Figures 7.4 and 7.5 show the observability metrics computed for a range of q and, as before, for a trade-off between control and noise. As in our last example, there is a local minimum in the estimation condition number. Note that the condition number can range quite high (to very poorly conditioned estimation) values, but that the median condition number stays close to the condition number with pure control.

While the noise analysis here might seem superfluous, given that control in Σ_4 is able to actuate the same parts of the dynamics as the noise, in general, this situation need not be the case. While these systems were structured to have similar noise and control inputs, so that noise and control influences on observability could be compared, in many systems, noise and control actuate the dynamics differently. Noise, in general, can induce observability of



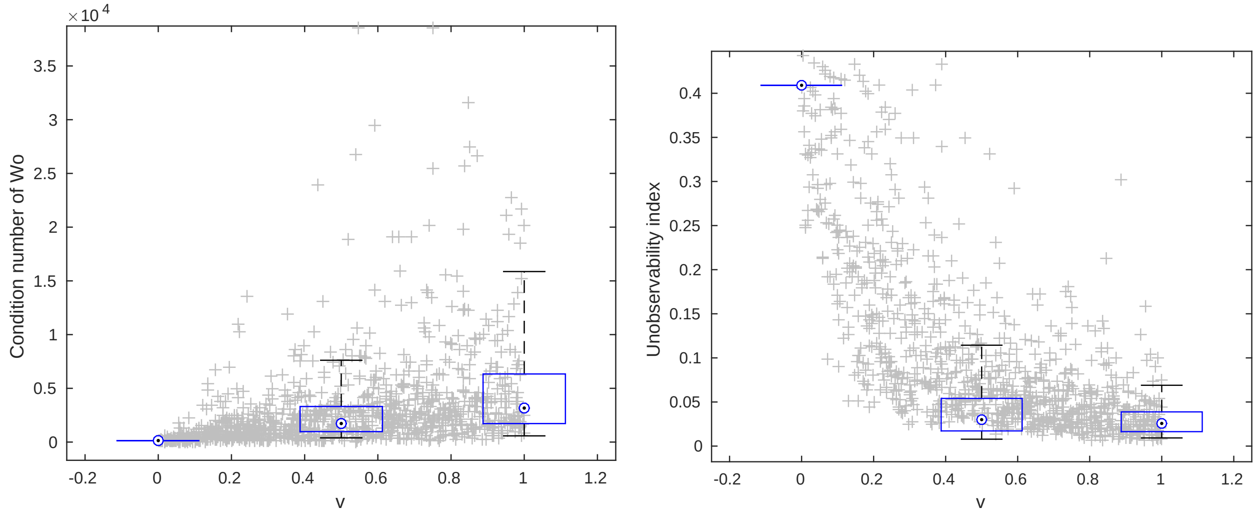
(a) (b)

Figure 7.3: 1000 sample runs with one highlighted run illustrate that acceleration noise can render the initial unicycle heading of 45° observable up to 180°. The forward and reverse headings are clearly visible in the histogram of directions of the final position of the vehicle from the origin.



(a) (b)

Figure 7.4: Estimation condition number for the unicycle system has a minimum near $q = 0.008$. As before, too much and too little noise contribute to poor conditioning. Increasing noise monotonically decreases the unobservability index for the unicycle system.



(a) (b)

Figure 7.5: Control provides better conditioning than mixed or pure noise for the unicycle, while increasing noise monotonically decreases the median unobservability index in the unicycle system.

states that cannot be excited by the available controls.

It may also appear that in this example we have added noise only to the state calculated to provide the most benefit. However, Σ_4 may be physically representative of, for example, a unicycle robot with electrical noise in its drive motors. Furthermore, the intuition that leads us to this experiment still holds in general, even if we add noise to other states as we can see from Figure 7.3. We do not consider process noise in the position states for this system, because those states are kinematically related to the heading/velocity states, and therefore not subject to internal forces.

7.3 Quadrotor Dynamics

Finally, we examine the application of the empirical observability Gramian to a six degree of freedom quadrotor, which we model as a rigid body subject to control torques along all three axes, a body-fixed control force perpendicular to the plane of the rotors, and gravity in the inertial downwards direction. For our states, we use the inertial frame position, Euler angles, inertial frame velocity, and Euler angle rates, with the dynamics

$$\Sigma_5 : \quad \begin{aligned} dx &= \begin{bmatrix} x_{6:12} \\ \mathcal{M}(x_{1:6})^{-1}\left(\mathcal{Q}(x_{1:6}, u) - \mathcal{D}(x)x_{6:12}\right) \end{bmatrix} dt + \begin{bmatrix} 0_{6\times 1} \\ \mathcal{M}(x_{1:6})^{-1} \end{bmatrix} dW \\ y &= \begin{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \left(R(x_{4:6}) \begin{bmatrix} x_{6:7} \\ 0 \end{bmatrix} + E'(x_{4:6}) \begin{bmatrix} x_{10:11} \\ 0 \end{bmatrix} \frac{x_3}{\cos(x_4)\cos(x_5)} \right) \\ \frac{x_3}{\cos(x_4)\cos(x_5)} \\ x_{4:6} \\ x_{10:12} \end{bmatrix}, \end{aligned} \tag{7.8}$$

where

$$\mathcal{M}(x_{1:6}) = \begin{bmatrix} mI_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & E'^{T}(x_{4:6})JE'(x_{4:6}) \end{bmatrix} \tag{7.9}$$

$$\mathcal{Q}(x_{4:6}, u) = \begin{bmatrix} R^T(x_{4:6}) \begin{bmatrix} 0 \\ 0 \\ u_1 \end{bmatrix} - m \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} \\ u_{3:4} \end{bmatrix} \tag{7.10}$$

$$\mathcal{D}(x) = \begin{bmatrix} 0_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & E'^{T}(x_{4:6}, x_{10:12})JdE'(x_{4:6}, x_{10:12}) + dE'^{T}(x_{4:6}, x_{10:12})JE'(x_{4:6}, x_{10:12}) \end{bmatrix} \tag{7.11}$$

$$R(x_{4:6}) = \begin{bmatrix} c_{x_5}c_{x_6} & c_{x_5}s_{x_6} & -s_{x_5} \\ s_{x_4}s_{x_5}c_{x_6} - c_{x_4}s_{x_6} & s_{x_4}s_{x_5}s_{x_6} + c_{x_4}c_{x_6} & c_{x_5}s_{x_6} \\ c_{x_4}s_{x_5}c_{x_6} + s_{x_4}s_{x_6} & c_{x_4}s_{x_5}s_{x_6} - s_{x_4}c_{x_6} & c_{x_5}c_{x_4} \end{bmatrix} \tag{7.12}$$

$$E'(x_{4:6}) = \begin{bmatrix} 1 & 0 & -s_{x_5} \\ 0 & c_{x_4} & c_{x_5}s_{x_4} \\ 0 & -s_{x_4} & c_{x_5}c_{x_4} \end{bmatrix} \tag{7.13}$$

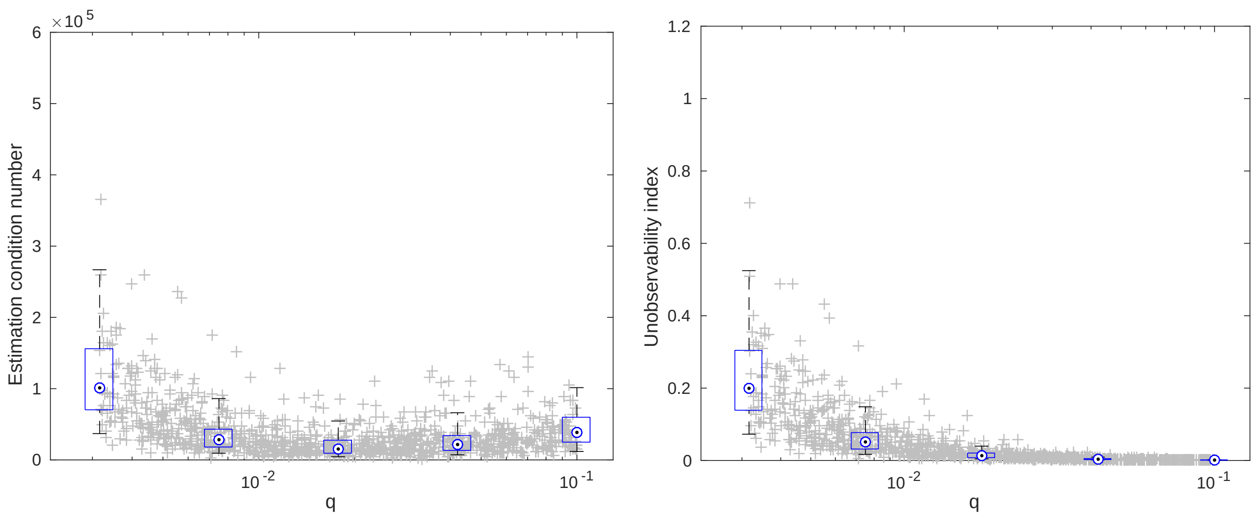
Table 7.1: The vehicle parameters for the Hummingbird quadrotor used in the simulations of this section.

J_{11}	0.0046 kg m^2	m	0.782 kg
J_{22}	0.0046 kg m^2	g	9.81 m/s^2
J_{33}	0.0082 kg m^2		

$$dE'(x_{4:6}) = \begin{bmatrix} 0 & 0 & -c_{x_5}x_{11} \\ 0 & -s_{x_4}x_{10} & -s_{x_5}s_{x_4}x_{11} + c_{x_5}c_{x_4}x_{10} \\ 0 & -c_{x_4}x_{10} & -s_{x_5}c_{x_4}x_{11} + c_{x_5}c_{x_4}x_{10} \end{bmatrix}, \tag{7.14}$$

where m is the vehicle mass, J is the vehicle rotational inertia, and g is the gravitational constant. We have also shortened $\sin(x)$ and $\cos(x)$ to s_x and c_x respectively. The measurement function corresponds to a body-fixed optic flow sensor and range finder facing along the negative x_3 body-axis and returning off of a plane at the inertial $x_3 = 0$ coordinate, and a body fixed inertial measurement unit (IMU). Note that the control input is assumed to be a body-fixed force along the vehicle x_3 axis and three independent torques. We assume that an appropriate calibration has been performed to allow the vehicle rotor velocities (the true control of such a vehicle) to be chosen to provide independent control over these quantities, and that the motor bandwidth is sufficiently high to allow the forces and torques to be adjusted effectively instantaneously. The vehicle parameters, Table 7.1 were chosen to match those of the Ascending Technologies Hummingbird quadrotor from the Nonlinear Dynamics and Controls Laboratory.

The nominal initial condition for the computation of the empirical observability Gramian of this system was chosen to be a steady level hover at an altitude of 1 m. We first compute the Gramian of the systems with no noise. In this case, the rank of the Gramian is 10. By examination, we can see that there is not enough information from the sensors to determine the inertial planar position of the vehicle in an absolute sense. The planar position of the vehicle does not influence any of the sensor readings, so the most we can determine is the



(a) (b)

Figure 7.6: Estimation condition number for the unicycle system has a minimum near $q = 0.008$. As before, too much and too little noise contribute to poor conditioning. Increasing noise monotonically decreases the unobservability index for the unicycle system.

relative position of the vehicle to the initial condition. Therefore, a the Gramian rank indicates that all of the remaining states of the vehicle are observable. We can confirm this understanding, by noting that the rank of the sub-matrix of the Gramian consisting of the last ten rows and columns of the empirical observability Gramian is also 10.

As with our previous two example systems, we can also compute the Gramian of the system with noise, and perform a Monte Carlo analysis of the impact of noise on the observability metrics of the system. Not that for this vehicle, unlike in the previous two cases, noise or control is not required for the system to be observable.[1] Noise was scaled by a parameter, q across the Monte Carlo runs. As with the unicycle vehicle, increasing noise adds energy to the output, decreasing the unobservability index, as we can see from Figure 7.6b.

Interestingly, however, we can see from Figure 7.6a the quadrotor system also exhibits a

[1]Because the planar position states cannot be measured from the given sensors with or without noise, we will consider the observability of only the remaining ten states in this section. Thus, we compute the observability metrics of the empirical observability Gramian from the sub-matrix containing the last ten rows and columns, and will say that the system is observable when the last ten states are observable.

local minimum in the estimation condition number for a value of q near 0.02. In the case of the previous two systems this result was somewhat expected: when the noise in the system was zero, the system was unobservable and the estimation condition number was infinite, and when the noise increased we expected the estimation problem to become more poorly conditioned as the dynamics of the system were drowned out by noise. Here, however, the system is observable even when the noise is zero. The presence of noise has actually improved the observability of the system.

These results provide some guidance to system designers. They suggest that by accounting for noise already present in the system, we may find that our system is more observable than expected. This could allow for the use of cheaper or fewer sensors if it turns out that there is no need to enhance the observability of a designed system.

These results also suggest that adding noise to an observable system may increase the observability of the system. The noise in Σ_5 was added to the linear and angular acceleration. Recall that the control inputs for the system were one force and three torques. Therefore, the control inputs could be used to replicate the effects of process noise, or to increase the amount of noise experienced by the system. The advantage of using control input noise to enhance observability is that zero-mean white noise is simple to generate on top of whatever deterministic control policy is in use, and will have minimal *net* impact on the trajectory of the of the system. The noise will cause the velocity states to have a random walk component with zero mean distance from the origin. The drift of the states would have to be compensated by a stabilizing controller, however.

Finally, these results show that the effects of process noise on observability that were observed on the relatively simple systems of the first two sections of this chapter extend to more realistic systems with more complicated dynamics. The system Σ_5 has twelve states and accurately models the dynamics and available sensors of a vehicle in widespread use in many laboratories around the world, and observability metrics of the system also exhibit the responses to noise that were observed in the simpler example systems, Σ_3 and Σ_4.

Chapter 8

CONCLUSION

Returning now to the themes from the introduction, we can say that we have successfully addressed (at least partially) the questions and goals that we set for ourselves. In particular, we have shown that the empirical observability Gramian can provide a unified approach to the observability of discrete and continuous-time, deterministic and stochastic, linear and nonlinear systems. For linear systems, in both discrete and continuous-time, the empirical observability Gramian reduces to the corresponding linear observability Gramian with a rank (equivalently, non-zero minimum singular value) condition that is equivalent to observability. For nonlinear systems, in both discrete and continuous-time, we have given a minimum singular value condition for the empirical observability Gramian that is sufficient for weak observability (local weak, in the case of discrete-time systems). And for stochastic systems, we have shown that the rank of expected value of the empirical observability Gramian provides an equivalent condition for stochastic observability of linear stochastic systems, as well as a natural method of extending those definitions to nonlinear stochastic systems.

Furthermore, our numerical results from Chapter 7 demonstrate that our unobservable planar unicycle example from the introduction is observable in a stochastic sense when process noise is included in our analysis. The heading of the vehicle can be determined locally in a maximum likelihood sense when the system is actuated by noise in the acceleration and steering inputs. We also demonstrated that increasing the magnitude of the process noise in a system monotonically increases the local unobservability index of a system, but that the estimation condition number generally has a minimum value above which the noise will decrease the numerical conditioning of the filter. As we expected, noise driven actuation has similar effects on the observability of nonlinear stochastic systems as control driven actuation.

Based on the results in this dissertation, it appears that the empirical observability Gramian naturally extends the Gramian approach to observability from linear systems, whereas the Lie derivative algebra naturally extends the observability matrix approach. While the concepts of weak and local weak observability both reduce to the same concept for linear systems, the nonlinear generalizations of the observability Gramian and observability matrix each address a slightly different concept. The empirical observability Gramian connects more closely to estimation problems than the Lie observability algebra, through the connection to the Fisher information derived in Chapter 4.

The empirical observability Gramian approach to observability analysis of nonlinear systems has the advantage of being much simpler to compute for highly nonlinear systems. While neither the Gramian, nor the Jacobian of the Lie observability algebra can be computed in closed-form in general, the empirical observability Gramian, and the local unobservability index can be computed numerically, even when the system dynamics are presented as a black-box to the user. The results in this paper place a firmer footing under the growing use of the local unobservability index as a metric of observability for nonlinear systems. Additionally, the Gramian provides the ability to test nonlinear systems for an additional definition of observability, one that is somewhat more general than the local weak observability that is usually tested for.

The results of Chapters 6 and 7 are likely to invite comparisons to persistence of excitation in system identification. In order for the least squares method of system identification to converge to an equivalent system representation, the input to the system must be persistently exciting, a property possessed by Gaussian white noise inputs, among others. In both cases, we are attempting to learn more about aspects of the system from the outputs, with noise as a potentially useful input signal to tease out the required information. However, the role of noise in observability is distinct from its role in system identification. In system identification, it is the structure of the noise itself that makes it attractive, whereas in observability we are more interested in where the noise enters the system, which states it couples, and which states it can actuate. We focus only on noise inputs in the observability analysis in this dissertation because, in many systems, they can actuate states that cannot be reached by control inputs. As Figure 7.5a showed, noise is not always the most effective signal for enabling observability, nor are high bandwidth signals required to enable observability.

Another advantage of the empirical observability Gramian approach to observability is that the definitions of stochastic observability used in this paper are largely restricted to linear systems – they either directly depend on the linear system matrices or on the fundamental matrix of the linear system in the definition itself. Theorem 6.7 demonstrates that the Gramian could be used in an equivalent definition of stochastic observability, but one which would also apply to nonlinear stochastic systems. Given that most research in nonlinear stochastic observability currently centers around estimability and reconstructability, we have provided an alternate approach to observability of nonlinear stochastic systems.

Of course, the empirical Gramian approach to observability is not without weaknesses. Although the local unobservability index itself can be computed simply for highly nonlinear systems, in general the bound upon it to provide the sufficient condition for observability is not generally computable in closed-form. While the bound can be approximated using central-differencing methods, this procedure requires third-order derivatives (with their associated reduced numerical stability) to be computed at many time and spatial locations, which can be computationally expensive. We note that similar differencing methods for the Lie observability algebra can require even higher order numerical derivatives, however, and suffer greater numerical instabilities.

Furthermore, while the Gramian can be used to test for observability of a system at a particular state, and with a particular control, we must still find valid control inputs to test without knowing *a priori* if such inputs even exist or what form they might take. In contrast, the Lie observability algebra can inform our choice of control input to obtain observability without the requirement of exhaustive search. Some insight might be gleaned from a rank deficient Gramian by checking which states appear in the null space of the Gramian and determining what controls might actuate those states, but determining the correct form of the controls, and how long to apply them for, would not always be simple. Just as with the

Lie observability algebra, the non-necessity of the observability condition should be kept in mind at all times, to avoid falling into the trap of thinking the system unobservable simply because the Gramian is not full rank with a particular control and integration time. The Gramian observability condition is *equivalent* to observability only for linear deterministic and stochastic systems.

While we have addressed many of the questions we set out to ask, many open questions remain, and new areas of investigation have been opened by this work. In general, the problem of observability analysis becomes more challenging as the models used become more detailed and realistic. Linear systems have multiple straightforward tests for observability, while, as we mentioned in Chapter 2, testing observability for nonlinear deterministic models is substantially more challenging. The theorems in this dissertation provide tests for weak observability of nonlinear systems, as well as means of testing observability when process noise is included in models. However, challenges remain for observability analysis of models with yet more detail and realism. For example, one avenue of possible research would be the inclusion of sensor noise in these results. While we have neglected sensor noise to simplify our focus on process noise, most systems are subject to noisy sensors, which could increase the estimation condition number upon inclusion.

The empirical Gramian contains information about surrounding points in state space as a result of the perturbation techniques. Furthermore, the Gramian depends continuously on the state parameter, meaning that provided the inequality of Theorem 3.4 is met strictly (without equality), then the system should be observable in some region about the point of interest. Future work to determine the size of this region (presumably as a function of the system dynamics and the value of the local unobservability index) could allow the Gramian to be used a sufficient condition for observability on sets. Observability of the system as a whole could be determined by a tiling of such sets across the state domain.

Many inequalities were necessary to arrive at the bounds in Theorem 3.4. Some of these inequalities are potentially quite loose, and the cumulative effect of all the inequalities together produced a local unobservability index bound that is extremely conservative. It is possible that tighter bounds might be arrived at, if it were possible to determine the positive (or negative) definiteness of Δ^ε from the properties of W_o^ε, or if some of the inequalities in the proof could be tightened.

Another direction for research would be to investigate empirical Gramians of higher order (Gramians based on higher order central difference perturbations of the initial conditions). Such a higher order Gramian would presumably require more simulation to compute, but the finite ε approximation error for such a Gramian would potentially be $O(\varepsilon^4\tau)$ or better. This might be one path towards tighter bounds on the minimum singular value for observability.

Finally, all of observability is about taking information from the output of the system to determine the state of the system, using information about the dynamics to constrain our solutions. Intuitively, we see that a system can only be observable if there is mutual information between the state and the output of a system. Nevertheless, approaches to nonlinear stochastic observability that directly utilize mutual information as a criterion for observability (such as that taken by Liu and Bitmead [49]) do not clearly connect to more traditional definitions of observability. An important question remaining, is how we can unify these approaches, or at the least show under what conditions they do or do not agree. Is it possible to show that the Gramian is full rank only when there is non-zero mutual information between each state and the output? Are the singular values of the Gramian linked to the amount of mutual information between the state and the output?

Ultimately, we have advanced our understanding of sensing in an inherently stochastic world. However, we are brought to mind of the following quote (as we often are in studying scientific endeavors):

> We have not succeeded in answering all our problems. The answers we have found only serve to raise a whole set of new questions. In some ways we feel we are as confused as ever, but we believe we are confused on a higher level and about more important things. – Attributed variously

Sensing and estimation are ubiquitous in our world, both in engineering and in daily life, as is noise. This work brings partial understanding of how these phenomena interact, while leaving much more to be learned.

BIBLIOGRAPHY

[1] T. McGeer and J. Vagners, “Historic crossing: An unmanned aircraft’s Atlantic flight,” *GPS World*, vol. 10, pp. 24–30, February 1999.

[2] B. Wittenmark, “Adaptive dual control methods: an overview,” in *Proceedings of the IFAC Symposium on Adaptive Systems in Control and Signal Processing*, (Budapest, Hungary), pp. 67–72, 1995.

[3] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, “Stochastic resonance in climatic change,” *Tellus*, vol. 34, pp. 10–16, February 1982.

[4] C. Nicolis, “Stochastic aspects of climatic transitions - response to a periodic forcing,” *Tellus*, vol. 34, pp. 1–9, February 1982.

[5] A. R. Bulsara and F. E. Moss, “Single neuron dynamics: Noise-enhanced signal processing,” in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 1, (Singapore), pp. 420–425, November 1991.

[6] X. Godivier and F. Chapeau-Blondeau, “Noise-enhanced transmission of spike trains in the neuron,” *Europhysics Letters*, vol. 35, pp. 473–477, August 1996.

[7] S.-G. Lee, A. Neiman, and S. Kim, “Coherence resonance in a Hodgkin-Huxley neuron,” *Physical Review E*, vol. 57, pp. 3292–3297, March 1998.

[8] H. C. Tuckwell and R. Rodriguez, “Analytical and simulation results for stochastic FitzHugh-Nagumo neurons and neural networks,” *Journal of Computational Neuroscience*, vol. 5, pp. 91–113, March 1998.

[9] D. T. W. Chik, Y. Wang, and Z. D. Wang, “Stochastic resonance in a Hodgkin-Huxley neuron in the absence of external noise,” *Physical Review E*, vol. 64, pp. 021913.1–021913.6, July 2001.

[10] A. Longtin, *Nonlinear Dynamics in Physiology and Medicine*, vol. 25 of *Interdisciplinary Applied Mathematics*, ch. Effects of Noise on Nonlinear Dynamics, pp. 149–189. New York, NY, USA: Springer-Verlag New York, first ed., 2003.

[11] S. Lall, J. E. Marsden, and S. Glavaški, "Empirical model reduction of controlled nonlinear systems," in *Proceedings of the IFAC World Congress*, pp. 473–478, 1999.

[12] S. Lall, J. E. Marsden, and S. Glavaški, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *International Journal of Robust and Nonlinear Control*, vol. 12, pp. 519–535, May 2002.

[13] J. Hahn and T. F. Edgar, "An improved method for nonlinear model reduction using balancing of empirical Gramians," *Computers & Chemical Engineering*, vol. 26, pp. 1379–1397, October 2002.

[14] J. M. A. Scherpen, "Balancing for nonlinear systems," *Systems and Controls Letters*, vol. 21, pp. 143–153, August 1993.

[15] A. K. Singh and J. Hahn, "On the use of empirical Gramians for controllability and observability analysis," in *Proceedings of the American Control Conference*, (Portland, OR, USA), pp. 140–141, June 2005.

[16] A. J. Krener and K. Ide, "Measures of unobservability," in *Proceedings of the Joint IEEE Conference on Decision and Control and Chinese Control Conference*, (Shanghai, China), pp. 6401–6406, December 2009.

[17] B. T. Hinson, M. K. Binder, and K. A. Morgansen, "Path planning to optimize observability in a planar uniform flow field," in *Proceedings of the American Control Conference*, (Washington, D.C., USA), pp. 1392–1399, June 2013.

[18] T. Glotzbach, N. Crasta, and C. Ament, "Observability analyses and trajectory planning for tracking of an underwater robot using empirical Gramians," in *Proceedings of the IFAC World Congress*, (Cape Town, South Africa), pp. 4215–4221, August 2014.

[19] J. D. Quenzer and K. A. Morgansen, "Observability based control in range-only underwater vehicle localization," in *Proceedings of the American Control Conference*, (Portland, OR, USA), pp. 4702–4707, June 2014.

[20] Z. Yu, P. Cui, and S. Zhu, "Observability-based beacon configuration optimization for Mars entry navigation," *Journal of Guidance, Control, and Dynamics*, vol. 38, pp. 643–650, April 2015.

[21] C. Eberle and C. Ament, "Identifiability and online estimation of diagnostic parameters with in the glucose insulin homeostasis," *Biosystems*, vol. 107, pp. 135–141, March 2012.

[22] J. Qi, K. Sun, and W. Kang, "Optimal PMU placement for power system dynamic state estimation by using empirical observability Gramian," *IEEE Transactions on Power Systems*, vol. 30, pp. 2041–2054, September 2014.

[23] B. T. Hinson and K. A. Morgansen, "Flowfield estimation in the wake of a pitching and heaving airfoil," in *Proceedings of the American Control Conference*, (Montreal, Canada), pp. 1085–1091, June 2012.

[24] L. DeVries and D. A. Paley, "Observability-based optimization for flow sensing and control of an underwater vehicle in a uniform flowfield," in *Proceedings of the American Control Conference*, (Washington, D.C., USA), pp. 1386–1391, June 2013.

[25] W. Kang and L. Xu, "Partial observability for some distributed parameter systems," *International Journal of Dynamics and Control*, vol. 2, pp. 587–596, December 2014.

[26] S. King, W. Kang, and L. Xu, "Computational issues on observability and optimal sensor locations," in *Proceedings of the American Control Conference*, (Portland, OR, USA), pp. 4075–4080, June 2014.

[27] P. Batista, C. Silvestre, and P. Oliveira, "Single range aided navigation and source localization: Observability and filter design," *Systems and Controls Letters*, vol. 60, pp. 665–673, August 2011.

[28] R. Hermann and A. J. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, pp. 728–740, October 1977.

[29] I. Hwang, H. Balakrishnan, and C. Tomlin, "Observability criteria and estimator design for stochastic linear hybrid systems," in *Proceedings of the European Control Conference*, (Cambridge, UK), pp. 3317–3322, September 2003.

[30] M. Aoki, "On observability of stochastic discrete-time dynamic systems," *Journal of the Franklin Institute*, vol. 286, pp. 36–58, July 1968.

[31] Y. Sunahara, S. Airhara, and M. Shiraiwa, "The stochastic observability for noisy non-linear stochastic systems," *International Journal of Control*, vol. 22, pp. 461–480, March 1975.

[32] H.-F. Chen, "On stochastic observability and controllability," *Automatica*, vol. 16, pp. 179–190, March 1980.

[33] M. Mariton, "Stochastic observability of linear systems with Markovian jumps," in *Proceedings of the IEEE Conference on Decision and Control*, (Athens, Greece), pp. 2208–2209, December 1986.

[34] Y. Baram and T. Kailath, "Estimability and regulability of linear systems," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 26, (Los Angeles, CA, USA), pp. 1706–1713, December 1987.

[35] F. Gustafsson, "Stochastic observability and fault diagnosis of additive changes in state space models," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, (Salt Lake City, UT, USA), pp. 2833–2836, May 2001.

[36] V. Dragan and T. Morozan, "Stochastic observability and applications," *IMA Journal of Mathematical Control and Information*, vol. 21, no. 3, pp. 323–344, 2004.

[37] W. Zhang and B.-S. Chen, "On stabilizability and exact observability of stochastic systems with their applications," *Automatica*, vol. 40, pp. 87–94, January 2004.

[38] H. Sun and M. Li, "Exact observability/exact detectability and spectrum assignment of stochastic systems," in *Proceedings of the World Congress on Intelligent Control and Automation*, (Jinan, China), pp. 3786–3790, July 2010.

[39] H. Ting, Z. Weihai, and M. Hongji, "Stability and exact observability of discrete-time Markov jump systems with multiplicative noise," in *Proceedings of the Chinese Control Conference*, (Hefei, China), pp. 1634–1639, July 2012.

[40] L. Shen, J. Sun, and Q. Wu, "Observability and detectability of discrete-time stochastic systems with Markovian jump," *Systems and Controls Letters*, vol. 62, pp. 37–42, January 2013.

[41] Y. Subasi and M. Demirekler, "Quantitative measure of observability for linear stochastic systems," *Automatica*, vol. 50, pp. 1669–1674, June 2014.

[42] Y. Sunahara, K. Kishino, and S. Aihara, "On stochastic observability of non-linear discrete-time dynamical systems," *International Journal of Control*, vol. 19, no. 4, pp. 719–732, 1974.

[43] Y. Sunahara, S. Aihara, and K. Kishino, "On the stochastic observability and controllability for non-linear systems," *International Journal of Control*, vol. 22, pp. 65–82, February 1975.

[44] V. L. Bageshwar, D. Gebre-Egziabher, W. L. Garrard, and T. T. Georgiou, "Stochastic observability test for discrete-time Kalman filters," *Journal of Guidance, Control, and Dynamics*, vol. 32, pp. 1356–1370, July 2009.

[45] P. D. West and A. H. Haddad, "On the observability of linear stochastic switching systems," in *Proceedings of the American Control Conference*, vol. 2, (Baltimore, MD, USA), pp. 1846–1847, June 1994.

[46] J.-P. L. Cadre and C. Jauffret, "Discrete-time observability and estimability analysis for bearings-only target motion analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, pp. 178–201, January 1997.

[47] A. R. Liu and R. R. Bitmead, "Stochastic reconstructibility and estimability," in *Proceedings of the Joint IEEE Conference on Decision and Control and Chinese Control Conference*, (Shanghai, China), pp. 2339–2344, December 2009.

[48] A. R. Liu and R. R. Bitmead, "Observability and reconstructibility of hidden Markov models: Implications for control and network congestion control," in *Proceedings of the IEEE Conference on Decision and Control*, (Atlanta, GA, USA), pp. 918–923, December 2010.

[49] A. R. Liu and R. R. Bitmead, "Stochastic observability in network state estimation and control," *Automatica*, vol. 47, pp. 65–78, January 2011.

[50] A. R. Liu, *Stochastic Observability, Reconstructibility, Controllability, and Reachability.* PhD thesis, University of California, San Diego, San Diego, CA, USA, 2011.

[51] Y. Shen and H. Zhang, "Estimability of quantized linear systems," in *Proceedings of the International Conference on Wireless Communications and Signal Processing*, (Nanjing, China), pp. 1–6, November 2011.

[52] H. Zhang, Y. Tian, and L. Gao, "Stochastic observability of linear systems under access constraints," *Asian Journal of Control*, vol. 17, pp. 64–73, January 2015.

[53] Z.-Y. Li, Y. Wang, B. Zhou, and G.-R. Duan, "Detectability and observability of discrete-time stochastic systems and their applications," *Automatica*, vol. 45, pp. 1340–1346, May 2009.

[54] W. Zhang, H. Zhang, and B.-S. Chen, "Generalized Lyapunov equation approach to state-dependent stochastic stabilization/detectability criterion," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1630–1642, August 2008.

[55] V. A. Ugrinovskii, "Observability of linear stochastic uncertain systems," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 3, (Las Vegas, NV, USA), pp. 3035–3039, 2002.

[56] S. Hong, H.-H. Chun, S.-H. Kwon, and M. H. Lee, "Observability measures and their application to GPS/INS," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 97–106, January 2008.

[57] M. H. Lee, J. H. Lee, Y. H. Koh, H. G. Park, J. H. Moon, and S. P. Hong, "Observability and estimability analysis of the GPS and INS in the vehicle," *Journal of Mechanical Systems for Transportation and Logistics*, vol. 3, pp. 537–551, December 2010.

[58] V. Dragan and T. Morozan, "Observability and detectability of a class of discrete-time stochastic linear systems," *IMA Journal of Mathematical Control and Information*, vol. 23, pp. 371–394, September 2006.

[59] J. L. Crassidis and J. L. Junkins, *Optimal State Estimation of Dynamic Systems*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series, New York, NY: Chapman & Hall/CRC, 2004.

[60] S. Sugathadasa, C. Martin, and W. P. Dayawansa, "Boundedness of covariance estimates of extended Kalman filtering with directional measurements and linear state dynamics," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 4, (Orlando, FL, USA), pp. 3790–3795, December 2001.

[61] M. Zakai and J. Ziv, "Lower and upper bounds on the optimal filtering error of certain diffusion processes," *IEEE Transactions on Information Theory*, vol. 18, pp. 325–331, May 1972.

[62] E. Rohr, D. Marello, and M. Fu, "Kalman filtering with intermittent observations: Bounds on the error covariance distribution," in *Proceedings of the IEEE Conference on Decision and Control and European Control Conference*, (Orlando, FL, USA), pp. 2416–2421, December 2011.

[63] B. T. Hinson, N. D. Powel, J. D. Quenzer, and K. A. Morgansen, "Observability-based guidance algorithm for navigation using a single ranging beacon," in *Robotics: Science and Systems, Workshop on Autonomous Control, Adaptation, and Learning for Underwater Vehicles*, (Berkley, CA, USA), July 2014.

[64] N. D. Powel and K. A. Morgansen, "Empirical observability Gramian rank condition for observability of nonlinear systems with control," in *Proceedings of the Conference on Decision and Control*, (Osaka, Japan), December 2015.

[65] N. D. Powel and K. A. Morgansen, "Empirical observability Gramian for stochastic observability of nonlinear systems," *IEEE Transactions on Automatic Control*, in preparation.

[66] C.-T. Chen, *Linear System Theory and Design.* Oxford Series in Electrical and Computer Engineering, New York, NY: Oxford University Press, third ed., 1999.

[67] H. Nijmeijer and A. van der Schaft, *Nonlinear Dynamical Control Systems.* New York: Springer-Verlag, 1990.

[68] H. Khalil, *Nonlinear Systems.* Upper Saddle River, NJ: Prentice Hall, third ed., 2002.

[69] M. Anguelova, *Nonlinear Observability and Identifiability: General Theory and a Case Study of a Kinetic Model for S. cerevisiae.* PhD thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, April 2004.

[70] F. Albertini and D. D'Allessandro, "Observability and forward-backward observability of discrete-time nonlinear systems," *Mathematics of Control, Signals, and Systems*, vol. 15, pp. 275–290, November 2002.

[71] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems.* Texts in Applied Mathematics, New York, NY: Springer, first ed., 2005.

[72] J. E. Marsden, T. S. Ratiu, and R. Abraham, *Manifolds, Tensor Analysis, and Applications.* New York, NY: Springer-Verlag, third ed., 2001.

[73] P. Dorato, C. T. Abdallah, and V. Cerone, *Linear Quadratic Control: An Introduction.* Malabar, FL, USA: Krieger Publishing Company, 2000.