2.12 Notes and Comments

Basic Definitions

Various definitions of “system” or “machine” have been proposed in the literature, all attempting to model basically the same notion, but differing among themselves in mathematical technicalities. Some early references are [114], [230], and [444], but of course similar transition formalisms (with no controls nor outputs) have long been used in the theory of dynamical systems. Category-theoretic definitions also have been suggested, which make the definition of “linear” system or certain classes of “topological” systems very easy; see, for instance, [22] for pointers to that literature, as well as [314], Section 8.5. The area called general systems theory concerns itself with abstract systems defined by transition maps, typically with no extra algebraic or topological structure; see, for instance, [300] and [429].

Recently there has been a resurgence of interest in control theory in the use of automata models that had originated in computer science, logic, and operations research. In most of these studies, time corresponds not to “real” clock time but to instants at which the system changes in some special manner, as in queuing models when a new customer arrives. Work in this area falls under the general label of “control of discrete-event systems”; some references are [90], [405], and [433]. An approach combining discrete-event ideas together with more classical linear design, in the context of a special-purpose computer language, has been recently proposed in [46]; this paper models some of the logical operations as discrete-time systems over a finite field. More generally, the area of hybrid systems has become active in the mid 1990s, see for instance the conference proceedings [15]; one possible approach to combining automata and linear systems was studied in [362] (see also the author’s paper in [15]).

A conceptually very different foundation for system and control theory has been also proposed. Rosenbrock in [331] emphasized the fact that physical principles sometimes lead to mixtures of differential and algebraic equations, and state space descriptions are not necessarily the most appropriate; implicit differential equations are then used. Related to this is the literature on singular systems, or descriptor systems, defined (in the linear case) by equations of the type $\dot{E}x = Ax + Bu$ with the matrix $E$ not necessarily invertible; see, e.g., [83], [94], [192], and [277], as well as [82] for the related differential equation theory. The recent work of Willems has carried out this reasoning to the limit and proposed a completely different formalism for systems, based on not distinguishing a priori between inputs and outputs. Instead of “input/output” data, one then has an “observed time-series” which summarizes all i/o data without distinguishing between the two; the papers [309], [424], [425], and [426] provide an introduction to this literature.

The definitions given may be generalized to include what are often called multidimensional systems, for which the time set $T$ is $\mathbb{R}^k$ or $\mathbb{Z}^k$, with $k > 1$. (The terminology is most unfortunate, since we normally use the term “dimen-
sion” to refer to the number of dependent variables in the equations defining the dynamics, rather than to the number of independent variables.) Multidimensional systems are useful in areas such as picture processing or seismic data analysis, where different space directions may be thought of as “time" coordinates. Multidimensional systems are considerably more difficult to handle, since the difference and differential equations that appear with ordinary systems are replaced by partial difference and differential equations. For references to this area, see the book [57].

There is also a variation of the definition of system that allows for the possibility of the readout map $h$ to depend directly on $\omega(t)$. In other words, one has $h : T \times \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{Y}$. In the context of such more general systems, one would call state-output systems those that we have defined here. The use of non-state output systems causes some technical difficulties which are avoided with our definition, but one could model such more general systems using pairs $(\Sigma, \alpha)$, where $\Sigma$ is a system with outputs as defined in this Section and $\alpha : \mathbb{U} \times \mathbb{Y} \rightarrow \mathbb{Y}$ is a map.

In practice, it may often happen that control of a large system can only be achieved through local effects; this gives rise to the theory of large scale systems and corresponding decentralized control objectives; see [350] for a detailed treatment of many of the issues that arise there, including the use of decomposition techniques, as well as many references.

Continuous-time systems $\dot{x} = f(x, u)$ can be studied as differential inclusions ([26], [111], [93]). In general, given a set-valued map $F : \mathbb{X} \rightarrow 2^{\mathbb{R}^n}$, that is, a map that assigns to each point $x$ in $\mathbb{X} \subseteq \mathbb{R}^n$ a subset $F(x) \subseteq \mathbb{R}^n$, a solution of the differential inclusion $\dot{x} \in F(x)$ means, by definition, an absolutely continuous function $x : I \rightarrow \mathbb{X}$, defined on some interval, so that $\dot{x}(t) \in F(x(t))$ for almost all $t \in I$. To a system $(\Sigma) \dot{x} = f(x, u)$ one may associate the differential inclusion with $F_{\Sigma}(x) := \{f(x, u) \mid u \in \mathbb{U}\}$, and paths of the system become solutions of the latter. Conversely, there is a rich literature dealing with selection theorems which, under appropriate regularity assumptions on $F_{\Sigma}$, guarantee that every solution of $\dot{x} = F_{\Sigma}(x)$ arises in this manner, from some control $u$.

I/O Maps

There are sometimes cases in which the input affects the output with no delay, for instance in the “i/o behavior” of more general systems such as those mentioned above ($h$ depends in $u$) or in dealing with a memoryless i/o map, a map induced by an arbitrary mapping

$$h : \mathbb{U} \rightarrow \mathbb{Y}$$

between any two sets by the obvious rule

$$\lambda(\sigma, \omega)(t) := h(\omega(t)).$$

Unfortunately, there are technical difficulties in formulating this consistently in terms of response maps. For instance, when treating continuous-time systems
and behaviors one identifies controls that are equal almost everywhere, and this means that the instantaneous value \( \omega(t) \) is not well-defined. Also, for the empty control the output would be undefined, meaning that one cannot really talk about “instantaneous measurements” of states. Thus, a concept analogous to that of response cannot be used in this case. It is possible, however, to generalize the idea of i/o map, where strict causality is now weakened to the requirement (causality) that

\[
\bar{\lambda}(\mu, \sigma, \omega|_{[\sigma, \mu]})(t) = \bar{\lambda}(\tau, \sigma, \omega)(t)
\]

only for \( t \in [\sigma, \mu) \), rather than for all \( t \in [\sigma, \mu] \). Then a satisfactory theory can be developed for causal mappings. We prefer to deal here with the concept of a response, or equivalently, a (strictly causal) i/o map.

### Linear Discrete-Time Systems

In control applications, linear systems are of interest mainly in the case in which the field is \( \mathbb{K} = \mathbb{R} \), and sometimes, mostly for technical reasons, \( \mathbb{K} = \mathbb{C} \). However, finite fields do appear in communications and signal processing theory, as illustrated by the error detection example; in particular the theory of what are called convolutional codes is to a great extent the study of linear systems over such fields (see, for instance, [118] and [297]). Recently, work has been done on systems over fields of formal Laurent series, which provide a good model for certain perturbation results; see, for instance, [313].

The algebraic theory of linear systems is no more difficult when dealing with arbitrary fields than with \( \mathbb{R} \) or \( \mathbb{C} \), and so we will give most results for general \( \mathbb{K} \). On the other hand, it is also possible to define linear systems over rings rather than fields, with the same definition but using “module” instead of “vector space”; these are useful in various applications such as modeling the effect of fixed precision integer arithmetic or delay systems (see remarks below). Some results valid over fields generalize, typically with much harder proofs, to systems over rings, but many others do not. For references on systems over rings see, for instance, [61], as well as the early papers [332] and [333] that started most of this research.

### Linear Continuous-Time Systems and Behaviors

It is also of great interest to consider linear (or nonlinear) infinite dimensional systems, also called distributed systems. For instance, transmission delays give rise to what are often called delay-differential or hereditary models such as

\[
\dot{x}(t) = x(t - 1) + u(t),
\]

which reflects the effect of a retarded feedback; or one may study systems defined by partial differential equations, such as a heat equation with boundary control. Many such examples can be modeled by equations of the type \( \dot{x} = Ax + Bu \).
evolving in Banach spaces. There is a rich and extensive literature on these matters; see, for instance, [281] for optimal control, as well as the books [107], [133], or [147], or the paper [437], for more “system theoretic” questions, each using a somewhat different approach.

Sometimes it is possible to model usefully infinite dimensional systems as “systems over rings of operators.” For instance, one may write the above delay system as the “one dimensional” system

\[ \dot{x} = \sigma x + u, \]

where “\(\sigma\)” is a shift operator on an appropriate space of functions. Viewing this as the linear system \((\sigma, 1)\) over the polynomial ring \(\mathbb{R}[\sigma]\) permits using techniques —like Pole Shifting— from the theory of systems over rings, resulting in constructive synthesis procedures. This approach is due to Kamen; see, for instance, [61], [62], [79], [81], [233], [234], [305], [349], [357], [396], and [397] for more on the topic and related results, and [74] for the study of certain systems defined by partial differential equations in this manner.

**Volterra Expansions**

The Volterra expansion can be generalized to bilinear systems with more controls, and locally or in asymptotic senses to more general classes of nonlinear systems; see, for instance, [336] for details on this and the above example. (One approach for general nonlinear systems relies in first approximating nonlinear systems by bilinear ones, a process analogous to taking truncations of a Taylor expansion. There is also an alternative approach to bilinear approximation, valid in compact-open topologies of controls and based on the Stone-Weierstrass Theorem; see [381].)

Volterra series are important because, among other reasons, they exhibit the various homogeneous nonlinear effects separately. In the context of identification, they allow the separation of higher harmonic responses to periodic signals. When the original system is linear, every term except that for \(l = 1\) vanishes, and the Volterra series (2.65) is nothing more than the variation of parameters formula; more generally, the solution of the linearized system along \(\xi \equiv 0, \omega \equiv 0\) is obtained by simply dropping all terms except the first one. It is also possible to give Volterra expansions for smooth discrete-time systems, as also discussed in [336].

Volterra series were one of the most popular engineering methods for nonlinear systems modeling during the 1950s. Their theoretical study, especially in relation to state space models, was initiated during the early 1970s; see, for instance, [69] for a survey of the early literature.

An alternative approach to Volterra series is based on what are called *generating series*, or *Fliess series*. These are formal power series that essentially correspond to expanding the “kernels” \(W_i\), and can be interpreted in terms of mappings from jets of germs of differentiable inputs into jets of outputs. They
are closely related to work on the representation of solutions of differential equations, as in [89]; see [140] and references therein, as well as [199], [373], [241], and the excellent book [327].
3.10 Notes and Comments

Controllability of Time-Invariant Systems

Since the early work on state-space approaches to control systems analysis, it was recognized that certain nondegeneracy assumptions were useful, in particular in the context of optimality results. However, it was not until R.E. Kalman’s work (see, e.g., [215], [216], [218], and [231]) that the property of controllability was isolated as of interest in and of itself, as it characterizes the degrees of freedom available when attempting to control a system.

The study of controllability for linear systems has spanned a great number of research directions, and topics such as testing degrees of controllability, and their numerical analysis aspects, are still the subject of intensive research.

The idea in the proof of Corollary 3.2.7 was to use an ascending chain condition on the spaces $X_i$ to conclude a finite-time reachability result. Similar arguments are used often in control theory, under other hypotheses than finite dimensionality (or finiteness, as in Lemma 3.2.4). For instance, for discrete-time linear systems over Noetherian rings, one may use the same idea in order to conclude that $R(x) = R^T(x)$ for large enough $T$.

Controllability of continuous-time piecewise linear systems is studied in [267] and [411].

There is also an extensive literature on the controllability properties of infinite dimensional continuous-time or discrete-time linear systems; see, for instance, [107], [147], and [291]. For such systems it is natural to characterize “almost” reachability, where the set of states reachable from a given state is required only to be dense in the state space.

Algebraic Facts About Controllability

What we called the “Hautus condition” actually appeared first in [321], Theorem 1 on page 320, and in a number of other references, including [40]. However, [175] was the first to stress its wide applicability in proving results for linear systems, as well as in extending the criterion to stabilizability (asymptotic controllability) in [176]. Sometimes the condition is also referred to as the “PBH condition” because of the Belevitch and Popov contributions.

It should be emphasized that genericity of controllability was established only with respect to the set of all possible pairs. Often one deals with restricted classes of systems, and among these, controllable systems may or may not form an open dense subset. This gives rise to the study of structural controllability; see, for example, [298], and more generally the book [350] and references therein, for related questions.

In structural controllability studies, classes of systems are defined typically by constraints on the matrices $(A, B)$. For example, any system obtained from an equation of the type

$$\dot{x} + \alpha \dot{x} + \beta x = \gamma u$$
(as arises with a damped spring-mass system) via the introduction of state variables \( x_1 := x, x_2 := \dot{x} \) will have \( A_{11} = B_{11} = 0 \) and \( A_{12} = 1 \). These coefficients 0, 1 are independent of experimental data. Graph-theoretic methods are typically used in structural controllability.

### Controllability Under Sampling

Necessary and sufficient conditions for the preservation of controllability under sampling for multivariable \((m > 1)\) systems are known, but are more complicated to state than the sufficient condition in Theorem 4 (p. 104). See, for instance, [154].

The dual version of the Theorem, for observability (see Chapter 6), is very closely related to Shannon’s Sampling Theorem in digital signal processing.

Some generalizations to certain types of nonlinear systems are also known (see, for instance, [208] and [371]).

### More on Controllability of Linear Systems

The result in Exercise 3.5.7 is a particular case of the more general fact, for time-invariant continuous-time single-input \((m = 1)\) systems, that the operator that gives the optimal transfer from 0 to a state \( x \) in time \( \epsilon \) has an operator norm

\[
\|L^\#\| = O(\epsilon^{-n+\frac{1}{2}})
\]

for small \( \epsilon \). This is proved in [345], where the multiple-input case is also characterized.

### Bounded Controls

Theorem 6 (p. 119) can be found in [343]. A particular case had earlier been covered in [266], and the general case, as well as the obvious discrete-time analogue, were proved in an abstract algebraic manner in [364]. We based our presentation on [170], which in turn credits the result, in the stronger form given in Corollary 3.6.7, to the earlier thesis [196].

It is also possible to give characterizations of controllability under other constraints on control values. For instance, positive controls are treated in [60].

### First-Order Local Controllability

The problem of characterizing local controllability when the first-order test given in Theorem 7 fails is extremely hard, and constitutes one of the most challenging areas of nonlinear control research. It is often of interest to strengthen the definition of local controllability: For instance, about an equilibrium state \( x \) one might require that any state close to \( x \) be reachable from \( x \) in small time (small-time local controllability). Such stronger notions give rise to interesting variations of the nonlinear theory; see, e.g., [388] and references there for this and related issues.
Recurrent Nets

The systems studied in Section 3.8 are often called “continuous-time recurrent neural networks”. The motivation for the term comes from an interpretation of the vector equations for \( x \) in (3.28) as representing the evolution of an ensemble of \( n \) “neurons,” where each coordinate \( x_i \) of \( x \) is a real-valued variable which represents the internal state of the \( i \)th neuron, and each coordinate \( u_{i,i} = 1, \ldots, m \) of \( u \) is an external input signal. The coefficients \( A_{ij}, B_{ij} \) denote the weights, intensities, or “synaptic strengths,” of the various connections. The choice \( \theta = \tanh \), besides providing a real-analytic and globally Lipschitz right-hand side in (3.28), has major advantages in numerical computations, due to the fact that its derivative can be evaluated from the function itself (\( \theta' = 1 - \theta^2 \)). (Sometimes, however, other “activation functions” are used. One common choice is the function \( \sigma(x) = (1 + e^{-x})^{-1} \), which amounts to a rescaling of \( \theta \) to the range (0, 1).) Among the variants of the basic model (3.28) that have been studied in the literature are systems of the general form \( \dot{x} = -x + \theta_n(Ax + Bu) \), where the term \( -x \) provides stability. Recurrent nets (or these variants) arise in digital signal processing, control, design of associative memories (“Hopfield nets”), language inference, and sequence extrapolation for time series prediction, and can be shown to approximate a large class of nonlinear systems; see e.g. [237].

The implication \( 2 \Rightarrow 1 \) in Theorem 8 is from [375], where the reader may also find a solution to Exercise 3.8.11. The necessity of the condition \( B \in B_{n,m} \) was shown to the author by Y. Qiao. A characterization of a weak controllability property for the discrete-time analogue of these systems had been given earlier, in [12], which had also obtained partial results for the continuous-time problem.

Piecewise Constant Controls

Results similar to Proposition 3.9.1 are in fact true for much larger classes of systems. It is possible to establish, for instance, that piecewise constant controls are enough for any analytic continuous-time system. In [367], it is shown that under one weak extra assumption, polynomial controls can be used for controllable analytic systems. See [163] for a survey of such results. This section gave only very simple versions of these more general facts.
4.7 Notes and Comments

The systematic study of controllability of nonlinear continuous-time systems started in the early 1970s. The material in this chapter is based on the early work in the papers [255], [284], [391], and [185], which in turn built upon previous PDE work in [91] and [182]. Current textbook references are [199] and [311]. An excellent exposition of Frobenius’ Theorem, as well as many other classical facts about Lie analysis, can be found in [53].

Under extra conditions such as reversibility, we obtained an equivalence between controllability and accessibility. Another set of conditions that allows this equivalence is based on having suitable Hamiltonian structures, see, e.g., [51]. The advantage of considering accessibility instead of controllability is that, in effect, one is dealing with the transitivity of a group (as opposed to only a semigroup) action, since positive times do not play a distinguishing role in the accessibility rank condition. (Indeed, for analytic systems on connected state spaces, accessibility is exactly the same as controllability using possibly “negative time” motions in which the differential equation is solved backward in time.)

Also, for continuous-time systems evolving on Lie groups according to right-invariant vector fields, better results on controllability can be obtained; see, for instance, [126], [211], and [351]. Nonlinear global controllability problems are emphasized in [160], [161], and [162].

The question of providing necessary and sufficient characterizations for controllability is still open even for relatively simple classes such as bilinear continuous-time systems, but substantial and very deep work has been done to find sufficient conditions for controllability; see, for instance, [388], or [239]. It is worth remarking that it can be proved formally that accessibility is “easier” to check than controllability, in the computer-science sense of computational complexity and NP-hard problems; see [366], [240].

For discrete-time nonlinear accessibility, see for instance [206] and [208].

Example 4.3.13 is from [308], which introduced the model and pursued the analysis of controllability in Lie-algebraic terms.

Much recent work in nonlinear control has dealt with the formulation of explicit algorithms for finding controls that steer a given state to a desired target state, under the assumption that the system is controllable. This is known as the path-planning problem. See for instance the survey paper [251] and the many articles in the edited book [30].
5.10 Notes and Comments

Constant Linear Feedback

See Chapter 2 of [212] for an extensive discussion of controller and controllability forms, as well as related topics such as their use in analog computer simulations.

The Pole-Shifting Theorem has a long history. The proof for single-input systems, simple once the controller form is obtained, appears to have been discovered independently during the late 1950s. Kalman credits J. Bertram, ca. 1959, and Kailath credits also [329]. (In 1962, Harvey and Lee reported in [174] a weak version of the Theorem, asserting only stabilizability, as a consequence of the single-input case.) The general, multivariable, case, over the complexes, was first proved by Langenhop in [265] and independently by Popov ca. 1964, who discussed it in [321]. Over the real field, a complete proof was given by Wonham in [431]. The simple proof for general fields, by reduction to the single-input case, was given in [191]. The result that there is some $F_1$ so that $(A + BF_1, b)$ is controllable is known as Heymann’s Lemma; our proof is from [177]. Pole-placement problems for “singular” or “descriptor” systems $E\dot{x} = Ax + Bu$ are discussed for instance in [94].

For a solution of Exercise 5.1.11, see [430]; for Exercise 5.1.12, see [212], Section 3.2.

For any given pair $(A, B)$ and desired characteristic polynomial $\chi$, there are (when $m > 1$) many feedback matrices $F$ such that $\chi_{A + BF} = \chi$. As a set of equations on the $nm$ entries of $F$, assignability introduces only $n$ constraints, one for each of the coefficients of the desired characteristic polynomial. It is possible to exploit this nonuniqueness in various ways, for instance in order to impose other design requirements like the assignment of some eigenvectors, or the optimization of some criterion, such as the minimization of $\|F\|$. Furthermore, the Pole-Shifting Theorem does not say anything about the possible Jordan form of $A + BF$, that is, the possible similarity class of the closed loop $A$ matrix. The more precise result is sometimes called Rosenbrock’s Theorem, and is given in [331]; a purely state-space proof of this Theorem was given in [117] (see also [119]).

Design issues in control theory depend on the “zeros” of a system as well as its poles (the eigenvalues of $A$). The book [212] discusses various notions of “zero” for linear systems, a topic that will not be discussed here. Recently there has been work characterizing system zeros using techniques of commutative algebra; see [436] and references therein.

Feedback Equivalence

The Brunovsky form originated with [77]; see also [224], on which our discussion is based (but note that some of the formulas in the original paper are incorrect).
Feedback Linearization

The original formulation of the feedback linearization problem was given in [71], for the restricted case treated in Exercise 5.3.18. The problem in the general form, including the multiple-input case \((m > 1)\) is from [207] and [197], and is covered in detail in textbooks such as [199] and [311]. We restricted attention to the single-input case because this is sufficient for illustrating the techniques but is technically much simpler than the general case. (In general, instead of \(A_n\) and \(b_n\) in the controller form 5.3, one must base the proof on the Brunovsky form and Theorem 14. Frobenius' Theorem must also be generalized, to deal with nested families of distributions \(\Delta^1 \subseteq \Delta^2 \subseteq \ldots\) rather than single distributions, as part of the proof of the multiinput case.) Exercise 5.3.19 can be solved directly, not using feedback linearization ideas; see for instance [311], Theorem 5.3. Example 5.3.10 is from [378].

Disturbance Rejection and Relatively Invariant Subspaces

The notion of \((A, B)\)-invariant subspace and the applications to solving systems problems, were proposed independently by [39] and [432]. It forms the basis of the approach described in [430], which should be consulted for a serious introduction to the topic of relatively invariant subspaces and many control applications. See [178] for a different manner of presenting these concepts and relations with other formalisms.

There has been much research during the last few years attempting to generalize these techniques to nonlinear systems. See, for instance, the textbooks [199] and [311] for a treatment of that topic. The basic idea is to think of the space \(V\) not as a subset of the state space, but as what is sometimes called a “distribution” of subspaces in the tangent space to \(X\). That is to say, one thinks of \(V\) as an assignment of a (possibly different) space \(V(x)\) at each \(x \in X\). (See Definition 4.4.8.) Invariance then is defined with respect to the possible directions of movement, and an analogue of Lemma 5.4.3 is obtained, under strong extra assumptions, from the classical Frobenius' Theorem on partial differential equations (cf. Theorem 10 (p. 177).) See also [166], [167], [194], [200], [257], [302], and [303] in this context, including discrete-time nonlinear versions.

Stability and Other Asymptotic Notions

There are a large number of references on stability, most of them in the classical dynamical systems literature. The paper [63] gives an excellent bibliographical survey of those results on stability that are of major importance for control. Related to stability questions in feedback systems, it is worth pointing out that recently there have been studies of control laws in relation to chaotic and other highly irregular behavior; see, for instance, [29] and [341], as well as the relevant papers in the book [273] and the very early work already discussed in [213].

Attractivity is, in general, not equivalent to asymptotic stability, contrary to what happens for linear systems (Proposition 5.5.5); for a counterexample,
see, for instance, [169], pages 191-194. The weaker notion of attractivity is less interesting, partly because positive results usually will establish the stronger property anyway, and also because of considerations that arise in input/output design. It is also possible to introduce notions associated to simply “stability” (as opposed to asymptotic stability). For differential equations (no controls) this is defined as the property that states near \( x^0 \) should not give rise to trajectories that go far from \( x^0 \). Except in the context of proving an instability result, we did not discuss this notion.

**Stable and Unstable Modes**

For a solution of Exercise 5.6.6 see [430], Chapter 4. That book also has many applications of the concepts discussed in this section.

**Lyapunov and Control-Lyapunov Functions**

There are various results which provide converses to Theorem 17, at least for systems with no controls. The book [49] covers abstract dynamical systems. For continuous-time systems \( \dot{x} = f(x) \), with \( f(0) = 0 \) and evolving in \( \mathbb{R}^n \), and under weak assumptions on \( f \) (locally Lipschitz is sufficient), Massera's Theorem, given in [296], asserts that, if the system is asymptotically stable with respect to \( 0 \) then there is a smooth Lyapunov function \( V \), which satisfies the infinitesimal decrease condition \( \dot{V}(x) = L_f V(x) < 0 \) for all \( x \neq 0 \); if the system is globally asymptotically stable, then a proper \( V \) exists. In other words, Lemma 5.7.4 is necessary as well as sufficient in the case of systems with no controls. For other proofs see also [169], [262] (allows continuous \( f \); even though solutions may not be unique, a notion of asymptotic stability can be defined in that case), and [428]. The reader may also find a proof in [279], which extends Massera’s Theorem to smooth Lyapunov functions for systems subject to disturbances (robust stability), and is presented in the more general context of stability with respect to sets rather than merely equilibria.

For control systems, the situation with converses of Lemma 5.7.4 is far more delicate. The existence of \( V \) with the infinitesimal decrease property is closely tied to the existence of continuous feedback stabilizers; see [23], [372]. If differentiability is relaxed, however, one may always find control-Lyapunov functions for asymptotically controllable systems: see the result in [363], and the recent applications to discontinuous feedback design in [92].

In any case, all converse Lyapunov results are purely existential, and are of no use in guiding the search for a Lyapunov function. The search for such functions is more of an art than a science, and good physical insight into a given system plus a good amount of trial and error is typically the only way to proceed. There are, however, many heuristics that help, as described in differential equations texts; see, for instance, [334]. Another possibility is to build control-Lyapunov functions recursively, via “backstepping” as discussed in Lemma 5.9.5 and pursued systematically in [259].
For Exercise 5.7.17, see e.g. [323].

Example 5.7.12 implies that there exists a bounded feedback of the type $u = \theta(Fx)$ which stabilizes a double integrator. The result might not be surprising, since we know (cf. Exercise 3.6.8) that the double integrator ($A$ has all eigenvalues at zero) is null-controllable using bounded controls. It is a bit surprising, on the other hand, that, for the triple integrator $\dot{x}_1 = x_2$, $\dot{x}_2 = x_3$, $\dot{x}_3 = u$, while still true that every state can be controlled to zero using bounded controls, it is in general impossible to find any globally stabilizing feedback of the form $u = \theta(Fx)$, with $\theta$ a saturation function; see [150]. On the positive side, however, slightly more complicated feedback laws do exist in general, for stabilizing linear systems with saturating inputs; see [392].

The example in Exercise 5.7.21 is due to [295]. In the adaptive control literature one finds extensive discussions of results guaranteeing, for slow-enough time variations, that a time-varying system is stable if the “frozen” systems are.

**Linearization Principle for Stability**

Even though, as discussed in Section 5.9, the linearized condition in Theorem 19 is sufficient but not necessary, this condition does become necessary if one imposes exponential stability rather than just asymptotic stability; see, for instance, [372] and references therein, as well as [157].

Theorem 20 is basically Lyapunov’s Second Theorem on Instability, proved in 1892; it is often referred to (in generalized form) as “Chaetaev’s Theorem.”

**Introduction to Nonlinear Stabilization**

The question of existence of smooth and more generally continuous feedback laws in the nonlinear case is the subject of much current research; it is closely related to the existence of Lyapunov functions and to the existence of smooth solutions to optimal control problems. See, for instance, the survey papers [372] and [101], as well as [1], [4], [23], [27], [54], [72], [92], [99], [100], [109], [186], [187], [210], [238], [268], [294], [339], [342], [355], [403], [407], and [443]. The recent textbook [259] covers many questions of nonlinear feedback design. Discontinuous feedback of various types is, of course, also of interest, and is widespread in practice, but less has been done in the direction of obtaining basic mathematical results; see, for instance, [385] for piecewise analytic feedback, [362] for piecewise linear sampled feedback, [353] for pulse-width modulated control (a type of variable rate sampling), and [92] for a notion of feedback based upon the idea of sampling at arbitrarily fast rates. A major recent result in [99] showed that time-varying smooth feedback stabilizers exist for a large class of systems (see also the related work, and alternative proof, in [374]). There is also an extensive engineering literature, e.g., [406], on sliding mode control, in which discontinuity surfaces play a central role.

The line of work illustrated by Proposition 5.9.1 and Exercise 5.9.2 is based on [202], [210], and [268]. Exercise 5.9.4 is from [8].
The assumption in Lemma 5.9.5 about the existence of a suitable Lyapunov function is redundant, since stabilizability implies it (Massera’s Theorem, quoted above). The proof of this Lemma was given independently in [80], [249], and [403]. An alternative proof of the same result about stability of cascade systems is given in [368], using only elementary definitions. A local version of the result has been known for a long time; see, for instance, [415]. For closely related work, see also [250].

Proposition 5.9.10 is essentially from [23]. The proof that we gave, based on the use of the “universal” formula (5.56), is from [369]. The result in [23] is far more general than stated here, however, especially in that it includes consideration of control-value sets \( U \) which are proper subsets of \( \mathbb{R}^m \). For universal formulas that apply to more general sets \( U \), and for applications, see e.g. [199], [259], [278]. The material that centers around Proposition 5.9.10 can be stated in necessary and sufficient form, because the existence of a stabilizing feedback implies, via the converse Lyapunov theorem cited earlier, the existence of appropriate control-Lyapunov functions; see [23].

Theorem 22 was stated originally in [72]; see [443]. The proof given here is close to that for vector fields in [254]. See also [338].

There is also a wide literature regarding the stabilization of infinite dimensional linear systems; see, for instance, the survey article [324], as well as [113] and the general references cited for infinite dimensional linear systems theory.
6.9 Notes and Comments

Basic Observability Notions

The concept of observability originates both from classical automata theory and from basic linear systems. The material on final-state observability can be generalized to certain types of infinite systems, which in the continuous-time case give, because of Proposition 6.1.9, results about observability (see, for instance, [358] and [384]).

The text [52] has an extensive discussion of the topic of multiple experiments, final-state determination, and control for finite systems. Also, [97] addresses the observability question for such systems. In the context of VLSI design a central problem is that of testing circuits for defects, and notions of observability appear; see, for instance, [148] and the references therein.

Observability of Time-Invariant Systems

The concept of observability, its duality with controllability for linear systems, and the notion of canonical realization, all arose during the early 1960s. Some early references are [155], [215], and [320].

In fact, Kalman’s original filtering paper [214] explicitly mentions the duality between the optimization versions of feedback control (linear-quadratic problem) and observer construction (Kalman filtering). The term “canonical” was used already in [217], where one can find reachability and observability decompositions. For related results on duality of time-varying linear systems, see [435].

For more on the number of experiments needed for observability, and in particular Exercise 6.2.3, see, for instance, [97], [125], and especially [156]. See also [6] for continuous-time questions related to this problem. Results on observability of recurrent “neural” nets (the systems studied in Section 3.8, with linear output $y = Cx$) can be found in [14].

In the form of Exercise 6.2.12 (which, because of the simplifying assumption that the numbers $\omega_i$ are distinct, could also have been easily proved directly), one refers to the eigenvalue criterion for sampling as Shannon’s Theorem or the Sampling Theorem. It gives a sufficient condition to allow reconstruction of the signal $\eta$ and is one of the cornerstones of digital signal processing. The result can be generalized, using Fourier transform techniques, to more general signals $\eta$, containing an infinite number of frequency components.

It is possible to develop a large amount of the foundations of time-invariant systems based on the notion of observables associated to a system; see [373].

Linearization Principle for Observability

Often one refines Definition 6.4.1 to require that states be distinguishable without large excursions. One such possibility is to ask that for each neighborhood $W$ of $x^0$ there be a neighborhood $V$ so that every state in $V$ is distinguishable
Outputs from $x^0$ using a control that makes the resulting trajectory stay in $W$. This is more natural in the context of Lyapunov stability, and can be characterized elegantly for continuous-time smooth systems; see, for instance, [185] and [365]. Other nonlinear observability references are, for instance, [5], [7], [28], [139], and [310].

Realization Theory for Linear Systems

There are papers on realization of time-varying linear systems that use techniques close to those used for the time-invariant case. See, for instance, [235] and [434].

Recursion and Partial Realization

See, for instance, [180], [181], [192], [395], and the many references therein, as well as the early paper [225], for further results on families of systems.

Small perturbations of the Markov parameters will result in a nonrealizable sequence, since all of the determinants of the submatrices of $H$ are generically nonzero. In this context it is of interest to look for partial realizations, in which only a finite part of the sequence is matched; this problem is closely related to classical mathematical problems of Padé approximation. Lemma 6.6.9 is one result along these lines; see, for instance, [19] and the references therein, for a detailed treatment of partial realization questions and relations to problems of rational interpolation. The problem can also be posed as one of optimal approximation of Hankel operators; see the by now classical paper [3]. In addition, the procedures that we described are numerically unstable, but various modifications render them stable; a reference in that regard is [110]. A recursive realization procedure is given in [18], which permits realizations for additional data to make use of previous realizations.

Since a Markov sequence is specified in terms of an infinite amount of data, one cannot expect to solve completely the question of realizability unless some sort of finite description is first imposed. It is known however that, for arbitrary “computable” descriptions, the problem of deciding realizability is undecidable in the sense of logic, that is, there is no possible computer program that will always correctly determine, given a description of a Markov sequence, whether a realization exists; see, for instance, [356].

Rationality and Realizability

The ideas of realization and relations to input/output equations go back at least to the nineteenth century, in the context of using integrators to solve algebraic differential equations; see [400] as well as the extensive discussion in [212], Chapter 2. The relations between rationality and finite Hankel rank are also classical (for the scalar case) and go back to the work of Kronecker (see [152]).
Algebraic techniques for studying realizations of linear systems were emphasized by Kalman; see, for instance, [220], [221], and [223], as well as [228] for relations to econometrics.

Proposition 6.7.6 could also be proved using Laplace transform techniques; we used a direct technique that in principle can be generalized to certain nonlinear systems. For these generalizations, see, for instance, [360] and [421], which establish that certain types of i/o behaviors satisfy polynomial equations

\[ E(y(t + n), y(t + n - 1), \ldots, y(t), u(t + n - 1), \ldots, u(t)) = 0 \quad (6.44) \]

(or \( E(y^{(n)}(t), y^{(n-1)}(t), \ldots, y(t), u^{(n-1)}(t), \ldots, u(t)) = 0 \)) if and only if they are realizable, for discrete-time and continuous-time systems, respectively. In the nonlinear case, however, not every “causal” i/o equation necessarily gives rise to an i/o behavior, and this in turn motivates a large amount of research on such equations; see, for instance, [408]. Some authors, motivated by differential-algebraic techniques, have suggested that realizability should be defined in terms of i/o equations; see especially [141] and the references therein.

The reference [272] discusses i/o equations for nonlinear systems in the context of identification problems. Related material is also presented in [307], which uses i/o equations (6.44) in which \( E \) is not linear, nor polynomial, but instead is given by iterated compositions of a fixed scalar nonlinear function with linear maps. Numerical experience seems to suggest that such combinations are particularly easy to estimate using gradient descent techniques, and they are in principle implementable in parallel processors. The name “neural network” is used for this type of function \( E \), because of the analogy with neural systems: the linear combinations correspond to dendritic integrations of signals, and the scalar nonlinear function corresponds to the “firing” response of each neuron, depending on the weighted input to it.

Realization and i/o equations can also be studied in a stochastic context, when \( u(0), u(1), \ldots \) in (6.43) are random variables. In that case, especially if these variables are independent and identically distributed, (6.43) describes what is called an ARMA or autoregressive moving average model for the stochastic process, in the stochastic systems and statistics literature. (The outputs form a time series, and the realization is a Markov model for the series.) See, for instance, the “weak” and “strong” Gaussian stochastic realization problems studied, respectively, in [131], [132], [222], [409], and [10], [11], [280], and the related [344].

**Abstract Realization Theory**

It is possible to develop a unified theory of realization for discrete-time finite systems and linear finite dimensional systems, in the language of category theory; see, for instance, [22]. There are also many other approaches to realization theory, for instance for nonlinear systems evolving on manifolds in continuous-time ([199], [383], and [389]) or in discrete-time ([203]), systems on finite groups.
(75), polynomial discrete-time (360) and continuous-time (38) systems, bilinear systems (66, 198, 359, and 382), or infinite-dimensional linear systems (316, 437, and 438). The goal in these cases is to study realizations having particular structures—such as linear realizations when dealing with linear behaviors—and to relate minimality (in appropriate senses) to various notions of controllability and observability. It is also possible to base a realization theory on generalizations of Hankel matrix ideas to bilinear and other systems; see, for instance, 137 and 138, as well as applications of these ideas in 96. One also may impose physical constraints such as in the study of Hamiltonian control systems in 106 and 204.

In principle, the complete i/o behavior is needed in order to obtain a realization. However, if the system is “mixing” enough, so that trajectories tend to visit the entire state space, one may expect that a single long-enough record will be sufficient in order to characterize the complete behavior. In this sense one can view the recent work identifying dynamics of chaotic systems from observation data (see, e.g., 85 and 130) as closely related to realization theory (for systems with no controls), but no technical connections between the two areas have been developed yet.
7.6 Notes and Comments

Observers and Detectability

Asymptotic observability is obviously a necessary condition that must be satisfied if there is any hope of even asymptotically final-state distinguishing the state $x^0$ from other states. Any algorithm that uses only the output measurements will not be able to differentiate between $x^0$ and states in $X^0$ if the input happens to be constantly equal to $u^0$. So unless these states already converge to $x^0$, there is no possibility of consistent estimation. For linear systems, detectability as defined here turns out to be equivalent to the existence of observers, but in the general nonlinear case a stronger definition is required (see, for instance, [361]). Observers can be seen as a deterministic version of Kalman filters, introduced originally in [214] and [229]; the deterministic treatment seems to have originated with work of J.E. Bertram and R.W. Bass, but the first systematic treatment was due to Luenberger in the early 1960s; see [287] for references and more details. The construction of reduced order observers in Exercise 7.1.5 is taken from [287], who followed [159].

The terminology “strong” in Definition 7.1.3 is meant to emphasize that the observer must provide estimates for all possible controls, and starting at arbitrary initial states. It is possible—and even desirable—to weaken this definition in various ways, by dropping these requirements. For linear systems, all reasonable definitions (even requiring that there be some initial state for the observer that results in converging estimates when the zero control is applied to the original system) are equivalent in the sense that the existence of an observer of one type implies the existence of observers of the other types. For nonlinear systems, these different notions are of course not equivalent; see, for instance, [87]. Some references on observer construction for nonlinear systems are [87], [164], [209], and [253]. There are technical reasons for not making the definition slightly more general and allowing $\theta$ to depend on $\omega(t)$: This would mean that the instantaneous value of $\hat{\xi}(t)$ may not be well defined when dealing with continuous-time systems, as there $\omega$ is in general an arbitrary measurable function.

Dynamic Feedback

The design of dynamic controllers using a combination of an observer and a state feedback law is a classical approach. It turns out that even optimal solutions, when the (linear quadratic) problem is posed in the context of stochastic control, also have this separation property. For nonlinear stabilization using such techniques, see, for instance, [416] and [419]. For linear systems over rings, see, for instance, [129] and [246]; for infinite dimensional linear systems, see [201] and the references given therein.
External Stability for Linear Systems

The topic of external stability for linear infinite dimensional and for nonlinear systems has been the subject of much research. The books [115], [414], and [422] provide further results and references, and [410] was an influential early paper.

The notion of input-to-state stability (ISS) has recently become central to the analysis of nonlinear systems in the presence of disturbances. The definition is from [370]; for recent results see [376]. In the latter reference, it is shown that a system is ISS if and only if it is internally stable, complete, and has a finite nonlinear gain. The textbook [259] discusses at length applications of ISS and related properties to practical nonlinear feedback design problems. The “dual” notion of OSS is given in [377], where a generalization to systems with inputs as well as outputs (“input/output-to-state stability” or IOSS) is also considered.

For questions of i/o stability versus stability in a state-space setting for infinite dimensional linear systems see, e.g., [285], [324], and [439].

Frequency Domain Considerations

There is a very large number of textbooks dealing with complex variables techniques for the analysis of linear systems. A modern presentation of some of these facts, together with references to a bibliography too extensive to even attempt to partially list, can be found in [144].

Parametrization of Stabilizers

See, for instance, [417] and [145] for detailed developments of the theory of parametrizations of stabilizing compensators. Theorem 34 is due to Youla and coworkers, and was given in [441]. In its version for arbitrary $m, p$ it is the basis of many of the recent developments in multivariable linear control. Recently there have been attempts at generalizing such parametrizations to nonlinear systems as well as obtaining the associated coprime factorizations; see, for instance, [116], [171], [172], [258], [370], [398], and [412].
8.6 Notes and Comments

Dynamic Programming

The dynamic programming approach gained wide popularity after the work of Bellman; see, for instance, [41]. Almost every optimization or optimal control textbook contains numerous references to papers on the topic.

The Continuous-Time Case

A central problem in optimal control is the study of existence and smoothness properties. The question of existence often can be settled easily on the basis of general theorems; what is needed is the continuous dependence of $J$ on $\omega$, with respect to a topology on the space of controls which insures compactness. For bounded inputs and systems linear on controls, Theorem 1 (p. 57), part 2(ii), together with compactness in the weak topology, achieves this purpose. For more general systems, analogous results hold provided that one generalizes the notion of control to include “relaxed controls.” See, for instance, [134], [188], [266], or any other optimal control textbook; here we only proved results for a very special class of systems and cost functions, and for these existence can be established in an ad hoc manner.

Continuity of controls, and smoothness of Bellman functions, are a much more delicate matter. It turns out that many problems of interest result in a nonsmooth $V$. The regular synthesis problem in optimal control theory studies generalizations of the results that assumed smoothness to cases where $V$ is piecewise smooth in appropriate manners. See, for instance, [134] for more on this topic, as well as [50] and [386]. An alternative is to generalize the concept of solution of a partial differential equation: The notion of viscosity solution provides such a possibility, cf. [102]; or the related approach based on proximal subgradients can be used, cf. [93].

(Deterministic) Kalman Filtering

A purely deterministic treatment was pursued here so as not to require a large amount of preliminary material on stochastic processes. On the other hand, a probabilistic study is more satisfactory, as the cost matrices have a natural interpretation and formulas can be obtained that quantify the accuracy of the estimator (its covariance is given in fact by the solution to the Riccati equation).

There are excellent textbooks covering Kalman filtering, such as [108] or, for more of an engineering flavor, [16] and [263]. A good reference for practical implementation issues, as well as a detailed derivation in the discrete-time case, is provided by [153]. A very elegant discussion of the discrete-time problem is given by [86], from the point of view of recursive least-squares.

Much has been done regarding nonlinear filtering problems; for some of the theoretical issues involved see, for instance, [209], [293], [312], and the references therein.
Historically, the Kalman filter, introduced for discrete-time in [214] and for continuous-time in [229], appeared in the context of the general problem of “filtering” a signal corrupted by noise. As compared to older Wiener filtering ideas, the Kalman filter modeled the signal of interest as the state trajectory of a linear system driven by white noise. This allowed a highly computable method of solution as well as extensions to nonstationary processes (time-varying state models) and nonlinear systems.

The optimal estimation and optimal control problems can be combined into the stochastic optimal control problem of minimizing a cost criterion for a system such as the one in Exercise 8.3.9, by suitable choice of controls $u$, on the basis of noisy observations; the solution to this “LQG” problem — “linear quadratic Gaussian problem,” when all noises are assumed to be Gaussian processes — can be obtained from the solutions of the two separate deterministic linear quadratic and linear filtering problems, in much the same fashion as output stabilizers were obtained by combining observers and optimal state feedback laws; see [16] and [263], for instance.

**Linear Systems with Quadratic Cost**

The treatment of the Linear Quadratic Problem started with Kalman’s seminal paper [215]. The literature generated by this problem is immense, as many variants and numerical approaches have been tried. Two excellent texts that deal with such issues are [16] and [263]. The paper [423] discusses many important issues about Riccati equations, including the indefinite case, of interest in game theory as well as in modern $H_{\infty}$ optimization.

The infinite-dimensional linear case also has been studied in detail; see, for instance, [113] and [133].

**Infinite-Time Problems**

The computation of solutions to the ARE via Hamiltonian matrices was first suggested by [288] and [322] for the case of distinct eigenvalues; see [212], Section 3.4, for a discussion and many references to numerical techniques associated to this approach.

The result in Exercise 8.4.12 is from [112] (see also [64]).

It is interesting but not very surprising that quadratic problems for linear time-invariant systems give rise to linear solutions (a linear feedback law). When other criteria are used, or when “robust” design is desired, nonlinear or time-varying controllers may be superior, even for linear time-invariant systems. This type of question is explored in [244], [245], and related papers.

Substantial activity has taken place recently on the topic of $H_{\infty}$ optimization, dealing with a different optimization criterion than the linear-quadratic problem. This criterion, which corresponds to the minimization of an operator norm from external “disturbances” to outputs, is of great engineering interest. A book reference is [145], and recent research, resulting in a reduction to a Riccati
equation problem, is surveyed in [124]. That reference also shows how the linear-quadratic problem can be viewed as a minimization problem in an input/output context. The norm itself can be computed numerically; see, for instance, [59].

**Nonlinear Stabilizing Optimal Controls**

The main sense in which the nonlinear results given in Section 8.5 differ from the linear case is in the need to assume that $V$ exists. For linear systems, the ARE always has a solution (under the appropriate controllability conditions). In order to obtain necessary and sufficient results, one must introduce generalized solutions of various types (viscosity, proximal); see for instance [93] and the many references provided therein.
9.6 Notes and Comments

The material on optimization and the Calculus of Variations is classical; see, for instance, [48], [134], [151], [189], [247], [248], [402], [442]. Path minimization problems were posed and solved at least since classical times (Dido’s problem, to find a curve of a given length which encloses the largest possible area, appears in Virgil’s writings). However, the origins of the field may be traced to the challenge issued “to all mathematicians,” in 1696, by Johann Bernoulli, to solve the brachystochrone problem: “If in a vertical plane two points A and B are given, then it is required to specify the orbit AMB of the movable point M, along which it, starting from A, and under the influence of its own weight, arrives at B in the shortest possible time.” This problem, which was soon afterwards solved by Newton, Leibniz, and many other mathematicians of the time (including Johann and his brother Jakob), gave rise to the systematic study of the Calculus of Variations. The article [393] explains the historical development, taking the more general point of view afforded by optimal control theory.

A good reference for the material in Section 9.4 is [78]. Under mild conditions, it is possible to prove that generic controls are nonsingular for every state $x^0$; see [373] and [374], and the related papers [99] and [100].

Section 9.5 merely skims the surface of the subject of the Minimum (or Maximum) Principle. In particular, we did not provide a treatment of the more general case in which final-state constraints $K(x) = 0$ are present. However, the form of the result is easy to guess, and we state it precisely here; see [134] for a proof. Suppose that $K : X \rightarrow \mathbb{R}^r$ is also of class $C^1$. Assume that $\bar{\omega} \in L^u(\sigma, \tau)$ is optimal for $x^0$. Then, there exist a scalar $\nu_0 \in \{0, 1\}$ and a vector $\nu \in \mathbb{R}^r$, not both zero, so that the solution $\lambda : [\sigma, \tau] \rightarrow \mathbb{R}^n$ of the final-value problem $\dot{\lambda}(t) = -\nu_0 g_x(\bar{x}(t), \bar{\omega}(t)) - A(t)'\lambda(t)$ with $\lambda(\tau) = (\nu_0 p_x(\bar{x}(\tau))) + \nu K_x(\bar{x}(\tau))$ (cf. (9.13)) is so that $H(\bar{x}(t), \bar{\omega}(t), \nu_0, \nu(t)) = \min_{u \in U} H(\bar{x}(t), u, \nu_0, \lambda(t))$ for almost all $t \in [\sigma, \tau]$.

It is also possible to give even more general results, for instance, to deal with mixed initial/final state constraints, as well as non-fixed terminal time and time-optimal problems. For the latter, one must consider variations in which a piece of the control is deleted: such shorter controls cannot be optimal, and the directions generated in this manner lead to an appropriate Hamiltonian. (The only time-optimal problem considered in this text, cf. Chapter 10, is for linear systems, and can be developed in a simple and self-contained fashion using elementary techniques from convex analysis.)

Several extensions of the Maximum Principle have been developed during the past few years. These provide “high order” tests for optimality, and in addition permit the study of far more general classes of systems, including those in which the dynamics does not depend in a Lipschitz continuous manner (or is even discontinuous) on states. A promising direction, in [390], develops an approach to generalized differentials (“multidifferentials”), and proposes their use as a basis for a general nonsmooth version of the maximum principle; references to the extensive literature on the subject are also found there.
10.6 Notes and Comments

Weak convergence as defined here is a particular case of a concept that is standard in functional analysis. In general, if $N$ is a normed space and $N^*$ is its dual (the set of continuous linear functionals $N \rightarrow \mathbb{R}$), the weak-star topology on $N^*$ is the topology defined by the property that a net $\{x^*_k\}$ in $N^*$ converges to $x^* \in N^*$ if and only if $x^*_k(x) \rightarrow x^*(x)$ for all $x \in N$. In particular, we may take $N = L^1_m(0,T)$, the set of integrable functions $\varphi : [0,T] \rightarrow \mathbb{R}^m$ with the norm $\int_0^T \|\varphi(s)\| \, ds$. The space $L^\infty_m(0,T)$, with the norm $\|\cdot\|_\infty$, can be naturally identified with the dual of $N$ via

$$\omega(\varphi) := \int_0^T \varphi(s)' \omega(s) \, ds.$$ 

Thus, a net $\omega_k$ (in particular, a sequence) converges to $\omega$ in the weak-star sense if and only if $\int_0^T \varphi(s)' \omega_k(s) \, ds \rightarrow \int_0^T \varphi(s)' \omega(s) \, ds$ for all integrable functions $\varphi : [0,T] \rightarrow \mathbb{R}^m$. So what we called weak convergence is the same as weak-star convergence. Alaoglu’s Theorem (see e.g. [98], Theorem V.3.1) asserts that the closed unit ball in $N^*$ (or equivalently, any closed ball) is compact in the weak-star topology. This proves Lemma 10.1.2. However, we provided a self-contained proof.

In connection with Remark 10.1.12, we note that it is possible to relax the convexity assumption; see [188], which may also be consulted for many more results on the material covered in this chapter.

In connection with Remark 10.4.2, the reader is directed to [93].