The Dissertation Committee for Johnson Edward Hawes Carroll certifies that this is the approved version of the following dissertation:

Uniform Positive Recurrence and Long Term Behavior of Markov Decision Processes, with Applications in Sensor Scheduling

Committee:

Aristotle Arapostathis, Supervisor

Ross Baldick

Constantine Caramanis

W. Mack Grady

Raul G. Longoria

Uniform Positive Recurrence and Long Term Behavior of Markov Decision Processes, with Applications in Sensor Scheduling

by

Johnson Edward Hawes Carroll, B.A.; B.S.E.E.; B.S.Math; M.S.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2015

Acknowledgments

I would like to thank my wife Anita, who has supported and encouraged me far beyond my deserving. Without her support and understanding none of this would have been possible, and without her love and respect none of this would have been worthwhile.

I would like to thank my supervisor, Professor Ari Arapostathis, for his guidance, encouragement, and enthusiasm. His passion has inspired me to reach for and achieve heights of understanding that I would never have believed possible.

Uniform Positive Recurrence and Long Term Behavior of Markov Decision Processes, with Applications in Sensor Scheduling

Publication No. _____

Johnson Edward Hawes Carroll, Ph.D.
The University of Texas at Austin, 2015

Supervisor: Aristotle Arapostathis

In this dissertation, we show a number of new results relating to stability, optimal control, and value iteration algorithms for discrete-time Markov decision processes (MDPs). First, we adapt two recent results in controlled diffusion processes to suit countable state MDPs by making assumptions that approximate continuous behavior. We show that if the MDP is stable under any stationary policy, then it must be uniformly so under all policies. This abstract result is very useful in the analysis of optimal control problems, and extends the characterization of uniform stability properties for MDPs. Then we derive two useful local bounds on the discounted value functions for a large class of MDPs, facilitating analysis of the ergodic cost problem via the Arzelà-Ascoli theorem. We also examine and exploit the previously underutilized Harnack inequality for discrete Markov chains; one aim of this work was to discover how much can be accomplished for models with this property.

Convergence of the value iteration algorithm is typically treated in the literature under blanket stability assumptions. We show two new sufficient conditions for the convergence of the value iteration algorithm without blanket stability, requiring only geometric ergodicity under the optimal policy. These results form the theoretical basis to apply the value iteration to classes of problems previously unavailable.

We then consider a discrete-time linear system with Gaussian white noise and quadratic costs, observed via multiple sensors that communicate over a congested network. Observations are lost or received according to a Bernoulli random variable with a loss rate determined by the state of the network and the choice of sensor. We completely analyze the finite horizon, discounted, and long-term average optimal control problems. Assuming that the system is stabilizable, we use a partial separation principle to transform the problem into an MDP on the set of symmetric, positive definite matrices. A special case of these results generalizes a known result for Kalman filters with intermittent observations to the multiple-sensor case, with powerful implications.

Finally, we show that the value iteration algorithm converges without additional assumptions, as the structure of the problem guarantees geometric ergodicity under the optimal policy. The results allow the incorporation of adaptive schemes to determine unknown system parameters without affecting stability or long-term average cost. We also show that after only a few steps of the value iteration algorithm, the generated policy is geometrically ergodic and near-optimal.

Table of Contents

Chapter 1

Introduction

1.1 Summary of Contributions

This section serves as a guide to assist the reader in identifying the central results, as the dissertation contains a large number of supporting technical lemmas and theorems. The main contributions are highlighted below, with relevant theorem numbers noted in parentheses.

(a) For an MDP on a countable state space, we show that if all stationary policies are stable then the induced chains are *uniformly recurrent (Theorem 4.2.1)*. That is, one can find a uniform bound on any function integrated up to the time the chain hits a finite set. This abstract result is very useful in the analysis of optimal control problems, and extends a known result on uniform stability of MDPs. Next, substantial effort is usually expended in the literature to apply Arzela-Ascoli to the discounted value functions in order to pass to the ergodic optimality equation. We show that for a large class of problems this is unnecessary, by demonstrating *regularity properties of the discounted value functions (Theorem 4.2.2)*. An essential element of these two results is a version of the Harnack inequality for discrete Markov chains. Though the concept is not new and the result

not complex, the inequality has been underutilized in previous work. One of the goals of this study was to explore results for MDPs satisfying this property.

(b) In the literature, value iteration results are typically shown under blanket stability assumptions. We show *convergence of the value iteration (Theorems 5.3.1–5.3.2)* given geometric ergodicity under only the optimal policy, along with a norm-like running cost function. Without blanket stability, one can then consider the value iteration for systems previously unapproachable, including those like the linear quadratic system described next.

(c) Finally, for linear quadratic systems with Gaussian white noise and intermittent observations, we completely analyze the *optimal sensor scheduling problem (Theorems 7.3.1, 7.4.2, and 7.5.3)*. We also show a generalization and extension of a known result on the *critical loss rates for intermittent observations (Theorems 7.6.1–7.6.2)*. The structure of the system means that the optimal control guarantees geometric ergodicity, allowing us to prove *convergence of the value iteration (Theorems 8.3.1–8.3.2)* without additional assumptions, just as with a countable state space. Additionally, the system structure and method of proving the main results allow the incorporation of adaptive schemes to determine unknown system parameters and guarantee that only a few steps of the value iteration algorithm will produce a stable, near-optimal control.

1.2 Background and Motivation

We begin with a discrete time controlled dynamic process: at each time step the system state is observed, the controller chooses a control action, and a cost is incurred based on the state and control action. The system then evolves according to some transition rule (presumably dependent on the control action), and the process repeats. The goal of the controller is to select the control actions that will incur the least cost over some time horizon. Such a process can be called a decision process, and the formulation is astonishingly general.

The decision process model does not require that the state evolution be deterministic, nor that the observation be perfect. Indeed, many of the more potent and interesting results apply to systems with inherent randomness, such as economic forecasting, queuing theory, and population dynamics. When choosing control actions for a stochastic system, the goal is often to incur the least expected cost over the time horizon, though other stochastic rubrics are possible.

We are primarily interested in those decision processes which are also Markov. Formally, a process is Markov if, given the entire knowledge of the process up to the present time, only the system state and control *at* the present time is useful for predicting future system behavior. As an example, consider a simple linear system

$$x_{t+1} = -x_t + 1, \quad t = 0, 1, 2, \ldots.$$

When the system model is known exactly, knowing x_2 allows one to predict x_t for $t > 2$; *also* knowing x_0 and x_1 does not improve one's prediction. Markov processes are desirable in that they are ubiquitous, appearing in numerous scientific and engineering fields, and that they are computationally more tractable than non-Markov processes. Modeling future events does not require the storage or analysis of the entire state trajectory; only the current state must be considered. Also, though it is not assumed, we seek decision rules or policies that are Markov, so that the optimal control actions can be determined only from the current observation. When the optimal control policies are Markov, one can frequently calculate the optimal control action in "real time," at the moment of decision, via so-called dynamic programming algorithms.

The study of MDPs has its roots in sequential decision making methods developed in the 1940s [57], but the core of stochastic MDP analysis was developed in the 1950s with the group of researchers at RAND, most notably Richard Bellman [7, 8, 27]. His eponymous equation recursively calculates the expected cost of using a particular policy to choose control actions, and with the inclusion of a one-step minimization becomes the test of optimality. As research into MDPs expanded, it quickly incorporated infinite time horizons through discounting the future costs or averaging over a receding horizon, resulting in the average expected one-step cost. Analysis also incorporated countable and continuous state spaces, despite the lack of computational methods to implement the results. Research developments frequently followed a common pattern, beginning with restrictive assumptions and steadily expand-

ing to include more and more general results. An extensive survey of research and results is given in [2], and the various bibliographical notes in [47] also provide a thorough perspective on early and current research.

As significant advances in computation and analysis of MDPs continue to be made, the number of applications of MDPs expands accordingly. Systems posed abstractly by early researchers are now finding practical, computational uses, and the number of fields utilizing MDPs continues to grow. Discipline-specific texts now proliferate, including finance [18], management [54], artificial intelligence [55], and more.

1.3 Organization and Contents

In this work, we introduce new results on MDPs that expand the class of problems that can be analyzed. As mentioned, our focus is on average cost problems, but we also introduce several concepts and results that contribute to the overall body of MDP research. As the field of MDP research continues to expand and move forward, we fully expect future researchers, scientists, and engineers to find new and unexpected ways to apply this knowledge. Since the results presented here cover topics in various settings, we have arranged the subsequent chapters to reflect the conceptual grouping of results.

In Chapter 2, we formally introduce the general MDP model that will be utilized throughout the subsequent chapters. The basic structure and some intrinsic assumptions are discussed. We introduce stability in terms of MDPs, and define the fundamental optimization problems that allow the control of

MDPs: minimizing a cost function over a defined period of time. The most basic problem is the *finite horizon* problem, which adds the cost over a fixed number of time steps with a terminal penalty. The *infinite horizon, discounted cost* problem considers the cumulative cost for all future times, but multiplies the cost by a discount factor at each step, indicating that future costs are not as important as immediate, present costs. Finally, the *average cost* problem (also known as the *ergodic cost* problem) deals with the expected average cost over time without discount.

We then proceed to study MDPs on particular state spaces. First, Chapters 3–5 explore results in countable state spaces; that is, state spaces that are equivalent to the infinite set of positive integers $\mathbb{N}$. Chapter 3 defines the details of the MDP model with a countable state space, and explores some of the characteristics of Markov processes on such a space. The countable space is topologically very different from $\mathbb{R}^n$, and notions such as continuity and compactness take on different meanings. Many of our results are based on similar results for continuous processes on $\mathbb{R}^n$, so we introduce several structures and assumptions that will be utilized later to facilitate analysis on the countable state space. We also prove countable-space versions of several results from continuous diffusion processes. Notably, Harnack's inequality, though not difficult to prove, provides insight into the behavior of families of linear operators. We also show the discrete version of the Dirichlet problem, and present an appropriate version of Dynkin's formula that proves repeatedly useful for MDP analysis.

Chapter 4 begins with a result on the uniformity of recurrence properties for MDPs on countable spaces. Building on recent results in continuous diffusion processes, the result requires particular assumptions are made about the structure of the transition matrix, emulating in a general way the continuity properties needed to support the analysis. The same framework is also used to show local equicontinuity and a local uniform boundedness property of the discounted value function. Such a result facilitates the natural extension of the discounted cost problem to the more difficult average cost problem.

Directly finding the optimal average cost and corresponding control policy involves simultaneously solving for a constant and a function on the entire space, so is inherently intractable when the state space is infinite. In Chapter 5, we therefore seek conditions under which the well-known value iteration algorithm will converge to a solution. We assume that the cost function satisfies a near-monotone condition, which penalizes the system for moving away from a "central" set. We provide two new sufficient conditions for the convergence of the value iteration algorithm, neither of which rely on the blanket stability conditions commonly assumed in the literature. Instead, our first result assumes only that the value function is integrable with respect to the optimal cost (an assumption on the optimal policy only), and our second assumes that the cost function and value function have the same growth. These results greatly expands the applicability of the value iteration to new problems, and the structure of the assumptions motivates our extension to the linear systems considered in the following chapters.

Chapters 6–8 move to an entirely different state space: the product of a finite set and the set of symmetric positive definite matrices of fixed dimension. As detailed in Chapter 6, the problem is based on the control of a linear system with Gaussian noise and a quadratic cost function: the so-called LQG system. Our system is observed via a finite number of sensors over a congested network subject to random intermittency. At each time, the controller chooses the input to the linear system and the sensor to be scheduled next. The network congestion is modeled as a finite-state Markov process that evolves based on the sensor selected; the probability that an observation is lost depends on the network state and the chosen sensor. We derive a Kalman filter estimate of the linear system incorporating the intermittent observations and network congestion, and show some useful properties of the error covariance update operator.

In Chapter 7, we show that for any sensor scheduling policy, the optimal control for each of the optimal control problems consists of a predetermined linear gain with the estimate of the state. Combined with the Kalman filter estimate from the previous chapter, this allows the entire problem to be recast into an MDP on the product of the network states and the set of possible error covariances; that is, positive definite matrices. We derive new algebraic optimality conditions for each of the optimal control problems. Following the traditional method, we extend the horizon of the finite horizon problem to approach solutions to the discounted problem, then show that as the discount factor increases the limiting functions and policies are average cost optimal.

Of key importance is the concavity of the value function, which allows us to generate estimates and bounds based on the trace of the error covariance. We also present a special case of particular interest: with the network congestion depending only on the sensor scheduled (i.e., a single, constant network state), the system becomes a generalized version of a popular intermittent observation model. We show that when each sensor has a different loss rate, there is a critical hypersurface: the system is not stabilizable if and only if the vector of loss rates lies above the hypersurface.

Chapter 8 shows the value iteration algorithm converges for the LQG system, recast in terms of the error covariance. Unlike in Chapter 5, the LQG system intrinsically satisfies near-monotonicity and geometric ergodicity by virtue of the concavity of the value function. The proof of convergence then follows naturally along the same lines as in Chapter 5, and notably does not depend on the loss rates. This suggests that if the loss rates were not known, a straightforward estimator or adaptive algorithm could estimate the loss rates without affecting the long term average cost. Additionally, we show that the sub-optimal control policy found after finitely many steps of the value iteration algorithm is in fact a stable, near-optimal policy. Further, the convergence to the optimal average cost is geometric, a result with significant implications for computational effort.

Finally, Chapter 9 reviews the main contributions and discusses possible extensions for future research.

1.4 General Mathematical Notation

The following standard notation will be used throughout:

- $\mathbb{R}$ is the set of real numbers
- $\mathbb{R}_+$ is the set of non-negative real numbers
- $\mathbb{N}$ is the set of non-negative integers
- For a topological space $\mathcal{X}$, $\mathcal{C}(\mathcal{X})$ is the set of continuous real-valued functions on $\mathcal{X}$, and $\mathcal{C}_+(\mathcal{X}) \subset \mathcal{C}(\mathcal{X})$ be the set of non-negative functions in $\mathcal{C}(\mathcal{X})$. When $\mathcal{X}$ is finite or countable (as in $\mathcal{X} = \mathbb{N}$), continuity is superfluous and each $\varphi \in \mathcal{C}(\mathbb{N})$ is equivalently represented as a (possibly infinite) row vector.
- For a Borel space $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ is the set of probability measures on $\mathcal{X}$ endowed with the topology of weak convergence.
- $\mathbb{P}$ and $\mathbb{E}$ are the classical probability measure and expectation operator.

Chapter 2

Discrete Time Markov Decision Processes

2.1 Introduction

In this chapter we present the underlying system model that will be utilized throughout the rest of the dissertation. We begin by defining a Markov decision process (MDPs) on a general state space and define some of the essential properties of MDPs. Then, in the subsequent chapters, we will revise the portions of the model and properties relevant to the specific systems being considered. Though some of the definitions are somewhat convoluted compared to the versions that appear in later chapters, the underlying general model forms the link between the specific models studied later.

2.2 MDP Model

An MDP is an $(\mathbb{S} \times \mathbb{U})$-valued stochastic process $\{(X_n, U_n) : n \in \mathbb{N}\}$, where the *state space* $\mathbb{S}$ and the *control space* $\mathbb{U}$ are Borel spaces. We will refer to $\{X_n\}$ and $\{U_n\}$ respectively as the state process and control process, and unless otherwise specified we will assume that $\mathbb{U}$ is a compact metric space. The initial state is an $\mathbb{S}$-valued random variable with distribution $\mu \in \mathcal{P}(\mathbb{S})$, and the state process dynamics are governed by a transition kernel P that

depends on the control. For any $u \in \mathbb{U}$, $x \in \mathbb{S}$, and measurable set $A \subset \mathbb{S}$,

$$P^u(x, A) \;:=\; \mathbb{P}\big(X_{n+1} \in A \mid X_n = x, U_n = u\big)\,.$$

Intuitively, the probability of the process being in a particular set at a particular time are determined by the state and control of the process in the immediately previous time; this is the Markov property of the process. Note that for a fixed $u \in \mathbb{U}$, a transition kernel P^u can interchangeably be treated as:

- an operator on probability measures: $\mu[P^u](A) = \int_{\mathbb{S}} \mu(dx) P^u(x, A)$; and
- an operator on appropriately integrable functions on the state space: $P^u f(x) = \int_{\mathbb{S}} P^u(x, dy) f(y) = \mathbb{E}[f(X_1) \mid (X_0, U_0) = (x, u)]$.

In order to appropriately relate limits in the control space to limits in the induced probability distributions, we assume that transition probabilities P^u_{xy} are continuous in u. This assumption is fairly standard [16, 22], though some authors make this assumption unnecessary by considering only finite or countable control spaces, [1, 9, 52].

To simplify notation and analysis, we will assume that all control actions $u \in \mathbb{U}$ are possible from any state $x \in \mathbb{S}$. This does not affect the generality of results: if $(x, u) \in \mathbb{S} \times \mathbb{U}$ is an impossible state-action combination, the kernel $P^u(x, \cdot)$ can be changed to match $P^{u'}(x, \cdot)$ for some action $u' \in \mathbb{U}$ that *is* possible in state x. At each time n, the system state is X_n, the control U_n is chosen according to some decision criteria, and a cost $r(X_n, U_n)$

is incurred, where the cost function $r : \mathbb{S} \times \mathbb{U} \to \mathbb{R}$. Hence, the MDP is defined by the tuple $(\mathbb{S}, \mathbb{U}, P, \mu, r)$, and the decision-maker's task is to define the control process, usually with the goal of minimizing the cost in some manner.

Because we will frequently be conditioning on the initial distribution μ, we will use the shorter notation

$$\mathbb{E}_\mu[\,\cdot\,] = \mathbb{E}[\,\cdot\,|X_0 \sim \mu], \qquad \mathbb{P}_\mu(\,\cdot\,) = \mathbb{P}(\,\cdot\,|X_0 \sim \mu).$$

When μ is a Dirac measure δ_x (i.e., $\mathbb{P}(X_0 = x) = 1$) for some $x \in \mathbb{S}$, we will abuse notation by simply writing x instead of μ:

$$\mathbb{E}_x[\,\cdot\,] = \mathbb{E}[\,\cdot\,|X_0 = x], \qquad \mathbb{P}_x(\,\cdot\,) = \mathbb{P}(\,\cdot\,|X_0 = x).$$

2.3 Policies

For each $n \in \mathbb{N}$, we define the *history* of the state process up to n as the σ-algebra generated by the chain up to n and the control process up to $n-1$:

$$\mathcal{F}_n \;:=\; \sigma(X_0, \ldots, U_{n-1}, X_n).$$

The control process $U = \{U_0, U_1, \ldots\}$ is called *admissible* if for each n, U_n is $\mathcal{F}_n$-measurable, and we denote the set of all admissible controls $\mathfrak{U}$. A *policy* or *control strategy* $v = \{v_0, v_1, v_2, \ldots\}$ is a sequence of probability measures on $\mathbb{U}$ that govern the control process dynamics:

$$\mathbb{P}(U_n \in A) = v_n(A) \quad \text{for } A \text{ a measurable subset of } \mathbb{U}\,.$$

We denote the set of admissible policies Π, and note that each element v_n of an admissible policy v is the probability distribution of U_n on $\mathbb{U}$. Following the framework of [14], an admissible control U is called *randomized Markov* if for each $n \in \mathbb{N}$, $(U_n|X_n)$ is independent of $\{(X_m, U_m) : m < n\}$. Then the corresponding *Markov policy* v is can be treated as a sequence of functions $v_n : \mathbb{S} \to \mathcal{P}(\mathbb{U})$ such that the distribution of $(U_n|X_n)$ is given by $v_n(X_n)$. If additionally $v_0 = v_1 = v_2 = \cdots$, the policy and corresponding control are called *stationary Markov*. We denote the set of stationary Markov controls (policies) as $\mathfrak{U}_{sm}$ (Π_{sm}). With a slight abuse of notation, when $v \in \Pi_{sm}$ we will interchangeably refer to the policy and the set of component functions $\mathbb{S} \to \mathcal{P}(\mathbb{U})$ as v. When $\mathbb{U}$ is a compact metric space, $\mathcal{P}(\mathbb{U})$ is metrizable in the topology of weak convergence [13]. In the case where $\mathbb{S}$ is countable, [14] shows that Π_{sm} is compact, which will prove useful in several results. Notably, we will frequently use the notion of sequential compactness: every sequence $\{v_n\} \in \Pi_{sm}$ has a subsequence which converges to a policy $v \in \Pi_{sm}$.

For a stationary policy $v \in \Pi_{sm}$, define:

- P^v as the transition kernel where

$$P^v(x, A) \ := \ \mathbb{P}(X_{n+1} \in A \mid X_n = x),$$

  when the chain is controlled under the policy v.

- $\mathbb{P}_x^v$ as the probability measure on the canonical process space under control law $v \in \Pi_{sm}$, conditioned on $X_0 = x \in \mathbb{S}$, and

- $\mathbb{E}_x^v$ as the expectation operator on the same.

One can likewise define P^U, $\mathbb{P}_x^U$, and $\mathbb{E}_x^U$ for any particular control $U \in \mathfrak{U}$.

For a given policy $v \in \Pi_{sm}$ (or control $U \in \mathfrak{U}$), the MDP is simply a Markov chain on $\mathbb{S}$ with transition matrix P^v (P^U), and we will refer to it as such when the particular policy or control is explicit or clear from context. A Markov control (or policy) for which the distribution $(U_n|X_n)$ is a Dirac measure is called *precise*, and we denote the set of such controls (policies) as $\mathfrak{U}_{sd}$ (Π_{sd}).

2.4 Recurrence Properties and Exit Times

For any set $D \subset \mathbb{S}$, the *exit time* $\tau(D)$ is defined as

$$\tau(D) \;:=\; \min\{n \geq 0 : X_n \notin D\},$$

and the *first entry time* $\tau_e(D)$ as

$$\tau_e(D) \;:=\; \min\{n \geq 1 : X_n \in D\}.$$

We define the return time probability $L(x, A) := \mathbb{P}(\tau(A^c) < \infty)$, and say that a Markov chain is *ψ-irreducible* if there exists a measure ϕ on $\mathcal{B}(\mathbb{S})$ such that

$$\phi(A) > 0 \implies L(x, A) > 0 \text{ for all } x \in \mathbb{S}. \tag{2.1}$$

As detailed in [43], the name "ψ-irreducible" arises from the fact that if there exists a ϕ satisfying (2.1), there also exists a maximal (in the sense of largest support) probability measure ψ on $\mathcal{B}(\mathbb{S})$ also satisfying (2.1). Note that for

finite or countable state spaces, the traditional definition of irreducibility (via communicating classes) is naturally incorporated for any probability measure ψ supported on the whole space.

Without delving to deeply into the details, we will generally assume that under any admissible policy the induced chain is *aperiodic*. Intuitively, this means that the chain can return to any set of non-zero ψ-measure at irregular/acyclical times.

A set $A \in \mathcal{B}(\mathbb{S})$ is called *recurrent* if the expected number of times the chain revisits A is infinite; that is, for any $x \in A$,

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{I}_{X_n \in A} \;\middle|\; X_0 = x\right] = \infty.$$

The entire chain is called recurrent if it is ψ-irreducible and every set of non-zero ψ-measure is recurrent.

A chain with a transition kernel P is called *positive* if there exists an invariant probability measure $\pi \in \mathcal{P}(\mathbb{S})$; that is, $\pi(A) = \pi[P](A)$ for any $A \in \mathcal{B}(\mathbb{S})$, and a chain that is *positive recurrent* is equivalently called *stable*. Any recurrent chain has a unique (up to scalar multiples) invariant measure, but that measure may not be finite. To determine the existence of an invariant probability measure, we need the following definition. A set $A \in \mathcal{B}(\mathbb{S})$ is called *petite* if there exists a maximal irreducibility measure ψ such that

$$\sum_{n=0}^{\infty} \left(\tfrac{1}{2}\right)^{n+1} P^n(x, B) \geq \psi(B),$$

for all $x \in A$ and $B \in \mathcal{B}(\mathbb{S})$. Then, from [43], a ψ-irreducible chain is positive and recurrent if there exists there exists a petite set $C \in \mathcal{B}(X)$ with $\psi(C) > 0$ such that

$$\sup_{x \in C} \mathbb{E}[\tau_e(C) \mid X_0 = x] < \infty.$$

When considering an MDP with a policy $v \in \Pi$ that induces a stable Markov chain, we will frequently indicate the corresponding invariant probability measure as μ_v. Define $\mathfrak{U}_{ssm} \subset \mathfrak{U}_{sm}$ ($\Pi_{ssm} \subset \Pi_{sm}$) as the set of stationary Markov controls (policies) that induce a stable chain. We will refer to these controls and policies as stable.

For a function $h : \mathbb{S} \times \mathbb{U} \to \mathbb{R}$, we define the function $\bar{h} : \mathbb{S} \times \mathcal{P}(\mathbb{U}) \to \mathbb{R}$ by

$$\bar{h}(x, \mu) \ := \ \int_{\mathbb{U}} h(x, u) \mu(du), \quad \mu \in \mathcal{P}(\mathbb{U}).$$

Further, for a particular $v \in \mathfrak{U}_{sm}$, we treat v as a parameter and define

$$h_v(x) \ := \ \bar{h}(x, v(x)) = \int_{\mathbb{U}} h(x, u) v(du|x).$$

2.5 Minimal Cost Problems

We generally assume that the cost function r is bounded below, and without loss of generality that $r : \mathbb{S} \times \mathbb{U} \to \mathbb{R}_+$. Generality is maintained because, as will be clear in the coming sections, translating the cost function by a constant will not affect the choice of policy and will simply translate the overall cost as well.

2.5.1 Finite Horizon Control Problem

For a fixed time $N \in \mathbb{N}$, in addition to a running cost $r \in \mathbb{S} \times \mathbb{U} \to \mathbb{R}_+$, we can consider a terminal cost $r_N : \mathbb{S} \to \mathbb{R}_+$. Then for an admissible control $U \in \mathfrak{U}$, we define the finite-horizon cost as

$$J_N^U(x) \;:=\; \mathbb{E}_x^U \left[\sum_{t=0}^{N-1} r(X_t, U_t) + r_N(X_t) \right].$$

The *finite horizon control problem* is then to minimize J_N over all admissible controls:

$$J_N^* \;:=\; \inf_{U \in \mathfrak{U}} J_N^U.$$

2.5.2 Infinite Horizon Discounted Control Problem

As the finite time horizon is lengthened, the total cost may be unbounded. Hence, one approach to considering the cost over an infinite horizon is to introduce a discount factor $\alpha \in (0, 1)$. For a cost function $r \in \mathbb{S} \times \mathbb{U} \to \mathbb{R}_+$ and an admissible control $U \in \mathfrak{U}$, we define the α-discounted cost:

$$J_\alpha^U(x) \;:=\; \mathbb{E}_x^U \left[\sum_{t=0}^{\infty} \alpha^t r(X_t, U_t) \right].$$

As before, the *infinite horizon discounted control problem* is to minimize J_α over all admissible controls:

$$J_\alpha^* \;:=\; \inf_{U \in \mathfrak{U}} J_\alpha^U.$$

For brevity, we will sometimes refer to the infinite horizon discounted cost problem as simply the *discounted cost problem.*

2.5.3 Long Term Average Cost

For some situations, however, discounting future costs is not appropriate. In such cases, we consider the *long term average cost average cost*, also called the ergodic cost. With a running cost function $r : \mathbb{S} \times \mathbb{U} \to \mathbb{R}_+$ and an admissible control $U \in \mathfrak{U}$, the long term average cost is defined as:

$$J^U := \limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} \mathbb{E}[r(X_m, U_m)].$$

The *long term average cost control problem* is to minimize J over all admissible controls:

$$J^* := \inf_{U\in\mathfrak{U}} J^U.$$

Chapter 3

Countable State Space: Model, Assumptions, and General Results

3.1 Introduction

In this chapter and in Chapters 4–5, we consider Markov decision processes (MDPs) on a countable state space, sometimes referred to as a denumerable state space. Though the theory of MDPs developed first on finite spaces, several approachable problems intrinsically require an infinite set of states. For example, queuing problems may lead to fundamentally different results if the size of the queue is capped at any finite value. Infinite state spaces also allow the possibility of unstable behavior (e.g., $\mathbb{P}(X_n \in A) \to 0$ as $n \to \infty$ for any finite $A \in \mathbb{S}$) which is not possible with finitely many states.

In the following sections, we review some essential results about MDPs and Markov chains on countable state spaces and introduce some notation that will aid our later analysis. We identify or re-interpret some of the assumptions we will make on the MDP in the subsequent chapters, and describe how the structure of a countable-state MDP can be made to fundamentally mimic certain characteristics of $\mathbb{R}^n$. We also show some important results for countable state operators and chains which, though not very involved, are essential later

and are in some cases unique formulations.

3.2 Countable State Model and Notation

We will resist the temptation to explicitly replace $\mathbb{S}$ with $\mathbb{N}$ because the natural numbers hint at structure that may not be present. For example, consider an autonomous chain on $\mathbb{Z}$ with

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} p, & j = i - 1\,, \\ 1 - p, & j = i + 1\,. \end{cases}$$

The state 0 seems to hold a special place in $\mathbb{N}$, but in the example 0 is structurally indistinguishable from any other state. Further, to re-enumerate the states in the example to create an equivalent chain on $\mathbb{N}$ makes the transition probabilities awkward to define. Hence, we leave the enumeration of the states undefined until needed.

Even so, with a countable state space and an admissible control $u \in \mathbb{U}$, the transition kernel can be equivalently represented as an infinite stochastic matrix, or *transition probability matrix*: $P^u_{ij} = \mathbb{P}^u(X_{n+1} = j | X_n = i)$, for any states $i, j \in \mathbb{S}$. For a set $D \subset \mathbb{S}$ and a transition probability matrix P on $\mathbb{S}$, define a matrix ${}_DP$ by

$${}_DP_{ij} := \begin{cases} P_{ij} & \text{for } i, j \in D\,, \\ 0 & \text{otherwise.} \end{cases}$$

Note that if D is finite, ${}_DP$ is equivalent to a finite $(|D| \times |D|)$ matrix. We interchangeably use ${}_DP$ to refer to the infinite matrix with rows and columns of zeros defined above, and to the equivalent $|D| \times |D|$ matrix.

Further define a probability transition matrix $\overline{{}_D P}$ on $D \cup \{b\}$, by replacing D^c with a single absorbing state b:

$$\overline{{}_D P}_{ij} := \begin{cases} P_{ij} & \text{for } i, j \in D\,, \\ 0 & \text{for } i = b,\ j \in D\,, \\ \sum_{k \notin D} P_{ik} & \text{for } i \in D, j = b. \end{cases}$$

Clearly, a *truncated* MDP on $D \cup \{b\}$ with a transition kernel $\overline{{}_D P^u}$ will share several characteristics with an MDP on $\mathbb{S}$ with kernel P^u; notably, the exit time $\tau(D)$ will have the same distribution when both MDPs start in D.

The period of a state is the greatest common factor of possible return times. That is, for a state $i \in \mathbb{S}$, $\gcd\{n > 0 : \mathbb{P}(X_n = i | X_0 = i)\}$. A state is called aperiodic if it has period 1, and a Markov chain is said to be aperiodic if every state is aperiodic. We will assume throughout that:

Assumption 3.2.1. The MDP is an aperiodic Markov chain under any admissible $U \in \mathfrak{U}$.

Many of the results here can be adapted for chains with period N by replacing functions of the chain with the N-step average. Additionally, in many cases a periodic controlled Markov chain can be replaced with an approximate aperiodic chain that will lead to the equivalent conclusions and calculations [47, pp. 371]. In the current work, accounting for periodicity will unnecessarily complicate the analysis.

Finally, a matrix P (or, equivalently, a Markov chain governed by transition matrix P) is *irreducible* if for any $i, j \in \mathbb{S}$, there exists an $n \in \mathbb{N}$ such

that $P_{ij}^{(n)} > 0$. In other words, there is a non-zero probability of the chain reaching j from i in finitely many steps. Similarly, ${}_DP$ is called irreducible if for any $i, j \in D$, there exists an $n \in \mathbb{N}$ such that ${}_DP_{ij}^{(n)} > 0$.

Assumption 3.2.2. The MDP is an irreducible Markov chain under any admissible $U \in \mathfrak{U}$.

The assumption of irreducibility can also be relaxed under some circumstances (see, for example, [16, Section V4] and [24]), but again such assumptions complicate the analysis. We will rather assume irreducibility and leave extensions up to the reader.

3.3 Structural Assumptions

In order to apply concepts from continuous analysis, one requires a discrete space that behaves in some sense like a continuous space. In topological terms, a countable space with the discrete topology is intrinsically unlike a $\mathbb{R}^n$, whereas a discrete lattice with the taxicab metric *is* similar to $\mathbb{R}^n$ space in fundamental ways.

In the same way, evolution of a Markov chain on a discrete space may be entirely dissimilar to a continuous diffusion process without appropriate assumptions on the transition probabilities. The following assumptions define the structure of Markov chains that are sufficiently similar to allow the translation of some of the analyses of continuous processes. We point out examples of the type of processes that these assumptions exclude, but also note that

a very rich set of processes *are* allowed under all of the assumptions. These assumptions will be used at various places throughout the following analyses, and will be indicated either in the appropriate theorem or at the beginning of the chapter.

3.3.1 Finitely Many Transitions

The following assumption is used to restrict the chain dynamics to trajectories that behave in some sense like continuous trajectories.

Assumption 3.3.1. For any state $i \in \mathbb{S}$:

(i) the set $\{j \in \mathbb{S} : P_{ij}^u > 0 \text{ for some } i \in \mathbb{S}, u \in \mathbb{U}\}$ is finite, and

(ii) the set $\{i \in \mathbb{S} : P_{ij}^u > 0 \text{ for some } j \in \mathbb{S}, u \in \mathbb{U}\}$ is finite.

A restatement of Assumption 3.3.1 (i)/(ii) is that there are at most finitely many transitions into/out of any particular state. A key implication of Assumption 3.3.1 is that for any finite set $A \in \mathbb{S}$, there exists a finite set $B \supset A$ such that for any $v \in \mathfrak{U}_{sm}$, $\mathbb{P}^v(X_1 \in A | X_0 \in B^c) = \mathbb{P}^v(X_1 \in B^c | X_0 \in A) = 0$. In other words, the chain cannot reach A from B^c or B^c from A without an intermediate step in $A^c \cap B$; this approximates the behavior of a continuous process. This assumption is essentially a strengthening of a sufficient condition for stability under local perturbations as described in [16, VI, Lemma 1.1].

Some easily defined chains can violate Assumption 3.3.1, such as a chain governed by $P_{0i} = (1/2)^i$, $P_{ii-1} = 1$, where $0 \in \mathbb{S}$ is some particular state. A

chain of this or similar form can move "infinitely far" in a single step, and such behavior is problematic in some of the later analysis.

3.3.2 Filtration

For each $v \in \Pi_{sm}$, define

$$\mathcal{H}^v := \{G \subset \mathbb{S} : G \text{ finite, } {}_GP^v \text{ and } {}_{G^c}P^v \text{ irreducible}\} ,$$

$$\mathcal{H} := \bigcap_{v \in \Pi_{sm}} \mathcal{H}^v = \{G \subset \mathbb{S} : G \subset \mathcal{H}^v \text{ for all } v \in \Pi_{sm}\} .$$

Using this notation we can state another assumption which ensures behavior analogous to continuous processes:

Assumption 3.3.2. There is a filtration $\mathcal{G} = \{G_k\} \subset \mathcal{H}$, satisfying $G_0 \subset G_1 \subset G_2 \subset \cdots$, and $\bigcup_{n=0}^{\infty} G_k = \mathbb{S}$.

A random walk on $\mathbb{Z}$ violates Assumption 3.3.2. For example, consider the simple random walk $P_{i,i-1} = P_{i,i+1} = 0.5$. The only finite sets $K \in \mathbb{Z}$ such that ${}_KP$ is irreducible are sets of consecutive integers: $K_{i,N} = \{i, i+1, \ldots, i+N\}$. However, ${}_{K^c_{i,N}}P$ cannot be irreducible, as the chain must go through $K_{i,N}$ to get from $i-1$ to $i+N+1$. On the other hand, random walks on $\mathbb{Z}^n$ for $n > 1$ can satisfy Assumption 3.3.2, as in multiple dimensions the random walker can walk "around" any finite set.

Assumption 3.3.2 has the following immediate implication:

Corollary 3.3.3. *Assumption 3.3.2 is equivalent to the following: for any finite $D \subset \mathbb{S}$, there exists a $G \in \mathcal{H}$ such that $D \subset G$.*

When Assumption 3.3.2 is invoked, we will frequently be using the sets $G \in \mathcal{G}$ as "neighborhoods" to show various results about the system trajectory. Therefore, we will need the following:

Definition 3.3.4. Given a set $D \in \mathbb{S}$, we say that a state $i \in \mathbb{S}$ is $\mathcal{G}$-*separated* from D if there is a $G \in \mathcal{G}$ such that $D \subset G$ and $i \in G^c$.

3.3.3 Structural Results

For an infinite, non-negative matrix P (i.e., with non-negative entries), recall that if ${}_DP$ is irreducible for some $D \subset \mathbb{S}$, then for any $i, j \in D$, ${}_DP_{ij}^{(n)} > 0$ for some finite $n > 0$. We say $i \rightsquigarrow j$ in D if ${}_DP_{ij}^{(n)} > 0$ for some finite $n > 0$, and so D is irreducible if and only if $i \rightsquigarrow j$ for every $i, j \in D$. Equivalently, for $i, j \in D$, there is a finite chain $\{k_1, k_2, \ldots, k_{n-1}\} \subset D$ such that the product $P_{ik_1}P_{k_1k_2}\cdots P_{k_{n-1}j} > 0$. We say that this chain connects i to j in D. For a set $D_0 \subset D \subset \mathbb{S}$, $i \rightsquigarrow D_0$ in D indicates that $i \rightsquigarrow j$ in D for some $j \in D_0$. We can also, for any $i, j \in D$, select a (not necessarily unique) shortest chain of length N, where $N = \min\{n > 0 : {}_DP_{ij}^{(n)} > 0\}$.

Lemma 3.3.5. *Let Assumptions 3.3.1 and 3.3.2 hold. For any $G_k \in \mathcal{G}$, there is a $m > k$ such that ${}_{G_m}P^v \cap {}_{G_k^c}P^v$ is irreducible for all $v \in \Pi_{sm}$.*

Proof. Without loss of generality, let $k = 0$. Let G be the smallest G_m such that $P_{ij}^v = P_{ji}^v = 0$ for all $i \in G_0$, $j \in G^c$, and all $v \in \Pi_{sm}$ (i.e., it takes at least two steps for the chain to enter G^c starting in G_0, and visa versa). If ${}_{G\cap G_0^c}P^v$ is not irreducible for all $v \in \Pi_{sm}$, let $\widehat{G} \subset \mathbb{S}$ be the smallest set containing G

such that ${}_{\widehat{G}\cap G_0^c}P^v$ is irreducible for all $v \in \Pi_{sm}$, and let $\overline{G} \in \mathcal{G}$ be the smallest G_m containing $\widehat{G}$. (To construct such a $\widehat{G}$, for any $i \not\rightsquigarrow j$ in $G \cap G_0^c$, find the shortest chain connecting $i \rightsquigarrow j$ in G_0^c. Let $\widehat{G}$ be the union of G and all such chains.)

Suppose ${}_{\overline{G}\cap G_0^c}P$ is not irreducible for all $v \in \Pi_{sm}$. Then there must be states $i, j \in \overline{G} \cap G_0^c$ such that $i \not\rightsquigarrow j$ in $\overline{G} \cap G_0^c$. By virtue of the various irreducible matrices, either $i \in \overline{G} \cap \widetilde{G}^c$ and $i \not\rightsquigarrow \widetilde{G} \cap G_0^c$ in $\overline{G} \cap G_0^c$, or $j \in \overline{G} \cap \widetilde{G}^c$ and $\widetilde{G} \cap G_0^c \not\rightsquigarrow j$ in $\overline{G} \cap G_0^c$. But $i \rightsquigarrow j$ in $\overline{G}$, so there must be an $k \in \overline{G} \cap \widetilde{G}^c$ and a $\ell \in G_0$ such that either $P_{k\ell}^v > 0$ or $P_{\ell k}^v > 0$ for all $v \in \Pi_{sm}$. However, by construction both of these probabilities must be zero, so the supposition must be false. Therefore choosing $G_m = \overline{G}$ satisfies the claim. □

Corollary 3.3.6. *Let Assumptions 3.3.1 and 3.3.2 hold. Any filtration $\mathcal{G}$ defined as in Assumption 3.3.2 has a subfiltration $\{G_k\}$ such that ${}_{G_{k+1}\cap G_k^c}P$ is irreducible for each $k \in \mathbb{N}$.*

3.4 General Results for Countable Operators

Here we present several results relating to linear operators on a countable state space, framing the results in a manner similar to results in partial differential equations. Though these results are not particularly complex, this presentation is significant in supporting the understanding and development of results in other chapters. We derive a simple version of Harnack's inequality for our context, and versions of the Dirichlet problem and Dynkin's inequality particularly suited to the problems addressed later.

3.4.1 Harnack's Inequality

Other researchers have derived more complex Harnack inequalities for use in more complex discrete scenarios. For example, [17] derives a parabolic Harnack inequality for continuous-time Markov processes on a countable space. In [38], the authors derive and utilize a Harnack inequality for continuous-time controlled Markov processes in a framework otherwise quite similar to the one presented here. However, this general and simple presentation of Harnack's inequality is uniquely valuable in our MDP context, and does not require that the chain be irreducible or aperiodic.

A function $\varphi \in \mathcal{C}_+(\mathbb{S})$ is called $(P-I)$-*harmonic* on $D \subset \mathbb{S}$ if

$$(P-I)\varphi(i) = 0 \quad \forall i \in D.$$

Lemma 3.4.1 (Harnack's Inequality)**.** *Let $D \subset \mathbb{S}$ be finite, and let P be an infinite non-negative matrix on $\mathbb{S}$ such that ${}_DP$ is irreducible. Suppose $\varphi \in \mathcal{C}_+(\mathbb{S})$ is $(P-I)$-harmonic on D. Then the following hold:*

(i) *Either $\varphi > 0$ or $\varphi = 0$ on D; and*

(ii) *There is a constant $C_H > 1$ depending only on D and P, such that $\varphi(i) \leq C_H \varphi(j)$ for every $i, j \in D$.*

Proof. (i): Suppose $\varphi(i) = 0$ for some $i \in D$. $0 = \varphi(i) = P\varphi(i) = \sum_j P_{ij}\varphi(j)$. Then $P_{ij} > 0 \Rightarrow \varphi(j) = 0$. Iterating this argument, $\varphi(j) = 0$ for any $j \in D$ satisfying $i \rightsquigarrow j$, which by irreducibility is all of D. Hence $\varphi(i) = 0$ for any

$i \in D$ implies $\varphi = 0$ on D, and equivalently $\varphi(i) > 0$ for any $i \in D$ implies $\varphi > 0$ on D.

(ii): If $\varphi(i) = 0$ for $i \in D$ then the claim is trivially true for any $C_H > 1$, so the same constant C_H identified for $\varphi > 0$ will suffice. For $\varphi > 0$ on D, suppose ${}_DP$ is aperiodic and let $n = |D|$, the number of states in D. Define

$$p \;:=\; \min_{i,j \in D} {}_DP_{ij}^{(n)},$$

and note that $p > 0$. Then for any $i, j \in D$, we have

$$\begin{aligned} \varphi(j) = P^{(n)}\varphi(j) \geq {}_DP^{(n)}\varphi(j) \;&=\; \sum_{k \in D} {}_DP_{jk}^{(n)}\varphi(k) \\ &\geq\; {}_DP_{ji}^{(n)}\varphi(i) \geq p\varphi(i)\,. \end{aligned} \tag{3.1}$$

Note that (3.1) with $j = i$ implies that $p \leq 1$, and that if $p = 1$ then (3.1) will also hold for any $p \in (0, 1)$. Since i and j were chosen arbitrarily from D, $C_H = p^{-1}$ satisfies the requirement.

If ${}_DP$ is periodic with period d ($\leq n$), let $\widehat{{}_DP} := \frac{1}{n}\sum_{m=1}^{n} {}_DP^{(m)}$. It follows that $\widehat{{}_DP}$ is aperiodic and irreducible, so we can choose

$$p = \min_{i,j \in D} \widehat{{}_DP}_{ij}, \tag{3.2}$$

and again $0 < p < 1$. Then for any $i, j \in D$, we have

$$\begin{aligned} \varphi(j) \;&=\; \frac{1}{n}\sum_{m=1}^{n} P^{(m)}\varphi(j) \\ &\geq \widehat{{}_DP}\varphi(j) \end{aligned}$$

$$= \sum_{k \in D} \widehat{{}_D P}_{jk} \varphi(k) \geq \widehat{{}_D P}_{ji} \varphi(i) \geq p\varphi(i), \tag{3.3}$$

and just as before we can guarantee $p < 1$ and let $C_H = p^{-1}$. Indeed, the definition in (3.2) is sufficient even when ${}_D P$ is aperiodic. □

An identical result can be shown for functions that are $(P - I)$-superharmonic on a finite $D \in \mathbb{S}$; that is, a function $\varphi \in \mathcal{C}_+(\mathbb{S})$ such that

$$(P - I)\varphi(i) \leq 0 \quad \forall i \in D.$$

Corollary 3.4.2 (Harnack for superharmonic functions)**.** *Let $D \subset \mathbb{S}$ be finite, and let P be an infinite non-negative matrix on $\mathbb{S}$ such that ${}_D P$ is irreducible. Suppose $\varphi \in \mathcal{C}_+(\mathbb{S})$ satisfies $(P - I)\varphi \leq 0$ on D. Then there is a constant $C_H > 1$ depending only on D and P, such that $\varphi(i) \leq C_H \varphi(j)$ for every $i, j \in D$.*

Proof. With $P\varphi \leq \varphi$, the first equality in (3.1) and in (3.3) is replaced with an inequality, and the rest of the proof follows. □

Now, recalling that P^v is the probability transition matrix induced by policy $v \in \Pi_{sm}$, we can show a more general result:

Lemma 3.4.3 (Harnack for all controls)**.** *Let $D \subset \mathbb{S}$ be finite such that ${}_D P^v$ is irreducible for every $v \in \Pi_{sm}$. Suppose that for some $v \in \Pi_{sm}$, $\varphi \in \mathcal{C}_+(\mathbb{S})$ is $(P^v - I)$-superharmonic on D. Then there is a constant $C_H > 1$ depending only on D, such that $\varphi(i) \leq C_H \varphi(j)$ for every $i, j \in D$.*

Proof. Lemma 3.4.1 proves that for each v there is a constant $C_H^v > 1$ satisfying $\varphi^v(i) \leq C_H^v \varphi^v(j)$ for every $i, j \in D$. Elements of P^v depend continuously on v, so from (3.2) C_H^v also depends continuously on v. Therefore, since Π_{sm} is compact, $C_H = \sup_{v \in \Pi_{sm}} C_H^v$ exists and satisfies the requirement. □

It is worth noting that removing the dependence on the matrix P will not work for $(P-I)$-harmonic functions (or $(P-I)$-superharmonic functions) unless, as with P^v, the operators meeting the irreducibility requirement form (or are continuously indexed by elements of) a compact space. However, as the proof of Lemma 3.4.1 indicates, a global Harnack constant for superharmonic functions *can* be found if the class of operators has a uniform lower bound on the minimum averaged n-step probability defined in (3.2). More formally, for some $\delta > 0$ and finite $D \subset \mathbb{S}$, let

$$\mathfrak{h}(\delta, D) := \left\{ \text{matrices } P \geq 0 \text{ on } \mathbb{S} : {}_D P \text{ is irreducible, } \min_{i,j \in D} \widehat{{}_D P}_{ij} \geq \delta \right\}.$$

We call a set of infinite matrices belonging to $\mathfrak{h}(\delta, D)$ *uniformly irreducible* on D. The proof of the following lemma follows exactly as the others.

Lemma 3.4.4. *Let $D \subset \mathbb{S}$ be finite and $\delta > 0$. Suppose $\varphi \in C(\mathbb{S})$, $\varphi \geq 0$, and φ is $(P-I)$-superharmonic on D for some $P \in \mathfrak{h}(\delta, D)$, then there exists a constant $C_H = \delta^{-1} > 1$ depending only on D such that for any $i, j \in D$, $\varphi(i) \leq C_H \varphi(j)$.*

Uniformly irreducible matrices can also be identified directly: for an infinite non-negative matrix P, if ${}_D P$ is irreducible and the smallest non-zero entry of ${}_D P$ is greater than some $\gamma > 0$, then $P \in \mathfrak{h}(\gamma^{|D|}, D)$.

Remark 3.4.5. As an example of how this insight can be useful, consider the α-discounted cost J_α^v defined in Section 2.5.2. In Chapter 4, we show that for $v \in \Pi_{sm}$,

$$(\alpha P^v - I)J_\alpha^v = -c_v.$$

For a fixed $\alpha \in (0,1)$, we can use Lemma 3.4.3 to find a Harnack constant C_H^α. If we bound α below, say $\alpha \geq 1/2$, then the matrices (αP^v) are uniformly irreducible and have a Harnack constant $C_H^{1/2^+}$.

In this particular case, however, we can also find a Harnack constant for $\alpha \in (0, 1/2)$. Let $i, j \in D$, and let $\bar{D} = \{k \in \mathbb{S} : P_{ik}^v > 0 \text{ for some } i \in D\}$. Then

$$\begin{aligned} C_H^{1/2^-} &:= \frac{J_\alpha^v(i)}{J_\alpha^v(j)} = \frac{c_v(i) + \alpha P^v J_\alpha^v(i)}{c_v(j) + \alpha P^v J_\alpha^v(j)} \\ &\leq \frac{c_v(i) + \alpha \max_{k\in\bar{D}} J_\alpha^v(k)}{c_v(j)} \\ &\leq \frac{\max_{k\in D} c_v(k) + \frac{1}{2}\max_{k\in\bar{D}} J_{1/2}^v(k)}{\min_{k\in D} c_v(k)}. \end{aligned}$$

Hence, we can let $C_H = \max\{C_H^{1/2^-}, C_H^{1/2^+}\}$, so $J_\alpha^v(i) \leq C_H J_\alpha^v(j)$ for all $i, j \in D$ and C_H depends only on D (and not on P, v, or α).

3.4.2 A Dirichlet Problem

The following lemma is a discrete version of the Dirichlet problem for irreducible Markov chains.

Lemma 3.4.6 (Dirichlet)**.** *Let $D \subset \mathbb{S}$ be finite, and let P be an irreducible*

countable stochastic matrix. For any $h \in C(D)$, $g \in C(D^c)$,

$$(P - I)\varphi = -h \text{ on } D, \qquad \varphi = g \text{ in } D^c, \tag{3.4}$$

has a unique solution.

Proof. Any solution is clearly uniquely specified on D^c. Define the following:

$${}_D\varphi(i) = \begin{cases} \varphi(i) & \text{for } i \in D, \\ 0 & \text{for } i \in D^c, \end{cases} \qquad \overline{P}_{ij} = \begin{cases} P_{ij} & \text{for } i \in D,\ j \in D^c, \\ 0 & \text{otherwise.} \end{cases}$$

The problem $(P-I)\varphi = -h$ on D can be rewritten as $({}_DP - I)_D\varphi = -h - \overline{P}g$. Since P is irreducible, ${}_DP^m \to 0$ as $m \to \infty$. (This is equivalent to saying $\mathbb{P}_i(\tau(D) < \infty) = 1$ for all $i \in D$ for a Markov chain with transition matrix P.) Therefore (see, e.g., [50], lemma B1), $({}_DP - I)^{-1}$ exists, so

$$\varphi = \begin{cases} ({}_DP - I)^{-1}(-h - \overline{P}g) & \text{on } D\,, \\ g & \text{on } D^c, \end{cases}$$

is the unique solution of (3.4). □

Next, a lemma that the limit of a sequence of controls induces a limit of solutions of the Dirichlet problem described above.

Lemma 3.4.7. *Let* $v_n \to v^* \in \Pi_{sm}$, *and let* $D \subset \mathbb{S}$ *be finite. If* φ_n *solves*

$$(P^{v_n} - I)\varphi_n = -h \text{ on } D, \qquad \varphi_n = g \text{ in } D^c,$$

then $\varphi_n \to \varphi^*$ *where* φ^* *solves*

$$(P^{v^*} - I)\varphi^* = -h \text{ on } D, \qquad \varphi^* = g \text{ in } D^c.$$

Proof. $v_n \to v^*$ in Π_{sm}, so $P^{v_n} \to P^{v^*}$ element-wise, and there is a unique φ^* that solves

$$(P^{v^*} - I)\varphi^* = -h \text{ on } D, \qquad \varphi^* = g \text{ in } D^c.$$

Consider $\psi_n = \varphi_n - \varphi^*$. $\psi_n = 0$ on D^c, and on D we have

$$\begin{aligned}(P^{v_n} - I)\psi_n &= (P^{v_n} - I)\varphi_n - (P^{v_n} - I)\varphi^* \\ &= -h - (P^{v_n} - I)\varphi^* \\ \xrightarrow[n\to\infty]{} &- h - (P^{v^*} - I)\varphi^* = -h - (-h) = 0.\end{aligned}$$

Since $\psi_n = 0$ on D^c, $({}_DP^{v_n} - I)\psi_n \to 0$ everywhere. Therefore, either $\psi_n \to 0$ everywhere or at least one eigenvalue of $({}_DP^{v_n} - I)$ approaches 0. But $({}_DP^{v_n} - I) \to ({}_DP^{v^*} - I)$, which has nonzero eigenvalues. Since eigenvalues depend continuously on the matrix elements, we can find N large enough that the eigenvalues of $({}_DP^{v_n} - I)$ are bounded away from zero for $n > N$. Then because no eigenvalues of $({}_DP^{v_n} - I)$ approach zero, $\psi_n \to 0$ on D, and therefore $\varphi_n \to \varphi^*$. □

3.4.3 Dynkin's Formula

We first state Dynkin's formula as traditionally presented:

Theorem 3.4.8 (Dynkin's formula, Theorem 11.3.1 [43])**.** *Let f be a real-valued on $\mathbb{S}$, let τ be a stopping time. Define another stopping time*

$$\tau^n := \min\left\{n, \tau, \min\{k \geq 0 : f(X_k) \geq n\}\right\}.$$

For each $i \in \mathbb{S}$ and $n \in \mathbb{Z}_+$,

$$\mathbb{E}_i[f(X_{\tau^n})] = f(i) + \mathbb{E}_i\left[\sum_{m=1}^{\tau^n} \mathbb{E}[f(X_m)|\mathcal{F}_{m-1}] - f(X_{m-1})\right].$$

In the current context, we will frequently wish to apply Dynkin's formula to a function not only of $\mathbb{S}$ but also of time:

$$f : \mathbb{S} \times \mathbb{N} \to \mathbb{R}; \qquad f(i,n) = f_n(i).$$

To accomplish this, we define an augmented Markov chain $\mathbf{Y}$ which takes values on $(\mathbb{S} \times \mathbb{N})$ as follows:

- $Y_n = (X_n, T_n)$, where T_n takes values on $\mathbb{N}$;
- $\mathbb{P}\big(Y_{n+1} = (j,m) \mid Y_n = (i,n'), U_n = u\big) = P^u(i,j)\mathbb{I}_{m=n'+1}$;
- $\mathbf{Y}$ is initialized with $Y_0 = (X_0, 0)$. Combined with the transition rule, $Y_n = (X_n, n)$ almost surely.

Now we can slightly abuse notation to say $f(Y_n) = f(X_n, n) = f_n(X_n)$, and use the following corollary.

Corollary 3.4.9. *Let f be a positive function on $\mathbb{S} \times \mathbb{N}$, let τ be a stopping time. Define another stopping time*

$$\tau^n \;:=\; \min\left\{n, \tau, \min\{m \geq 0 : f_m(X_m) \geq n\}\right\}.$$

For each $i \in \mathbb{S}$ and $n \in \mathbb{N}$,

$$\mathbb{E}_i[f_{\tau^n}(X_{\tau^n})] \;=\; f_0(i) + \mathbb{E}_i\left[\sum_{m=0}^{\tau^n - 1} \mathbb{E}[f_{m+1}(X_{m+1}) \mid \mathcal{F}_m] - f_m(X_m)\right].$$

Note that in the definition of τ^n, the third component,

$$\min\{m \geq 0 : f_m(X_m) \geq n\},$$

is included to ensure that $\sum_{m=0}^{\tau^n-1} f_m(X_m)$ is essentially bounded (by n^2). However, if $f_m(X_m)$ is almost surely bounded for $m < \tau$, then the third component is unnecessary. For example, if $\tau = \tau(D)$, the exit time from some finite set, and f_m is uniformly bounded on D for $m < \tau \wedge n$, then for $m < \tau$ we also have $f_m(X_m) \leq \max_D f_m$ and $\sum_{m=0}^{\tau^n-1} f_m(X_m)$ is essentially bounded by $n(\max_D f_m)$. In such a case, we can simply drop the third component and define

$$\tau^n \;:=\; \min\{n, \tau\}.$$

Finally, the following formulation of Dynkin's formula will prove repeatedly useful, as it matches the structure used in Theorem 3.4.6 and the definition of $(P - I)$-harmonic functions.

Lemma 3.4.10. *Suppose $D \subset \mathbb{S}$ is finite, $\{X_n\}$ a Markov chain on $\mathbb{S}$ governed by an irreducible transition probability matrix P, and $h \in C_+(\mathbb{S})$. Then*

$$\varphi(i) \;=\; \mathbb{E}_i\left[\sum_{m=0}^{\tau(D)-1} h(X_m)\right]$$

is a solution of

$$(P - I)\varphi \;=\; -h \text{ on } D, \qquad \varphi = 0 \text{ in } D^c\,.$$

Proof. Let $g \in C_+(\mathbb{S})$ be bounded, and to simplify notation let $\tau = \tau(D)$. For any $T > 0$, Define $\tau^n = \min\{n, \tau\}$; hence, from Theorem 3.4.8 we get

$$\mathbb{E}_i[g(X_{\tau^n})] \;=\; g(i) + \mathbb{E}_i\left[\sum_{m=1}^{\tau^n} \mathbb{E}[g(X_m) \mid \mathcal{F}_{m-1}] - g(X_{m-1})\right],$$

for each $i \in \mathbb{S}$ and $n > \max_{i\in\mathbb{S}} g(i)$. Also let $\{\widehat{X}_n\}$ be another Markov chain on $\mathbb{S}$ governed by P.

By irreducibility, $\mathbb{P}(\tau < \infty) = 1$, and so letting $n \to \infty$ we get

$$\begin{aligned}
\mathbb{E}_i\left[g(X_\tau)\right] - g(i) &= \mathbb{E}_i\left[\sum_{m=1}^{\tau} \mathbb{E}\left[g(X_m) \;\middle|\; X_{m-1}\right]\right] - \mathbb{E}_i\left[\sum_{m=1}^{\tau} g(X_{m-1})\right] \\
&= \mathbb{E}_i\left[\sum_{m=0}^{\tau-1} Pg(X_m)\right] - \mathbb{E}_i\left[\sum_{m=0}^{\tau-1} g(X_m)\right] \\
&= \mathbb{E}_i\left[\mathbb{E}_{X_1}\left[\sum_{m=0}^{\tau-1} g(\widehat{X}_m)\right]\right] - \mathbb{E}_i\left[\sum_{m=0}^{\tau-1} g(X_m)\right] \\
&= (P-I)\left(\mathbb{E}_i\left[\sum_{m=0}^{\tau-1} g(X_m)\right]\right). \qquad (3.5)
\end{aligned}$$

Now let $g = \mathbb{I}_D h$. Because $g(X_\tau) = 0$ by construction and $\tau = 0$ for $X_0 \in D^c$,

$$\mathbb{E}_i\left[\sum_{m=0}^{\tau-1} g(X_m)\right] = \mathbb{E}_i\left[\sum_{m=0}^{\tau-1} h(X_m)\right] = \varphi(i)\,,$$

and so 3.5 becomes

$$(P-I)\varphi = -h \text{ on } D, \qquad \varphi = 0 \text{ in } D^c\,.$$

Uniqueness follows from Theorem 3.4.6. □

Remark 3.4.11. Other applications of (3.5) are also frequently useful. For example, consider nested finite sets $D \subset B \subset S$, and suppose φ solves $(P - I)\varphi = 0$ on $B \cap D^c$, $\varphi = 1$ on D, $\varphi = 0$ on B^c. Then with $\tau = \tau(B \cap D^c)$, $g = \varphi$ makes the right side of (3.5) zero, and so $\varphi(i) = \mathbb{P}_i\left(\tau(D^c) < \tau(B)\right)$. We will refer to Lemma 3.4.10 for all such implications.

Chapter 4

Countable State Space: Uniform Recurrence Properties

4.1 Introduction

Two results are presented in this chapter, both adapted from the field of continuous diffusion processes. First, we recall [15], in which Borkar showed a series of equivalent properties for Markov decision processes (MDPs) when $\Pi_{sm} = \Pi_{ssm}$; that is, when all stationary Markov policies induce stable chains. The entire theorem is too detailed to reproduce here in its entirety, but the strength of the result is indicated by the following three equivalent properties [15, Theorem 8.1]:

(v) *The set* $\{f_v(i, du) = \mu_v(i)v(i, du) : v \in \Pi_{sm}\}$ *of ergodic occupation measures is tight, where* μ_v *is the stationary distribution under* v*.*

(vii) *Let* $0 \in \mathbb{S}$ *be a designated zero state. There exists an unbounded* $h \to \mathbb{R}_+$ *such that*

$$\sup_{v \in \Pi_{sm}} \mathbb{E}_i^v \left[\sum_{n=1}^{\tau(\{0\}^c)} h(X_n) \right] < \infty \,.$$

(viii) *There exists a* $V : \mathbb{S} \to \mathbb{R}_+$*, a constant* $b > 0$*, a finite* $C \subset \mathbb{S}$*, and a*

function h as in (vii) above, such that for any $v \in \Pi_{sm}$,

$$\mathbb{E}[V(X_{n+1})|\mathcal{F}_n] \leq V(X_n) - h(X_n) + b\mathbb{I}_{X_n \in C}\,.$$

Echoing the corresponding results for controlled diffusion processes, begun in [15] and greatly extended in [3], we derive a more general condition that is equivalent to [15, Theorem 8.1, (vii)], above. Our result is a property called *uniform recurrence*, and says that a bound as in [15, Theorem 8.1, (vii)] holds for any particular finite set and all policies, then it holds for any finite set and for the supremum over policies. To adapt the result for the countable state space and discrete time, however, we require some of the assumptions formulated in Chapter 3 chosen to make the countable-state MDP behave like a continuous process in specific ways.

Next, we show a result involving the discounted cost J_α from Section 2.5.2. Under the same structural assumptions used in the first result, we show that under any stationary Markov policy, the set of functions $\{J_\alpha : \alpha \in (0,1)\}$ has bounded variation on finite sets; this result approximates equicontinuity. We also show that $(1-\alpha)J_\alpha$ is uniformly bounded on finite sets. Uniform bounds on particular forms of the discounted cost can facilitate analysis of the average cost, as in [2, 52], for example.

4.2 Main Results

4.2.1 Uniform Recurrence

The main result extending [15, Theorem 8.1] is the following:

Theorem 4.2.1. *Let Assumptions 3.3.1 and 3.3.2 hold, and assume* $\Pi_{sm} = \Pi_{ssm}$. *If for some* $h \in C_+(\mathbb{S} \times \mathbb{U})$, *some finite* $D \subset \mathbb{S}$, *and some* $i_0 \in \mathbb{S}$ *that is* $\mathcal{G}$*-separated from* D, *we have*

$$\mathbb{E}_{i_0}^v \left[\sum_{n=0}^{\tau(D^c)-1} h_v(X_n) \right] < \infty \qquad \forall v \in \Pi_{ssm} \,.$$

Then for any finite $B \subset \mathbb{S}$, $i \in B^c$,

$$\sup_{v \in \Pi_{ssm}} \mathbb{E}_i^v \left[\sum_{n=0}^{\tau(B^c)-1} h_v(X_n) \right] < \infty \,.$$

4.2.2 Regularity of Discounted Value Functions

We also show that under the same structural assumptions, uniform bounds can be placed on the discounted cost function J_α. Note that this result does not require $\Pi_{sm} = \Pi_{ssm}$. Recall that for a set $D \in \mathbb{S}$ and a function $f : \mathbb{S} \to \mathbb{R}$,

$$\operatorname*{osc}_G f := \max_{i \in D} f(i) - \min_{j \in D} f(j) \,.$$

Theorem 4.2.2. *Let Assumptions 3.3.1 and 3.3.2 hold, and let* $G \in \mathcal{G}$. *There exists a constant* C_0 *depending only on* G *such that for all* $v \in \Pi_{ssm}$ *and* $\alpha \in (0, 1)$,

$$\operatorname*{osc}_G J_\alpha^v \;\le\; C_0 \frac{\varrho_v}{\mu_v(G)} \left(1 + \frac{1}{\mu_v(G)} \right) \,, \tag{4.1}$$

$$\sup_G (1-\alpha) J_\alpha^v \;\le\; C_1 \frac{\varrho_v}{\mu_v(G)} \,. \tag{4.2}$$

4.3 Supporting Lemmas

Lemma 4.3.1. *Let $D \subset \mathbb{S}$ be a finite set. Then*

$$\sup_{v \in \Pi_{sm}} \max_{i \in D} \mathbb{E}_i^v [\tau(D)] < \infty \,.$$

Proof. Since $D \subset G \Rightarrow \tau(D) \leq \tau(G)$, it suffices to show that for any $G \in \mathcal{G}$, $i \in G$, that $\sup_{v \in \Pi_{sm}} \mathbb{E}_i^v [\tau(G)] < \infty$. Let $G \in \mathcal{G}$. For any fixed $i \in G$, $v \in \Pi_{sm}$, $\mathbb{E}_i^v [\tau(G)] < \infty$ by irreducibility of P^v. (See [50, Appendix B]). Suppose claim is false. Then there exists an $i \in G$, $\{v_m\} \subset \Pi_{sm}$ such that $\mathbb{E}_i^{v_m} [\tau(G)] \to \infty$ as $m \to \infty$. Since Π_{sm} is compact, $v_m \to v^* \in \Pi_{sm}$. For any $v \in \Pi_{sm}$, let φ^v be the unique solution of

$$(P^v - I)\varphi^v = -1 \text{ on } G \,, \qquad \varphi^v = 0 \text{ on } G^c.$$

From Lemma 3.4.10, $\varphi^v(i) = \mathbb{E}_i^v [\tau(G)]$, and from Lemma 3.4.7, $\varphi^{v_n} \to \varphi^{v^*}$ which is bounded on G, contradicting the supposition. Hence, claim is true. □

Lemma 4.3.2. *For any finite sets $D \subset \mathbb{S}$ and $\Gamma \subset D^c$, we have*

$$0 < 1 \leq \inf_{v \in \Pi_{sm}} \min_{i \in \Gamma} \mathbb{E}_i^v [\tau(D^c)] \,,$$

$$\max_{i \in \Gamma} \mathbb{E}_i^v [\tau(D^c)] < \infty \qquad \forall v \in \Pi_{ssm} \,.$$

Proof. First is trivial, second is precisely stability. □

Lemma 4.3.3. *Let $D \subset \mathbb{S}$ be finite, $D \subseteq G \in \mathcal{G}$. Then*

$$\inf_{v \in \Pi_{sm}} \min_{i \in G} \mathbb{P}_i^v \big(\tau(D^c) < \tau(G)\big) > 0 \,.$$

Proof. For any particular $i \in G$, $v \in \Pi_{sm}$, $\mathbb{P}_i^v\big(\tau(D^c) < \tau(G)\big) > 0$ by irreducibility of ${}_G P^v$.

Suppose false. Then there exists $i \in G$ and $\{v_m\} \in \Pi_{sm}$ such that

$$\mathbb{P}_i^{v_m}\big(\tau(D^c) < \tau(G)\big) \xrightarrow[m\to\infty]{} 0\,.$$

Dropping to a subsequence if needed, $v_m \to v^* \in \Pi_{sm}$. For each $v \in \Pi_{sm}$, let φ^v be the unique solution of

$$(P^v - I)\varphi^v = 0 \text{ on } G \cap D^c\,, \qquad \varphi^v = 0 \text{ on } G^c\,, \qquad \varphi^v = 1 \text{ on } D\,.$$

From Lemma 3.4.10, $\varphi^v(i) = \mathbb{P}_i^{v_m}\big(\tau(D^c) < \tau(G)\big)$, and from Lemma 3.4.7 we have $\varphi_m \to \varphi^*$ which is non-zero on $G \cap D$, contradicting the supposition. □

Lemma 4.3.4. *Let $D \subset \mathbb{S}$ be finite, $h \in C_+(\mathbb{S})$, and $v \in \Pi_{sm}$ such that*

$$\mathbb{E}_i^v\left[\sum_{n=0}^{\tau(D^c)-1} h(X_n)\right] < \infty \qquad \forall x \in D^c\,.$$

Then for any $B \subset \mathbb{S}$,

$$\mathbb{E}_i^v\left[\sum_{n=0}^{\tau(B^c)-1} h(X_n)\right] < \infty \qquad \forall x \in B^c\,.$$

Proof. For a stopping time τ, define

$$\beta_i^v[\tau] := \mathbb{E}_i^v\left[\sum_{n=0}^{\tau-1} h_v(X_n)\right]\,.$$

It suffices to prove the claim for finite B. Further, for any finite B and D, we can choose $G \in \mathcal{G}$ such that $B \cup D \subset G$; then $\beta_i^v[\tau(G^c)] \le \beta_i^v[\tau(D^c)] < \infty$,

and $B \subset G \in \mathcal{G}$. So let $B \subset G \in \mathcal{G}$ and assume $\beta_i^v[\tau(G^c)] < \infty$ for all $i \in G^c$. Choose $G_1 \in \mathcal{G}$ such that $G \subset G_1$, and let $h_1 = \max_{i \in G_1} h(i) < \infty$. Define the stopping times $\widehat{\tau}_0 := \min\{n \geq 0 : X_n \in G\}$ and, for $k \geq 0$,

$$\begin{aligned} \widehat{\tau}_{2k+1} &:= \min\{n > \widehat{\tau}_{2k} : X_n \in G_1^c\}\,, \\ \widehat{\tau}_{2k+2} &:= \min\{n > \widehat{\tau}_{2k+1} : X_n \in G\}\,. \end{aligned}$$

Clearly, $\widehat{\tau}_k - \widehat{\tau}_{k-1} \geq 1$, $\beta_i^v[\widehat{\tau}_0] = \beta_i^v[\tau(G^c)] < \infty$. For $\beta_i^v[\widehat{\tau}_{2k}] < \infty$,

$$\begin{aligned} \beta_i^v[\widehat{\tau}_{2k+1}] &\leq \beta_i^v[\widehat{\tau}_{2k}] + \max_{j \in G} \beta_j^v[\tau(G_1)] \\ &\leq \beta_i^v[\widehat{\tau}_{2k}] + h_1 \max_{j \in G} \mathbb{E}_j^v\left[\tau(G_1)\right] < \infty \end{aligned}$$

by Lemma 4.3.1, and with $\partial G_1 := \{i \in G_1^c : P_{ji}^v > 0 \text{ for some } j \in G_1\}$

$$\beta_i^v[\widehat{\tau}_{2k+2}] \leq \beta_i^v[\widehat{\tau}_{2k+1}] + \max_{j \in \partial G_1} \beta_j^v[\tau(G^c)] < \infty$$

by assumption. So each $\beta_j^v[\widehat{\tau}_{2k}] < \infty$, and $\widehat{\tau}_k \uparrow \infty$.

Let $\varphi(i) = \mathbb{P}_i^v\left(\tau(G_1) < \tau(B^c)\right)$, which is the unique solution of

$$(P^v - I)\varphi = 0 \text{ on } G_1 \cap B^c\,, \qquad \varphi = 0 \text{ on } B\,, \qquad \varphi = 1 \text{ on } G_1^c\,.$$

Define

$$\begin{aligned} p_0 &:= \max_{i \in G_1 \cap B^c} \varphi(i) = \max_{i \in G_1 \cap B^c} \mathbb{P}_i^v\big(\tau(G_1) < \tau(B^c)\big) \\ &= 1 - \min_{i \in G_1 \cap B^c} \mathbb{P}_i^v\big(\tau(G_1) > \tau(B^c)\big) < 1\,, \end{aligned}$$

where the last step uses Lemma 4.3.3. By the strong Markov property,

$$\mathbb{P}_i^v\big(\tau(B^c) > \widehat{\tau}_{2k}\big) \leq p_0 \mathbb{P}_i^v\big(\tau(B^c) > \widehat{\tau}_{2k-2}\big) \leq \cdots \leq p_0^k\,.$$

So for any $i \in G_1 \cap B^c$, we have

$$\begin{aligned}
\beta_i^v[\tau(B^c)] &\leq \sum_{k=1}^{\infty} \mathbb{E}_i^v \left[\mathbb{I}_{\widehat{\tau}_{2k-2} < \tau(B^c) \leq \widehat{\tau}_{2k}} \sum_{n=0}^{\widehat{\tau}_{2k}-1} h(X_n) \right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}_i^v \left[\mathbb{I}_{\widehat{\tau}_{2k-2} < \tau(B^c) \leq \widehat{\tau}_{2k}} \left(\sum_{n=0}^{\widehat{\tau}_0-1} h(X_n) + \sum_{\ell=1}^{k} \sum_{n=\widehat{\tau}_{2\ell-2}}^{\widehat{\tau}_{2\ell}-1} h(X_n) \right) \right] \\
&= \beta_i^v[\widehat{\tau}_0] + \sum_{k=1}^{\infty} \sum_{\ell=1}^{k} \mathbb{E}_i^v \left[\mathbb{I}_{\widehat{\tau}_{2k-2} < \tau(B^c) \leq \widehat{\tau}_{2k}} \sum_{n=\widehat{\tau}_{2\ell-2}}^{\widehat{\tau}_{2\ell}-1} h(X_n) \right] \\
&= \beta_i^v[\widehat{\tau}_0] + \sum_{\ell=1}^{\infty} \mathbb{E}_i^v \left[\mathbb{I}_{\widehat{\tau}_{2k-2} < \tau(B^c)} \sum_{n=\widehat{\tau}_{2\ell-2}}^{\widehat{\tau}_{2\ell}-1} h(X_n) \right] \\
&\leq \beta_i^v[\widehat{\tau}_0] + \sum_{\ell=1}^{\infty} p_0^{\ell-1} \max_{j \in G} \mathbb{E}_j^v \left[\sum_{n=0}^{\widehat{\tau}_2-1} h(X_n) \right] \\
&\leq \beta_i^v[\widehat{\tau}_0] + \frac{1}{1-p_0} \max_{j \in G} \beta_j^v[\widehat{\tau}_2] < \infty \,.
\end{aligned}$$

Note that $G_1 \in \mathcal{G}$ can be chosen arbitrarily large, so $\beta_i^v[\tau(B^c)] < \infty$ for any $i \in B^c$. □

Note the useful special case of the previous lemma for $h = 1$, in which case the summations are replaced by the exit times themselves.

Lemma 4.3.5. *Let $D \subset \mathbb{S}$ be finite, $h \in C_+(\mathbb{S})$, and $\{v_k\} \subset \Pi_{sm}$ a sequence of policies such that*

$$\lim_{k \to \infty} \mathbb{E}_i^{v_k} \left[\sum_{n=0}^{\tau(D^c)-1} h(X_n) \right] < \infty \qquad \forall x \in D^c.$$

Then for any $B \subset \mathbb{S}$,

$$\lim_{k\to\infty} \mathbb{E}_i^{v_k}\left[\sum_{n=0}^{\tau(B^c)-1} h(X_n)\right] < \infty \qquad \forall i \in B^c.$$

Proof. As in Lemma 4.3.4, it suffices to prove the assertion for $D = G \in \mathcal{G}$ and $B \subset G$. Following a familiar argument structure, define

$$\begin{aligned} p_v &:= \max_{i\in G} \mathbb{P}_i^v\left(\tau(G) < \tau(B^c)\right), \\ \partial G &:= \left\{i \in G^c : P_{ji}^v > 0 \text{ for some } j \in G \text{ and any } v \in \Pi_{sm}\right\}. \end{aligned}$$

Then for any k we have

$$\begin{aligned} \beta_{i_0}^{v_k}[\tau(B^c)] &\le \beta_{i_0}^{v_k}[\tau(G^c)] + \max_{i\in G} \beta_i^{v_k}[\tau(G\cup B^c)] \\ &\qquad + \sum_{\ell=1}^{\infty} p_{v_k}^{\ell}\left(\max_{i\in\partial G} \beta_i^{v_k}[\tau(G^c)] + \max_{i\in G} \beta_x^{v_k}[\tau(G\cup B^c)]\right) \\ &\le \sum_{\ell=0}^{\infty} p_{v_k}^{\ell}\left(\max_{i\in\{i_0\}\cup\partial G} \beta_i^{v_k}[\tau(G^c)] + \max_{i\in G} \beta_i^{v_k}[\tau(G\cup B^c)]\right) \\ &= \frac{1}{1-p_{v_k}}\left(\max_{i\in\{i_0\}\cup\partial G} \beta_i^{v_k}[\tau(G^c)] + \max_{i\in G} \beta_i^{v_k}[\tau(G\cup B^c)]\right). \end{aligned}$$

From Lemma 4.3.3, p_{v_k} is bounded away from 1 uniformly in k, so taking limits as $k \to \infty$ on both sides of the inequality proves the result. □

Lemma 4.3.6. *Let $D \subset \mathbb{S}$ be finite, $v \in \Pi_{sm}$, and $h \in C_+(\mathbb{S})$, and suppose $f(i) := \mathbb{E}_i^v\left[\sum_{n=0}^{\tau(D^c)-1} h(X_n)\right]$ is finite at some $i_0 \in \mathbb{S}$ that is $\mathcal{G}$-separated from D (i.e., there exists $G \in \mathcal{G}$ such that $D \subset G$ and $i_0 \in G^c$). Then $f(i)$ is finite for all $i \in D^c$, and f is the minimal non-negative solution of*

$$(P^v - I)f = -h \text{ on } D^c, \qquad f = 0 \text{ on } D.$$

Proof. First, we will show that $\varphi(i) := \mathbb{E}_i^v \left[\sum_{n=0}^{\tau(G^c)-1} h(X_n) \right]$ is finite for all $i \in G^c$, then that $f(i)$ is finite for all $i \in D^c$, and finally that f is minimal.

Clearly, $\varphi(i_0) \leq f(i_0) < \infty$. Let $j \in G^c$, and choose $\{G_k\}_{k=0}^{\infty} \subset \mathcal{G}$ such that $\bigcup_k G_k = \mathbb{S}$, $G \cup \{i_0, j\} \subset G_0 \subset G_1 \subset \cdots$, and each ${}_{G_k \cup G^c}P$ is irreducible.

For $m = 0, 1, \cdots$, let φ_m solve

$$(P^v - I)\varphi_m = -h \text{ on } G_m \cap G^c\,, \qquad \varphi_m = 0 \text{ on } G \cup G_m^c\,.$$

From Lemma 3.4.10,

$$\varphi_m(i) = \mathbb{E}_i^v \left[\sum_{n=0}^{\tau(G^c)\wedge\tau(G_m)-1} h(X_n) \right].$$

Clearly, $0 \leq \varphi_m \leq \varphi_{m+1}$, and $\varphi_m(i_0) \leq \varphi(i_0) < \infty$.

For each m, let $\psi_{m+1} = \varphi_{m+1} - \varphi_m$. By construction, $(P^v - I)\psi_m = 0$ on $G_0 \cap G^c$. Let $\bar{\psi}_m = \sum_{k=1}^{m} \psi_m = \varphi_m - \varphi_0$. Then $(P^v - I)\bar{\psi}_m = 0$ on $G_0 \cap G^c$ and $\bar{\psi}_m(i_0) \leq \varphi(x_0) < \infty$, so (using Lemma 3.4.1) $\bar{\psi}_m(j) \leq C_H \bar{\psi}_m(i_0) < \infty$. Then $\bar{\psi}_m(j) \uparrow \bar{\psi}(j)$, and since j was arbitrarily chosen in G^c, $\bar{\psi}_m \uparrow \bar{\psi} \in C(\mathbb{S})$ uniformly on finite subsets of G^c. Let $\varphi = \bar{\psi} + \varphi_0$, which satisfies the original definition and is finite for every $i \in G^c$. $f(i)$ must therefore be finite at every $i \in D^c$ by direct application of Lemma 4.3.4.

To show that f is minimal, we now define

$$f_m(i) \;:=\; \mathbb{E}_i^v \left[\sum_{n=0}^{\tau(D^c)\wedge\tau(G_m)-1} h(X_n) \right] \leq f(i) < \infty\,.$$

For any $\bar{f} \in C(\mathbb{S})$ that solves $(P^v - I)\bar{f} = -h$ on D^c, $\bar{f} = 0$ on D, $\bar{f} \geq 0$, note that $(P^v - I)(\bar{f} - f_m) = 0$ on $G_m \cap G^c$ and $(\bar{f} - f_m) = \bar{f} \geq 0$ on $G_m^c \cup G$.

P^v is a positive operator and irreducibility guarantees that $P^v_{ij} > 0$ for some $i \in G^c_m \cup G$ and some $j \in G^c_m \cup G$, so non-negativity of $(\bar{f} - f_m)$ will percolate throughout $G_m \cap G^c$ for every m. Since f is clearly the pointwise limit of f_m, $(\bar{f} - f) \geq 0$ everywhere, and so f is minimal. □

Lemma 4.3.7. *Let $G \in \mathcal{G}$, $\widehat{\tau}_0 = 0$, and inductively for $k = 0, 1, \ldots$*

$$\begin{aligned} \widehat{\tau}_{2k+1} &= \min\{n > \widehat{\tau}_{2k} : X_n \in G^c\}\,, \\ \widehat{\tau}_{2k+2} &= \min\{n > \widehat{\tau}_{2k+1} : X_n \in G\}\,. \end{aligned} \tag{4.3}$$

Clearly, for any $k \geq 0$, $\widehat{\tau}_{k+1} - \widehat{\tau}_k \geq 1$, and $\mathbb{P}^v\left(\widehat{\tau}_{k+1} - \widehat{\tau}_k < \infty\right) = 1$ for every $v \in \Pi_{sm}$.

Define $\widetilde{X}_n = X_{\widehat{\tau}_{2n}}$, $n \geq 1$. $\widetilde{X}_n$ is an ergodic Markov chain on G (though not necessarily on all of G). Under $v \in \Pi_{ssm}$, there exists $\delta \in (0,1)$ (which does not depend on v) such that if we define $\widetilde{P}^v(\cdot,\cdot)$ and $\widetilde{\mu}_v$ to be the transition kernel and invariant distribution of $\widetilde{X}_n$, then for all $i \in G$

$$\begin{aligned} \|\widetilde{P}^{v\,(n)}(i,\cdot) - \widetilde{\mu}_v(\cdot)\|_{TV} &\leq \delta^n \qquad \forall n \in \mathbb{N}\,, \\ \delta \widetilde{P}^v(i,\cdot) &\leq \widetilde{\mu}_v(\cdot)\,. \end{aligned} \tag{4.4}$$

Proof. Let $v \in \Pi_{ssm}$ and note that for any $i \in G$ such that $P^v_{ji} = 0$ for all $j \in G^c$ (i.e., the "interior" of G) we have $\widetilde{P}^v(\cdot, i) = 0$, so we can proceed only considering those states $i \in G$ that have non-zero probability $P^v_{ji} > 0$ for some $j \in G^c$ (i.e., the "incoming boundary" of G):

$$\partial G := \{i \in G : P^v_{ji} > 0 \text{ for some } j \in G^c\}\,.$$

Because ${}_{G^c}P^v$ is irreducible, $\widetilde{P}^v(i,j) > 0$ for all $i, j \in \partial G$; hence $\widetilde{X}_n$ is ergodic and has stationary distribution $\widetilde{\mu}_v(i)$ supported on ∂G.

Let $\{G_m\}_{m=1}^{\infty} \subset \mathcal{G}$ such that $G \subset G_1 \subset G_2 \subset \cdots$, $\bigcup_m G_m = \mathbb{S}$, and G_1 is large enough that $P^v_{ij} = 0$ when $i \in G$, $j \in G^c$. (I.e., it takes at least two steps for the chain to move from G to G_1^c.) For $h \in C(G)$, $h \geq 0$, let ψ_m be the unique solution of $(P^v - I)\psi_m = 0$ on $G_m \cap G^c$, $\psi_m = h$ on G, $\psi_m = 0$ on G_m^c:

$$\psi_m(i) \;=\; \mathbb{E}_i^v \left[h(X_{\tau(G^c)}) \mathbb{I}_{\tau(G^c) < \tau(G_m)} \right].$$

For each $i \in \mathbb{S}$, by the Riesz representation theorem, there exists a measure $q_{1,m}(i, \cdot)$ on G such that

$$\psi_m(i) = \sum_{j \in G} q_{1,m}(i,j) h(j)\,.$$

Note that for $i \in G$, $q_{1,m}(i, \cdot) = \mathbb{I}_{\{i\}}(\cdot)$. For $i \in G^c$, $j \in G$, $q_{1,m}(i,j) \uparrow q_1(i,j) = \mathbb{P}_i^v \left(X_{\tau(G^c)} = j \right)$.

Now let $h_2 \in C(G^c)$, $h_2 \geq 0$, and let φ solve $(P^v - I)\varphi = 0$ on G, $\varphi = h_2$ on G^c. Then by the Riesz representation theorem,

$$\varphi(i) = \mathbb{E}_i^v \left[h_2(X_{\widehat{\tau}_1}) \right] = \sum_{j \in G^c} q_2(i,j) h_2(j).$$

As before, $q_2(i, \cdot) = \mathbb{I}_{\{i\}}(\cdot)$ for $i \in G^c$. For any $i \in \mathbb{S}$, $q_2(i,j) = \mathbb{P}_i^v \left(X_{\widehat{\tau}_1} = j \right)$. For any fixed $j \in G^c$, we can choose $h_2(i) = \mathbb{I}_{\{j\}}(i)$ and solve the Dirichlet problem above to get $\varphi(i) = q_2(i,j)$. Then, from Harnack (Lemma 3.4.3), for all $i, i' \in G$, $j \in G^c$, there is a $C_H > 1$ such that $q_2(i,j) \leq C_H q_2(i',j)$. So, noting that $\widetilde{P}^v(i, \cdot) = \sum_{j \in G^c} q_2(i,j) q_1(j, \cdot)$, any fixed $i_0 \in G$ yields

$$\widetilde{P}^v(i, \cdot) \geq C_H^{-1} \widetilde{P}^v(i_0, \cdot) \qquad \forall i \in G\,.$$

This implies that $\widetilde{P}^v$ is a contraction under the TV norm, and

$$\left\| \sum_{i \in G} (\mu(i) - \mu'(i)) \widetilde{P}^v(i, \cdot) \right\|_{TV} \leq (1 - C_H^{-1}) \|\mu - \mu'\|_{TV} \text{ for } \mu, \mu' \in \mathcal{P}(G) .$$

So (4.4) holds with $\delta = (1 - C_H^{-1})$. □

Lemma 4.3.8. *Let $v \in \Pi_{ssm}$ and $G \in \mathcal{G}$. For each $k \in \mathbb{N}$, define $\widehat{\tau}_k$ as in (4.3), along with the induced chain $\widetilde{X}_n$, transition matrix $\widetilde{P}^v$, and invariant distribution $\widetilde{\mu}_v$ on G. Define $\eta_v \in \mathcal{P}(\mathbb{S})$ by*

$$\sum_{\mathbb{S}} f \eta_v = \frac{\sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} f(X_n) \right] \widetilde{\mu}_v(i)}{\sum_{i \in G} \mathbb{E}_i^v [\widehat{\tau}_2] \, \widetilde{\mu}_v(i)} . \tag{4.5}$$

Then η_v is the invariant distribution of X under v (i.e., $\eta_v P^v = \eta_v$).

Proof. Define the measure μ_v by

$$\sum_{i \in \mathbb{S}} g(i) \mu_v(i) = \sum_{i \in \mathbb{S}} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} g(X_n) \right] \widetilde{\mu}_v(i) \qquad \text{for } g \in C_b(\mathbb{S}) .$$

Let $s \geq 0$. For any $f \in C_b(\mathbb{S})$, we have

$$\begin{aligned}
\mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} \mathbb{E}_{X_n}^v [f(X_s)] \right] &= \mathbb{E}_i^v \left[\sum_{n=0}^{\infty} \mathbb{I}_{t < \widehat{\tau}_2} \mathbb{E}_i^v \left[f(X_{s+n}) \mid \mathcal{F}_n^X \right] \right] \\
&= \mathbb{E}_i^v \left[\sum_{n=0}^{\infty} \mathbb{E}_i^v \left[\mathbb{I}_{n < \widehat{\tau}_2} f(X_{s+n}) \mid \mathcal{F}_n^X \right] \right] \\
&= \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} f(X_{s+n}) \right] .
\end{aligned}$$

Since $\widetilde{\mu}_v$ is stationary at $\widehat{\tau}_{2k}$,

$$\sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=\widehat{\tau}_2}^{\widehat{\tau}_2 + s - 1} f(X_n) \right] \widetilde{\mu}_v(i) = \sum_{i \in G} \mathbb{E}_i^v \left[\mathbb{E}_{X_{\widehat{\tau}_2}}^v \left[\sum_{n=0}^{s-1} f(X_n) \right] \right] \widetilde{\mu}_v(i)$$

$$= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{s-1} f(X_n) \right] \widetilde{\mu}_v(i) \,.$$

Combining yields

$$\begin{aligned}
\sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} f(X_{s+n}) \right] \widetilde{\mu}_v(i) &= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 + s - 1} f(X_n) - \sum_{n=0}^{s-1} f(X_n) \right] \widetilde{\mu}_v(i) \\
&= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 + s - 1} f(X_n) - \sum_{n=\widehat{\tau}_2}^{\widehat{\tau}_2 + s - 1} f(X_n) \right] \widetilde{\mu}_v(i) \\
&= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} f(X_n) \right] \widetilde{\mu}_v(i) \\
&= \sum_{i \in \mathbb{S}} f(i) \mu_v(i) \,.
\end{aligned}$$

Then with $g(i) = \mathbb{E}_i^v [f(X_s)]$,

$$\begin{aligned}
\sum_{i \in \mathbb{S}} \mathbb{E}_i^v [f(X_s)] \, \mu_v(i) &= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} \mathbb{E}_{X_n}^v [f(X_s)] \right] \widetilde{\mu}_v(i) \\
&= \sum_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}_2 - 1} f(X_{s+n}) \right] \widetilde{\mu}_v(i) \\
&= \sum_{i \in \mathbb{S}} f(i) \mu_v(i) \,.
\end{aligned}$$

So μ_v is invariant for X, and $\eta_v := \dfrac{\mu_v}{\mu_v(\mathbb{S})}$ is an invariant probability measure. Since $v \in \Pi_{ssm}$, X is positive recurrent and irreducible, and so η_v is unique. $\square$

4.4 Proofs of Main Results

Proof of Theorem 4.2.1. Let $G_0 \in \mathcal{G}$ be the set that $\mathcal{G}$-separates D from i_0, and note that

$$\mathbb{E}_{i_0}^{v}\left[\sum_{n=0}^{\tau(G^c)-1} h_v(X_n)\right] < \infty \qquad \forall v \in \Pi_{ssm}\,.$$

For a stopping time τ, define

$$\beta_i^v[\tau] \;:=\; \mathbb{E}_i^v\left[\sum_{n=0}^{\tau-1} h_v(X_n)\right].$$

Suppose that the claim does not hold for G_0. Then there exists a sequence of policies $\{v_m\} \subset \Pi_{sm}$ such that $\beta_{i_0}^{v_m}[\tau(G_0^c)] \uparrow \infty$ as $m \to \infty$. Drop to a subsequence of $\{v_m\}$: choose a $v_0 \in \Pi_{ssm}$ such that $\beta_{i_0}^{v_0}[\tau(G_0^c)] > 2$. Find a sequence of sets $\{G_k\} \subset \mathcal{G}$ such that $\bigcup_k G_k = \mathbb{S}$, $G_0 \cup \{i_0\} \subset G_k$ and ${}_{G_k \cap G_0^c}P$ irreducible for each k. Noting that $\beta_{i_0}^{v_0}[\tau(G_0^c) \wedge \tau(G_k)] \uparrow \beta_{i_0}^{v_0}[\tau(G_0^c)]$ as $k \to \infty$, choose $\widehat{G}_1 \in \{G_k\}$ such that $\beta_{i_0}^{v_0}[\tau(G_0^c)] \le 2\beta_{i_0}^{v_0}[\tau(G_0^c) \wedge \tau(\widehat{G}_1)]$. Let

$$\partial\widehat{G}_1 = \{i \in \widehat{G}_1^c | P_{ij}^v > 0 \text{ for some } j \in \widehat{G}_1,\, v \in \Pi_{ssm}\}\,;$$

$\partial\widehat{G}_1$ is finite by Assumption 3.3.1, and $p_1 := \inf_{v \in \Pi_{ssm}} \mathbb{P}_{i_0}^v\left(\tau(G_0^c) > \tau(\widehat{G}_1)\right)$ is strictly positive by the irreducibility of ${}_{\widehat{G}_1}P^v$ and of P^v.

Note that Lemmas 4.3.4 and 4.3.6 imply that for any $i \in \partial\widehat{G}_1$,

$$\beta_i^{v_m}[\tau(\widehat{G}_1^c)] \uparrow \infty \text{ as } m \to \infty\,;$$

if not, the lemmas would imply that the claim *does* hold for G_0. Choose $v_1 \in \{v_m\}$ such that $\min_{i \in \partial\widehat{G}_1} \beta_i^{v_1}[\tau(\widehat{G}_1^c)] > 8p_1^{-1}$, and let

$$\widehat{v}_1(i) = \begin{cases} v_0 & \text{for } i \in \widehat{G}_1\,, \\ v_1 & \text{for } i \in \widehat{G}_1^c\,. \end{cases}$$

Clearly, $\widehat{v}_1 \in \Pi_{sm}$. Combining the above, we obtain

$$\beta_{i_0}^{\widehat{v}_1}[\tau(G_0^c)] \;\geq\; \mathbb{P}_{i_0}^{v}\left(\tau(G_0^c) > \tau(\widehat{G}_1)\right)\left(\min_{i\in\partial\widehat{G}_1}\beta_i^{v_1}[\tau(\widehat{G}_1^c)]\right) > 8\,.$$

As before we can choose $\widehat{G}_2 \in \{G_k\}$ such that $\widehat{G}_1 \cup \partial\widehat{G}_1 \subset \widehat{G}_2$ and $\beta_{i_0}^{\widehat{v}_1}[\tau(G_0^c) \wedge \tau(\widehat{G}_2)] > 4$.

Now we can see the induction needed: suppose that $\widehat{v}_{k-1} \in \Pi_{ssm}$ and $\widehat{G}_k \in \{G_k\}$ such that $\beta_{i_0}^{\widehat{v}_{k-1}}[\tau(G_0^c) \wedge \tau(\widehat{G}_k)] > 2^k$. Choose $v_k \in \Pi_{ssm}$ such that

$$\min_{i\in\partial\widehat{G}_k}\beta_i^{v_k}[\tau(\widehat{G}_k^c)] > 2^{k+2}\left(\inf_{v\in\Pi_{ssm}}\mathbb{P}_{i_0}^{v}\left(\tau(G_0^c) > \tau(\widehat{G}_k)\right)\right)^{-1},$$

which is always possible, as above. Define

$$\widehat{v}_k(i) = \begin{cases}\widehat{v}_{k-1} & \text{for } i \in \widehat{G}_k \\ v_k & \text{for } i \in \widehat{G}_k^c\,.\end{cases}$$

As before, choose $\widehat{G}_{k+1} \in \{G_k\}$ such that $\widehat{G}_k \cup \partial\widehat{G}_k \in \widehat{G}_{k+1}$ and $\beta_{i_0}^{\widehat{v}_k}[\tau(G_0^c)] \leq 2\beta_{i_0}^{\widehat{v}_k}[\tau(G_0^c) \wedge \tau(\widehat{G}_{k+1})]$, so $\beta_{i_0}^{\widehat{v}_k}[\tau(G_0^c) \wedge \tau(\widehat{G}_{k+1})] > 2^{k+1}$.

Each $\widehat{v}_k$ agrees with $\widehat{v}_{k-\ell}$ on $\widehat{G}_{k-\ell+1}$, and the sequence $\widehat{v}_k$ converges to a control $\widehat{v} \in \Pi_{ssm}$ that agrees with $\widehat{v}_k$ on $\widehat{G}_k$ for each $k \geq 1$. Thus $\beta_{i_0}^{\widehat{v}}[\tau(G_0^c) \wedge \tau(\widehat{G}_k)] > 2^k$ for all $k \geq 0$, and so $\beta_{i_0}^{\widehat{v}}[\tau(G_0^c)] = \infty$ which contradicts the original assumption.

Since the claim holds for G_0 and i_0, Lemmas 4.3.4, 4.3.5, and 4.3.6 imply that the claim also holds for any finite $B \in \mathbb{S}$ and any $i \in B^c$. □

Remark 4.4.1.

$$J_\alpha^v(i) \;=\; \mathbb{E}_i^v\left[\sum_{n=0}^{\infty}\alpha^n c_v(X_n)\right] = \mathbb{E}_i^v\left[c_v(X_0) + \sum_{n=1}^{\infty}\alpha^n c_v(X_n)\right]$$

$$= \mathbb{E}_i^v [c_v(X_0)] + \mathbb{E}_i^v \left[\sum_{n=1}^{\infty} \alpha^n c_v(X_n) \right]$$

$$= c_v(i) + \alpha \mathbb{E}_i^v \left[\sum_{n=0}^{\infty} \alpha^n c_v(X_{n+1}) \right].$$

Therefore,

$$J_\alpha^v = c_v + \alpha P^v J_\alpha^v \implies (\alpha P^v - I) J_\alpha^v = -c_v \,.$$

Now, if $\varrho_v = \mu_v c_v = \sum_{i \in \mathbb{S}} \mu_v(i) c_v(i)$ is finite, then:

$$\begin{aligned} \frac{\varrho_v}{1-\alpha} &= \varrho_v \sum_{n=0}^{\infty} \alpha^n = \mu_v c_v \sum_{n=0}^{\infty} \alpha^n \\ &= \mu_v \sum_{n=0}^{\infty} (P^v)^n c_v \alpha^n \\ &= \sum_{i \in \mathbb{S}} \mu_v(i) \mathbb{E}_i^v \left[\sum_{n=0}^{\infty} \alpha^n c_v(X_n) \right] \\ &= \mu_v J_\alpha^v \,. \end{aligned}$$

Proof of Theorem 4.2.2. Let $\widehat{\tau} := \min\{n > \tau(G) : X_n \in G\}$. Then for $i \in G$,

$$\begin{aligned} J_\alpha^v(i) &= \mathbb{E}_i^v \left[\sum_{n=0}^{\infty} \alpha^n c_v(X_n) \right] \\ &= \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n) + J_\alpha^v(X_{\widehat{\tau}}) - (1 - \alpha^{\widehat{\tau}}) J_\alpha^v(X_{\widehat{\tau}}) \right] . \end{aligned} \tag{4.6}$$

From Lemma 4.3.7, there exists a $\delta \in (0,1)$ depending only on G such that

$$\operatorname*{osc}_G \left(\mathbb{E}_{(\cdot)}^v [J_\alpha^v(X_{\widehat{\tau}})] \right) \leq \delta \operatorname*{osc}_G J_\alpha^v. \tag{4.7}$$

Then from (4.6) and (4.7),

$$\operatorname*{osc}_G J_\alpha^v \leq \max_{i \in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n) \right] + \max_{i \in G} \mathbb{E}_i^v \left[(1 - \alpha^{\widehat{\tau}}) J_\alpha^v(X_{\widehat{\tau}}) \right]$$

$$+ \max_{i\in G} \mathbb{E}_i^v \left[J_\alpha^v(X_{\widehat{\tau}})\right] - \min_{i\in G} \mathbb{E}_i^v \left[J_\alpha^v(X_{\widehat{\tau}})\right]$$

$$\leq \max_{i\in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n)\right] + \max_{i\in G} \mathbb{E}_i^v \left[(1-\alpha^{\widehat{\tau}}) J_\alpha^v(X_{\widehat{\tau}})\right] + \delta \operatorname*{osc}_G J_\alpha^v \,,$$

and therefore

$$(1-\delta) \operatorname*{osc}_G J_\alpha^v \leq \max_{i\in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n)\right] + \max_{i\in G} \mathbb{E}_i^v \left[(1-\alpha^{\widehat{\tau}}) J_\alpha^v(X_{\widehat{\tau}})\right]. \quad (4.8)$$

For any $i \in G$,

$$\begin{aligned} \mathbb{E}_i^v \left[(1-\alpha^{\widehat{\tau}}) J_\alpha^v(X_{\widehat{\tau}})\right] &\leq \mathbb{E}_i^v \left[\frac{1-\alpha^{\widehat{\tau}}}{1-\alpha}\right] \max_{j\in G} (1-\alpha) J_\alpha^v(j) \\ &\leq \mathbb{E}_i^v \left[\widehat{\tau}\right] \max_{j\in G} (1-\alpha) J_\alpha^v(j) \,, \end{aligned} \quad (4.9)$$

and from Remark 4.4.1 we get the estimate

$$\min_G (1-\alpha) J_\alpha^v \leq \frac{\varrho_v}{\mu_v(G)} \,.$$

Note that $(\alpha P^v - I) J_\alpha^v = -c_v < 0$, which, as detailed in Remark 3.4.5 implies the existence of a constant $C_1 > 1$ depending only on G such that

$$\max_G J_\alpha^v \leq C_1 \min_G J_\alpha^v \,.$$

Therefore

$$\max_G (1-\alpha) J_\alpha^v \leq C_1 \frac{\varrho_v}{\mu_v(G)} \,, \quad (4.10)$$

which proves (4.2). Let

$$\overline{J_\alpha^v} := \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n)\right],$$

and note that $\overline{J}_\alpha^v$ also satisfies

$$(\alpha P^v - I)\overline{J}_\alpha^v \;=\; -c_v\,.$$

Hence, following an identical argument to Remark 3.4.5, we find that the constant C_1 also satisfies

$$\max_G \overline{J}_\alpha^v \;\leq\; C_1 \min_G \overline{J}_\alpha^v.$$

Then using the bound

$$\min_{i\in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} c_v(X_n)\right] \;\leq\; \varrho_v \max_{i\in G} \mathbb{E}_i^v[\widehat{\tau}]$$

derived from (4.5), we get

$$\max_{i\in G} \mathbb{E}_i^v \left[\sum_{n=0}^{\widehat{\tau}-1} \alpha^n c_v(X_n)\right] \;\leq\; C_1 \varrho_v \max_{i\in G} \mathbb{E}_i^v[\widehat{\tau}]\,. \tag{4.11}$$

Finally, $\mathbb{E}_i^v[\widehat{\tau}]$ is $(P^v - I)$-superharmonic, and so Lemma 3.4.3 guarantees a constant C_2 such that

$$\mathbb{E}_i^v[\widehat{\tau}] \;\leq\; C_2 \mathbb{E}_j^v[\widehat{\tau}] \quad \forall i,j \in G, v \in \Pi_{sm}\,. \tag{4.12}$$

Applying Lemma 4.3.8 with $f(\cdot) = \mathbb{I}_G(\cdot)$ and using Lemma 4.3.1 to find a constant C_3 yields

$$\begin{aligned}\min_{i\in G} \mathbb{E}_i^v[\widehat{\tau}]\mu_v(G) \;&\leq\; \sum_{i\in G} \mathbb{E}_i^v[\widehat{\tau}]\widetilde{\mu}_v(i)\mu_v(G) = \sum_{i\in G} \mathbb{E}_i^v[\tau(G)]\widetilde{\mu}_v(i) \\ &\leq\; \max_{i\in G} \mathbb{E}_i^v[\tau(G)] \leq C_3 < \infty\,. \end{aligned} \tag{4.13}$$

Then combining (4.8) with (4.9), (4.10), (4.11), (4.12), and (4.13) we get

$$(1-\delta)\operatorname*{osc}_G J_\alpha^v \;\le\; \left(1+\frac{1}{\mu(G)}\right)\frac{C_1C_2C_3\varrho_v}{\mu_v(G)}\,. \tag{4.14}$$

Since the constants δ, C_1, C_2, and C_3 depend only on G, we can indeed rewrite (4.14) in the form of (4.6). □

Chapter 5

Countable State Space: Value Iteration

5.1 Introduction

In this chapter, we consider the value iteration and relative value iteration algorithms to determine the average cost and optimal policy for the average cost problem in Section 2.5.3. The optimal policy is can be found via the *value function* $V : \mathbb{S} \to \mathbb{R}_+$ which satisfies the *average cost optimality equation* (ACOE):

$$V(i) \;=\; \min_{u \in \mathbb{U}} \left[r(i,u) + P^u V(i) \right] - \varrho^*, \quad i \in \mathbb{S}. \tag{5.1}$$

A stationary policy v^* is optimal if it satisfies

$$v^*(i) \in \arg\min_{u \in \mathbb{U}} \left[r(i,u) + P^u V(i) \right], \quad i \in \mathbb{S}.$$

We therefore take $v^* \in \Pi_{sm}$ to be a selector from the minimizer. For an infinite state space, it is generally not feasible to solve the ACOE directly, so a common approach is to a find a sequence of functions that might converge to the value function. Two closely related sequences frequently considered are given by the relative value iteration (RVI),

$$\varphi_{n+1}(i) \;=\; \min_{u \in \mathbb{U}} \left[r(i,u) + P^u \varphi_n(i) \right] - \varphi_n(0), \tag{5.2}$$

and the value iteration (VI)

$$\overline{\varphi}_{n+1}(i) \;=\; \min_{u\in\mathbb{U}} [r(i,u) + P^u\overline{\varphi}_n(i)] - \varrho^*, \quad \overline{\varphi}_0 = \varphi_0. \tag{5.3}$$

We seek conditions and initial values φ_0 that will ensure the VI and RVI converge to a valid solution of (5.1). Our main assumption will be that the cost function is *near-monotone*; that is, it satisfies the following condition:

Assumption 5.1.1.

$$\left\{ i \in \mathbb{S} : \min_{u\in\mathbb{U}} r(i,u) \le \varrho^* + \delta \right\} \text{ is finite for some } \delta \in (0,1). \tag{5.4}$$

The near-monotone condition encourages stable behavior by penalizing the system for moving away from a "central" set, and also implies that ϱ^* is finite. We will also impose additional assumptions relating the cost function to the value function.

Note that by standard dynamic programming iteration (see, e.g., [10]), the VI (5.3) can be written in the following stochastic form:

$$\overline{\varphi}_n(i) \;=\; \inf_{U\in\mathfrak{U}} \mathbb{E}_i^U \left[\varphi_0(X_n) + \sum_{k=0}^{n-1} (r(X_k,U_k) - \varrho^*) \right]. \tag{5.5}$$

This representation makes it clear that $\overline{\varphi}_n$ is simply the n-horizon optimal control problem with running cost $(r - \varrho^*)$ and terminal cost φ_0. If the MDP is appropriately ergodic, in the long-term the minimizing policy in the VI will approach the optimal policy, and $\overline{\varphi}_n$ will converge to a solution of (5.1) (i.e., the optimal value function plus a constant). Also, by (5.5), if $\overline{\varphi}_n$ converges to

a solution of (5.1) then

$$\inf_{U\in\mathfrak{U}} \mathbb{E}_i^U \left[\sum_{k=0}^{n-1}(r(X_k,U_k)-\varrho^*)\right]$$

also converges to a solution of (5.1). Hence, we may say that the VI is truly representative of the long-term asymptotics of the finite horizon problem.

On the other hand, the RVI is normalized at every step, so although the minimizing policy will approach the optimal policy, the relative value function φ_n may converge while moving arbitrarily far away from the optimal value function. The distinction is clearly visible in the following relationships, proved later in Lemma 5.4.1:

$$\begin{aligned}
\bar{\varphi}_n(i) &= \varphi_n(i) - n\varrho^* + \sum_{m=0}^{n-1}\varphi_m(0), \\
\varphi_n(i) &= \bar{\varphi}_n(i) - \bar{\varphi}_{n-1}(0) + \varrho^*, \qquad \text{for all } i\in\mathbb{S},\ n\geq 1.
\end{aligned}$$

The first relationship shows that the RVI might converge while the cumulative sum of normalizing terms makes the VI diverge. However, the second relationship clearly implies that if the VI converges, then the RVI must also.

Though VI algorithms can be traced back to sequential decision models [53], developed around the time that dynamic programming was being formalized, results for countably many states and average cost were not developed until the 1970s and later. For non-finite state spaces, results showing the convergence of the VI algorithm rely on strong blanket stability assumptions. [34] showed that the value iteration converges when $V-\varphi_0$ is bounded, thereby limiting the one-step behavior of the algorithm. Sennott [51] instead assumed

that the value function is bounded above by another function that is integrable with respect to the optimal stationary distribution, a condition that is difficult to verify. Aviv and Federgruen in [6] address the VI by first showing convergence under a strong blanket stability assumption involving the optimal value function. They then show a sufficient condition involving an order function in lieu of the optimal value function, but the condition still requires blanket stability for policies. Hence, in the literature, convergence of the VI depends on blanket atability assumptions that are overly restrictive or difficult to verify.

Other efforts have focused almost exclusively on the more tractable RVI rather than the VI. An assumption similar to that of [51] is used in [52], combined with various conditions based on several others' frameworks, to show convergence of the average cost. The sufficient conditions include some from [21] requiring bounded costs, as well as a near monotone condition (5.4) from Borkar, as in [16]. Authors in [22] proved that the RVI converges for unbounded costs when one assumes that there is a global (in the sense of all possible controls) Lyapunov function. In [20], Cavazos-Cadena showed that the RVI converges under a slightly stronger version of the near monotone condition. Rather than the single set defined in (5.4), the author requires that all of the sub-level sets of the cost function are finite:

$$\left\{ i \in \mathbb{S} : \min_{u \in \mathbb{U}} r(i,u) \leq b \right\} \text{ is finite for any } b > 0.$$

In a related work, [24] argue convergence by initializing the RVI with a regular

policy v_0, a function $V_0 > 0$, and a constant ϱ_0 that satisfy

$$P^{v_0} V_0 \leq V_0 - r^{v_0} + \varrho_0.$$

With such an initialization, each step of the RVI algorithm yields a regular policy and a Lyapunov function, thereby guaranteeing convergence to a regular policy with Lyapunov stability. However, finding an initial policy and corresponding function effectively requires solving an equation (or inequality) of precisely the type which the value iteration algorithm is intended to avoid. Ultimately, though, all of these results avoid convergence of the VI, which for the models used is not guaranteed.

In this work, we present two new sufficient conditions for the convergence of the VI algorithm that do not require a uniform stability condition. Our weaker assumption is that the value function is integrable with respect to the optimal invariant distribution. Note that this requires stability under the optimal policy only, not in general. Under this condition, the VI algorithm converges when initialized with a function that is similar in growth to the value function. We also assert a stronger condition requiring that the value function grow no faster than the cost function. While somewhat restrictive, various problems with near-monotone cost do satisfy this requirement structurally, including problems with quadratic-like costs. Under this assumption, initializing with a constant function will guarantee convergence. Our approach adapts the controlled diffusion results in [4], for the countable state space, but unlike in Chapter 4 does not require onerous structural assumptions on the transition probabilities.

In the next section we describe our main assumptions and lay out some additional notation needed for the ensuing results. Then, in Section 5.3 we present the main results and discuss some of the implications thereof. Proofs are deferred until Sections 5.4–5.5, where we show a number of supporting lemmas before proving the main theorems.

5.2 Assumptions and Additional Notation

Throughout this chapter, we will assume that the Markov decision process (MDP) is irreducible and aperiodic (see Assumptions 3.2.1–3.2.2) for all admissible controls. Since the transition kernel is a stochastic matrix, if V solves (5.1) then so does $V + c$ for any $c \in \mathbb{R}$. We therefore fix a particular solution V that solves (5.1) with $\min_{i\in\mathbb{S}} V(i) = 1$. We then define for $f : \mathbb{S} \to \mathbb{R}$ the norm

$$\|f\|_V \;:=\; \sup_{i\in\mathbb{S}} \frac{|f(i)|}{V(i)},$$

and the set

$$\mathcal{O}_V \;:=\; \{f : \mathbb{S} \to \mathbb{R} :\; \|f\|_V < \infty, f \geq 0\}.$$

Let $\mathfrak{v} = \{\mathfrak{v}_m, m \in \mathbb{N}\}$ be a selector from the minimizer in (5.3) corresponding to a solution $\overline{\varphi}$. Note that $\mathfrak{v}$ is also a selector from the minimizer in (5.2) when φ and $\overline{\varphi}$ are initialized with the same φ_0. At the n^{th} step of the VI, define the (nonstationary) Markov control

$$\hat{v}^n \;:=\; \{\hat{v}^n_m = \mathfrak{v}_{n-m}, m \in \mathbb{N}, m < n\}. \tag{5.6}$$

If the cost function r is replaced with $r + c$ for some $c \in \mathbb{R}$, the resulting

average cost will simply be $\varrho^* + c$ and the optimal policy will be unchanged. Hence, without loss of generality we will assume $\min_{\mathbb{S}\times\mathbb{U}} r = 1$. To simplify analysis, we will also occasionally use

$$\bar{r} := r - \varrho^*.$$

Assumption 5.2.1. There exist positive constants θ_1 and θ_2 such that

$$\min_{u\in\mathbb{U}} r(i,u) \geq \theta_1 V(i) - \theta_2 \quad \forall i \in \mathbb{S}.$$

Since V is positive everywhere, without loss of generality we assume $\theta_1 \in (0,1)$, which will be useful in proving some essential results.

When a policy v induces a stable process, we denote by μ_v the corresponding invariant probability distribution on $\mathbb{S}$. The existence of an optimal invariant distribution μ_{v^*} is shown in [16] to be a consequence of the near-monotone assumption. Clearly, when the average cost ϱ^* is finite, $\varrho^* = \mu_{v^*}[r] < \infty$. Hence, the following assumption is also asserting a very general structural similarity between r and V.

Assumption 5.2.2. There exists an optimal invariant probability distribution μ_{v^*} such that

$$\mu_{v^*}[V] = \sum_{i\in\mathbb{S}} V(i)\mu_{v^*}(i) < \infty\,.$$

Equation 5.2.1 implies 5.2.2 because

$$\mu_{v^*}[V] \leq \frac{\mu_{v^*}[\min_{u\in\mathbb{U}} r(\cdot,u)] + \theta_2}{\theta_1} \leq \frac{\mu_{v^*}[r(\cdot,v^*(\cdot))] + \theta_2}{\theta_1} = \frac{\varrho^* + \theta_2}{\theta_1} < \infty\,.$$

5.3 Main Results

Using the notation of dynamical systems, we consider the semi-cascades $\overline{\Phi}_n[\varphi_0]$ of (5.3) and $\Phi_n[\varphi_0]$ of (5.2). Let $\mathcal{E}$ denote the set of solutions of the ACOE in (5.1). Recalling that the solution of (5.1) is unique up to a constant, define

$$\mathcal{E} \;:=\; \{V + c : c \in \mathbb{R}\}.$$

For any particular $c \in \mathbb{R}$, we define the set

$$\mathcal{G}_c \;:=\; \{h : \mathbb{S} \to \mathbb{R} :\; \|h\|_V < \infty, h - V \geq c\}.$$

Theorem 5.3.1. *Suppose Assumption 5.2.2 holds and $\varphi_0 \in \mathcal{G}_c$ for some $c \in \mathbb{R}$. Then $\overline{\Phi}_n[\varphi_0]$ converges to $c_0 + V \in \mathcal{E}$ for some $c_0 \in \mathbb{R}$ such that*

$$0 \leq c_0 \leq \mu_{v^*}[\varphi_0 - V], \tag{5.7}$$

and $\Phi_n[\varphi_0]$ converges to $V - V(0) + \varrho^*$.

Under the stronger assumption 5.2.1, the same convergence can be shown with notably relaxed conditions on the initial function φ_0:

Theorem 5.3.2. *Suppose Assumption 5.2.1 holds and $\varphi_0 \in \mathcal{O}_V$. Then the semi-cascade $\overline{\Phi}_n[\varphi_0]$ converges to a point $c_0 + V \in \mathcal{E}$ satisfying*

$$-\frac{\varrho^* + \theta_2}{\theta_1} \;\leq\; c_0 \;\leq\; \|\varphi_0\|_V \frac{\varrho^* + \theta_2}{\theta_1}, \tag{5.8}$$

and therefore $\Phi_n[\varphi_0]$ converges to $V - V(0) + \varrho^*$ as $t \to \infty$.

5.4 Supporting Lemmas

We begin by proving a number of essential, intermediate results.

Lemma 5.4.1. *The solutions φ and $\overline{\varphi}$ of* (5.2) *and* (5.3)*, respectively, satisfy*

$$\overline{\varphi}_n(i) \;=\; \varphi_n(i) - n\varrho^* + \sum_{m=0}^{n-1} \varphi_m(0)\,, \tag{5.9}$$

$$\varphi_n(i) - \varphi_n(0) \;=\; \overline{\varphi}_n(i) - \overline{\varphi}_n(0)\,, \tag{5.10}$$

$$\varphi_n(i) \;=\; \overline{\varphi}_n(i) - \overline{\varphi}_{n-1}(0) + \varrho^*\,. \tag{5.11}$$

for all $i \in \mathbb{S}$ and $n \geq 1$.

Proof. Let φ be a solution of (5.2) and suppose (5.9) is true for a particular $n \in \mathbb{N}$. Then

$$\begin{aligned}
\overline{\varphi}_{n+1}(i) \;&=\; \min_{u\in\mathbb{U}} [r(i,u) + P^u\overline{\varphi}_n(i)] - \varrho^* \\
&=\; \min_{u\in\mathbb{U}} \left[r(i,u) + P^u \left(\varphi_n(i) - n\varrho^* + \sum_{m=0}^{n-1} \varphi_m(0) \right) \right] - \varrho^* \\
&=\; \min_{u\in\mathbb{U}} [r(i,u) + P^u\varphi_n(i)] - n\varrho^* + \sum_{m=0}^{n-1} \varphi_m(0) - \varrho^* \\
&=\; \varphi_{n+1}(i) + \varphi_n(0) - (n+1)\varrho^* + \sum_{m=0}^{n-1} \varphi_m(0) \\
&=\; \varphi_{n+1}(i) - (n+1)\varrho^* + \sum_{m=0}^{n} \varphi_m(0)\,.
\end{aligned}$$

Since (5.9) is trivially satisfied for $n = 0$, it must also be true for all $n \geq 0$. (5.10) then follows directly from (5.9). Also from (5.9), and using (5.10),

$$\overline{\varphi}_n(i) - \overline{\varphi}_{n-1}(i) \;=\; \varphi_n(i) - \varphi_{n-1}(i) + \varphi_{n-1}(0) - \varrho^*$$

$$= \varphi_n(i) - \overline{\varphi}_{n-1}(i) + \overline{\varphi}_{n-1}(0) - \varrho^*.$$

and rearranging yields (5.11). □

Lemma 5.4.2. *For each $n \geq 0$ and $i \in \mathbb{S}$, with $\hat{v}^n$ as in (5.6), $\overline{\varphi}$ satisfies*

$$P^{\hat{v}^n}(\overline{\varphi}_n(i) - V(i)) \;\leq\; \overline{\varphi}_{n+1}(i) - V(i) \;\leq\; P^{v^*}(\overline{\varphi}_n(i) - V(i))\,, \tag{5.12a}$$

$$\mathbb{E}_i^{\hat{v}^n}[\varphi_0(X_n) - V(X_n)] \;\leq\; \overline{\varphi}_{n+1}(i) - V(i) \;\leq\; \mathbb{E}_i^{v^*}[\varphi_0(X_n) - V(X_n)]\,. \tag{5.12b}$$

Proof. For the right inequality in (5.12a), from (5.3) and (5.1), we have

$$\begin{aligned} 0 \;&\leq\; r(i, v^*(i)) + P^{v^*}\overline{\varphi}_n(i) - \min_{u\in\mathbb{U}}\left[r(i,u) + P^u\overline{\varphi}_n(i)\right] \\ &=\; P^{v^*}\overline{\varphi}_n(i) - P^{v^*}V(i) - \min_{u\in\mathbb{U}}\left[r(i,u) + P^u\overline{\varphi}_n(i)\right] + r(i, v^*(i)) + P^{v^*}V(i) \\ &=\; P^{v^*}(\overline{\varphi}_n(i) - V(i)) - (\overline{\varphi}_{n+1}(i) - V(i)). \end{aligned}$$

For the left, again from (5.1) and using the definition of $\hat{v}$ with (5.3), we have

$$\begin{aligned} 0 \;&\leq\; r(i, \hat{v}^n(i)) + P^{\hat{v}^n}V(i) - \min_{u\in\mathbb{U}}\left[r(i,u) + P^uV(i)\right] \\ &=\; r(i, \hat{v}^n(i)) + P^{\hat{v}^n}\overline{\varphi}_n(i) - \min_{u\in\mathbb{U}}\left[r(i,u) + P^uV(i)\right] + P^{\hat{v}^n}V(i) - P^{\hat{v}^n}\overline{\varphi}_n(i) \\ &=\; (\overline{\varphi}_{n+1}(i) - V(i)) - P^{\hat{v}^n}(\overline{\varphi}_n(i) - V(i)). \end{aligned}$$

Extending to (5.12b) is accomplished by iterating (5.12a) and treating P^{v^*} and $P^{\hat{v}^n}$ as operators on $(\varphi_0 - V)$. □

The following result is well-known but reproduced here for completeness.

Lemma 5.4.3. *If a non-negative function $f : \mathbb{S} \to \mathbb{R}$ and transition probability kernel P satisfy*

$$Pf \leq \alpha + \beta f$$

for constants $\alpha \in \mathbb{R}$ and $\beta \in (0, 1)$, then a chain $\{X_0, X_1, \dots\}$ governed by P with $X_0 = i$ satisfies

$$\mathbb{E}_i[f(X_n)] \leq \frac{\alpha}{1-\beta} + \beta^n f(i) \qquad \forall i \in \mathbb{S}.$$

Proof. Using recursion, with a chain $\{X_i\}$ as described,

$$\mathbb{E}_i[f(X_1)] = Pf(i) \leq \alpha + \beta f(i) \leq \frac{\alpha}{1-\beta} + \beta f(i) \qquad i \in \mathbb{S}\,.$$

Then for $n \geq 2$,

$$\begin{aligned}
\mathbb{E}_i[f(X_{n-1})] \;&\leq\; \alpha \sum_{k=0}^{n-2} \beta^k + \beta^{n-1} f(i) \\
\implies \mathbb{E}_i[f(X_n)] \;&=\; P\mathbb{E}_i[f(X_{n-1})] \leq P\biggl(\alpha \sum_{k=0}^{n-2} \beta^k + \beta^{n-1} f(i)\biggr) \\
&=\; \alpha \sum_{k=0}^{n-2} \beta^k + \beta^{n-1} Pf(i) \leq \alpha \sum_{k=0}^{n-1} \beta^k + \beta^n f(i) \\
&\leq\; \alpha \sum_{k=0}^{\infty} \beta + \beta^n f(i) = \frac{\alpha}{1-\beta} + \beta^n f(i)\,.
\end{aligned}$$

□

Lemma 5.4.4. *Under Assumption 5.2.1,*

$$\mathbb{E}_i^{v^*}[V(X_n)] \leq \frac{\varrho^* + \theta_2}{\theta_1} + (1-\theta_1)^n V(i)\,.$$

Proof. Applying (5.1), we obtain

$$(P^{v^*} - I)V \;=\; P^{v^*}V - P^{v^*}V - r(\cdot, v^*) + \varrho^*$$

$$= \varrho^* - r(\cdot, v^*) \le \varrho^* + \theta_2 - \theta_1 V\,,$$

and thus,

$$P^{v^*} V \;\le\; \varrho^* + \theta_2 + (1-\theta_1) V\,.$$

Then an application of Lemma 5.4.3 yields the result. □

Lemma 5.4.5. *For any filtration* $\{D_\ell : \ell \in \mathbb{N}\}$ *of* $\mathbb{S}$*,*

$$\mathbb{E}_i^{\hat{v}^n}\left[\overline{\varphi}_{n-\tau(D_\ell)}(X_{\tau(D_\ell)})\mathbb{I}_{\tau(D_\ell)<n}\right] \xrightarrow[\ell\to\infty]{} 0\,.$$

Proof. Iterating (5.3) using the definition of $\hat{v}^n$, we get for any $n, \tau > 0$

$$\begin{aligned}\overline{\varphi}_n(i) \;&=\; \sum_{k=0}^{\tau\wedge n-1} P^{\hat{v}^n(k)}\bar{r}(i,\hat{v}^n(i)) + P^{\hat{v}^n(n)}\left(\mathbb{I}_{\tau\ge n}\,\varphi_0(i) + \mathbb{I}_{\tau<n}\,\overline{\varphi}_{n-\tau}(i)\right)\\ &=\; \mathbb{E}_i^{\hat{v}^n}\left[\sum_{k=0}^{\tau\wedge n-1}\bar{r}(X_k,\hat{v}^n(X_k)) + \mathbb{I}_{\tau\ge n}\,\varphi_0(i)\right]\\ &\qquad\qquad + \mathbb{E}_i^{\hat{v}^n}\left[\mathbb{I}_{\tau<n}\,\overline{\varphi}_{n-\tau}(i)\right]\,. \qquad (5.13)\end{aligned}$$

If $\tau = \tau(D_\ell)$ then

$$\mathbb{P}^{\hat{v}^n}\left(\tau(D_\ell) \ge n\right) \xrightarrow[\ell\to\infty]{} 1,$$

so the first term in (5.13) tends to the right-hand side of (5.5) by monotone convergence and the result follows. □

Lemma 5.4.6. *When* $\varphi_0 \in \mathcal{O}_V$*,* $\overline{\varphi}_n \ge -n\varrho^*$ *and satisfies*

$$\|\overline{\varphi}_n\|_V \le (1 + n\varrho^*)\max\{1, \|\varphi_0\|_V\}$$

for all $n \in \mathbb{N}$*.*

The following results and implications are restated here for clarity, and to direct the reader to the appropriate sources.

Lemma 5.4.7 (See [16, Chapter V])**.** *Under 5.4 and 3.2.2, there exists an optimal stationary policy v^* with an invariant distribution μ_{v^*}.*

Lemma 5.4.8. *Under Assumptions 5.4 and 3.2.2, the chain satisfies the conditions of the f-Norm Ergodic Theorem [43, Theorem 14.0.1] with $f(i) = r(i, v^*(i))$.*

Proof. Since r is finite-valued, 5.4 implies that $\varrho^* < \infty$. With optimal policy v^* and μ_{v^*} corresponding invariant distribution, let $f(i) = r(i, v^*(i))$. Then $\mu_{v^*}[f] = \varrho^* < \infty$, satisfying condition (i) of [43, Theorem 14.0.1]. □

Lemma 5.4.9. *Under Assumption 5.2.2, $\mathbb{E}_i^{v^*}[V(X_n)] \to \mu_{v^*}[V]$ as $n \to \infty$.*

Proof. Assumption 5.2.2 directly satisfies condition (i) of [43, Theorem 14.0.1], and the hypothesis is a direct consequence. □

A related essential result is the following:

Lemma 5.4.10. *Under Assumption 5.2.2, there exists a constant $M > 0$ such that*

$$\sup_{n \geq 0} \mathbb{E}_i^{v^*}[V(X_n)] \leq M(V(i) + 1), \qquad \forall i \in \mathbb{S}. \tag{5.14}$$

Proof. Let $B \subset \mathbb{S}$ be the finite set defined in 5.4, and recall ϱ^* and δ from the same definition. Also let $r^*(i) = r(i, v^*(i))$, and define a function $f : \mathbb{S} \to \mathbb{R}$

as

$$f(x) = \begin{cases} \dfrac{r^*(x) - \varrho^*}{\delta} & x \in B^c, \\ 1 & x \in B. \end{cases}$$

Then $f \geq 1$ and V/δ satisfies [43, Theorem 14.0.1, (iii)]:

$$\begin{aligned} P^{v^*}V - V = \varrho^* - r^* &\leq -\delta f + (\delta + \varrho^* - r^*)\mathbb{I}_B \\ &\leq -\delta f + (\delta + \varrho^*)\mathbb{I}_B. \end{aligned}$$

Assumption 5.2.2 with [43, Theorem 14.0.1] then further implies that there exists a constant $M_1 < \infty$ such that

$$\sum_{k=0}^{\infty} \|(P^{v^*})^k(i,\cdot) - \mu_{v^*}\|_{(f)} \leq M_1\big(\tfrac{V(i)}{\delta} + 1\big)\,,$$

where for any signed measure $\|\cdot\|_{(f)}$ is defined as

$$\|\nu\|_{(f)} := \sup_{g:|g|\leq f} |\nu[g]|\,.$$

If we define a new constant $M_2 = \max\{\varrho^* + \delta, \max_{x\in B} r^*(x)\}$, then $r^* \leq M_2 f$ on all of $\mathbb{S}$ and therefore $\|\cdot\|_{(r^*)} \leq \|\cdot\|_{(M_2 f)} \leq M_2\|\cdot\|_{(f)}$. Then using (5.1), for any $n \geq 0$,

$$\begin{aligned} \mathbb{E}_i^{v^*}[V(X_n)] &= (P^{v^*})^n V(i) = V(i) + \sum_{k=0}^{n-1} (P^{v^*})^k (P^{v^*}V(i) - V(i)) \\ &= V(i) + \sum_{k=0}^{n-1} (P^{v^*})^k (\varrho^* - r^*(i)) \\ &\leq V(i) + \sum_{k=0}^{\infty} |(P^{v^*})^k r^*(i) - \varrho^*| \\ &\leq V(i) + \sum_{k=0}^{\infty} \|(P^{v^*})^k(i,\cdot) - \mu_{v^*}\|_{(r^*)} \end{aligned}$$

$$
\begin{aligned}
&\leq V(i) + \sum_{k=0}^{\infty} M_2 \|(P^{v^*})^k(i,\cdot) - \mu_{v^*}\|_{(f)} \\
&\leq V(i) + M_2 M_1 \big(\tfrac{V(i)}{\delta} + 1\big) \\
&\leq V(i) + \frac{M_2 M_1}{\delta}(V(i)+1),
\end{aligned}
$$

Then (5.14) is satisfied with $M = \frac{(M_2 M_1)}{\delta} + 1$. □

5.5 Proofs of the Main Results

Proof of Theorem 5.3.1. Under Assumption 5.2.2, $\mathcal{G}_c$ is positively invariant for $\overline{\Phi}_n$ since (using Lemma 5.4.2)

$$
c \leq \overline{\varphi}_n - V \Rightarrow c = P^{\hat{v}^n} c \leq P^{\hat{v}^n}(\overline{\varphi}_n - V) \leq \overline{\varphi}_{n+1} - V
$$

and with Lemma 5.4.8, for all $i \in \mathbb{S}$ and $n \in \mathbb{N}$,

$$
\begin{aligned}
c \;&\leq\; \overline{\varphi}_{n+1}(i) - V(i) \leq \mathbb{E}_i^{v^*}[\varphi_0(X_n) - V(X_n)] \\
&\leq\; \|\varphi_0 - V\|_V \, \mathbb{E}_i^{v^*}[V(X_n)] \\
&\leq\; m_r \|\varphi_0 - V\|_V \, (V(i)+1)\,. \qquad (5.15)
\end{aligned}
$$

Since translating φ_0 by a constant translates the entire orbit $\{\overline{\Phi}_n[\varphi_0], n \geq 0\}$ by the same constant, without loss of generality assume $c = 0$.

From (5.5), for each $i \in \mathbb{S}$ and $n \in \mathbb{N}$,

$$
\overline{\Phi}_n[\varphi_0](i) \leq \mathbb{E}_i^{v^*}\left[\sum_{k=0}^{n-m-1} (r(X_k, v^*(X_k)) - \varrho^*) + \overline{\Phi}_m[\varphi_0](X_{n-m})\right] \qquad (5.16)
$$

for any $m \in \{0, \cdots, n\}$. Since $\overline{\Phi}_n[\varphi_0](i) - V(i) \geq 0$, and $\mu_{v^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is finite, then (5.16) with $m = n-1$ yields

$$\mu_{v^*}\left[\overline{\Phi}_n[\varphi_0]\right] \;\leq\; \mu_{v^*}\left[\overline{\Phi}_{n-1}[\varphi_0]\right].$$

Since the cascade remains in $\mathcal{G}_0$, the map $n \to \mu_{v^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is therefore non-increasing and bounded below, so must be constant on the ω-limit set of φ_0 under $\overline{\Phi}_n$, denoted $\omega(\varphi_0)$. Because (5.15) implies $\sup_{n\geq 0}\|\overline{\Phi}_n[\varphi_0]\|_V < \infty$, $\{\overline{\Phi}_n[\varphi_0]\}$ are uniformly bounded in the weighted norm. By a standard diagonal argument, it follows that the limit set $\omega(\varphi_0)$ is non-empty. Let $h \in \omega(\varphi_0)$, and define the non-negative (by Lemma 5.4.2) function

$$f(n,i) = P^{v^*}\left(\overline{\Phi}_{n-1}[h](i) - V(i)\right) - \left(\overline{\Phi}_n[h](i) - V(i)\right).$$

Then

$$\mathbb{E}_i^{v^*}\left[\sum_{m=0}^{n-1} f(n-m, X_m)\right] = \mathbb{E}_i^{v^*}\left[h(X_n) - V(X_n)\right] + V(i) - \overline{\Phi}_n[h](i). \quad (5.17)$$

Integrating with respect to the invariant distribution μ_{v^*} yields

$$\sum_{m=0}^{n-1}\sum_{i\in\mathbb{S}} f(n-m,i)\mu_{v^*}(i) \;=\; \sum_{i\in\mathbb{S}}\left(h(i) - \overline{\Phi}_n[h](i)\right)\mu_{v^*}(i) \quad \forall n \in \mathbb{N}. \quad (5.18)$$

Since both h and $\overline{\Phi}_n[h]$ are in $\omega(\varphi_0)$, the right-hand side of (5.18) is equal to zero and therefore $f(n,i) = 0$, (n,i)-almost everywhere. Using Lemma 5.4.9, (5.17) becomes

$$\lim_{n\to\infty} \overline{\Phi}_n[h](i) \;=\; V(i) + \mu_{v^*}\left[h - V\right].$$

Therefore $\omega(\varphi_0) \subset \mathcal{E} \cap \mathcal{G}_0$, and since $\mu_{v^*}\left[V - h\right]$ is a constant, the limit set must be a singleton. Because $\mu_{v^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is non-increasing in n, the inequality

(5.7) is satisfied. Therefore, by Lemma 5.4.1, $\Phi_n[\varphi_0]$ converges pointwise to $V - V(0) - \varrho^*$. $\square$

Proof of Theorem 5.3.2. For $\epsilon > 0$, let $\overline{\varphi}^\epsilon$ be the solution of (5.3) with initial data $\varphi_0 + \epsilon V$, and let $\{\hat{v}^n_\epsilon : n = 0, 1, \dots\}$ be the corresponding Markov control, as in (5.6). For convenience let $\alpha = (1 - \theta_1)$, $C = \frac{\varrho^* + \theta_2}{\theta_1}$, and let

$$f^\epsilon_n(i) \;:=\; \overline{\varphi}^\epsilon_n(i) - (1 - \alpha^n)(V(i) - C).$$

Noting that $(P^{\hat{v}^n_\epsilon} - I)V(i) \geq -r(i, \hat{v}^n_\epsilon) + \varrho^*$ from (5.1), we have

$$\begin{aligned}
F^\epsilon_n(i) \;&:=\; f^\epsilon_n(i) - P^{\hat{v}^n_\epsilon} f^\epsilon_{n-1}(i) \\
&=\; r(i, \hat{v}^n_\epsilon(i)) - \varrho^* - \theta_1 \alpha^{n-1}(V(i) - C) \\
&\qquad\qquad + (1 - \alpha^{n-1})(P^{\hat{v}^n_\epsilon} - I)(V(i) - C) \\
&\geq\; r(i, \hat{v}^n_\epsilon(i)) - \varrho^* - \theta_1 \alpha^{n-1}(V(i) - C) \\
&\qquad\qquad + (1 - \alpha^{n-1})(-r(i, \hat{v}^n_\epsilon(i)) + \varrho^* - C) \\
&=\; \alpha^{n-1} \left(-\theta_1 V(i) + \theta_2 + r(i, \hat{v}^n_\epsilon(i)) + C\right) \\
&\geq\; \alpha^{n-1} \left(-\theta_1 V(i) + \theta_2 + \theta_1 V(i) - \theta_2\right) \\
&=\; 0 \qquad \forall (i, n) \in \mathbb{S} \times \mathbb{N}\,.
\end{aligned}$$

Let $\{D_m : m \in \mathbb{N}\}$ be a filtration of $\mathbb{S}$; that is, each $D_m \subset \mathbb{S}$ is finite, $D_0 \subset D_1 \subset \cdots$, and $\bigcup_{m=0}^\infty D_m = \mathbb{S}$. Let

$$\tau^n_m \;=\; \min\{n, \tau(D_m)\},$$

and using Dynkin's formula from Corollary 3.4.9,

$$\begin{aligned} f_n^\epsilon(i) &= \mathbb{E}_i^{\hat{v}_\epsilon^n}\left[\sum_{k=0}^{\tau_m^n-1} F_{(n-k)}^\epsilon(X_k) + f_{(n-\tau_m^n)}^\epsilon(X_{\tau_m^n})\right] \\ &= \mathbb{E}_i^{\hat{v}_\epsilon^n}\left[\sum_{k=0}^{\tau_m^n-1} F_{(n-k)}^\epsilon(X_k) + (\varphi_0(X_n)+\epsilon V(X_n))\mathbb{I}_{\{n\le\tau(D_m)\}}\right] \\ &\qquad + \mathbb{E}_i^{\hat{v}_\epsilon^n}\left[f_{(n-\tau(D_m))}^\epsilon(X_{\tau(D_m)})\mathbb{I}_{\{n>\tau(D_m)\}}\right]. \end{aligned} \tag{5.19}$$

From Lemma 5.4.5 we have

$$\mathbb{E}_i^{\hat{v}_\epsilon^n}\left[f_{(n-\tau(D_m))}^\epsilon(X_{\tau(D_m)})\mathbb{I}\{n>\tau(D_m)\}\right] \xrightarrow[m\to\infty]{} 0 \quad \forall (n,i)\in\mathbb{N}\times\mathbb{S}\,. \tag{5.20}$$

Then letting $m\to\infty$ in (5.19), using Fatou's lemma and (5.20), we have $f_n^\epsilon(i)\ge 0$ for all $(n,i)\in\mathbb{N}$. By construction, $\overline{\varphi}^\epsilon \ge \overline{\varphi}$ and $\overline{\varphi}^\epsilon$ decreases with ϵ. So each $\overline{\varphi}^\epsilon$ satisfies

$$\begin{aligned} \overline{\varphi}_{n+1}^\epsilon(i) &= \min_{u\in\mathbb{U}}[r(i,u)+P^u\overline{\varphi}_n^\epsilon(i)] - \varrho^*, \\ \overline{\varphi}_0^\epsilon(i) &= \varphi_0(i)+\epsilon V(i)\,, \end{aligned}$$

and $\overline{\varphi}^\epsilon \downarrow \overline{\varphi}^0$ for some pointwise limit $\overline{\varphi}^0$. Clearly, $\overline{\varphi}_0^0=\varphi_0$, and so if we suppose that $\overline{\varphi}_n^0=\overline{\varphi}_n$ for some $n>0$, then

$$\begin{aligned} \overline{\varphi}_{n+1}^\epsilon(i)-\overline{\varphi}_{n+1}(i) &= \min_{u\in\mathbb{U}}[r(i,u)+P^u\overline{\varphi}_n^\epsilon(i)] - \varrho^* - \overline{\varphi}_{n+1}(i) \\ &= \min_{u\in\mathbb{U}}[r(i,u)+P^u\overline{\varphi}_n + P^u(\overline{\varphi}_n^\epsilon(i)-\overline{\varphi}_n(i))] - \varrho^* - \overline{\varphi}_{n+1}(i) \\ &\le r(i,\hat{v}_n(i)) + P^{\hat{v}_n}\overline{\varphi}_n(i) + P^{\hat{v}_n}(\overline{\varphi}_n^\epsilon(i)-\overline{\varphi}_n(i)) - \varrho^* - \overline{\varphi}_{n+1}(i) \\ &= P^{\hat{v}_n}(\overline{\varphi}_n^\epsilon(i)-\overline{\varphi}_n(i)) \xrightarrow[\epsilon\to 0]{} 0\,. \end{aligned}$$

Hence, inductively, $\overline{\varphi}_0 = \overline{\varphi}$ everywhere, and so

$$\overline{\varphi}_n(i) - \left(1-\alpha^n)(V(i) - \tfrac{\varrho^* + \theta_2}{\theta_1}\right) \;=\; \lim_{\epsilon \downarrow 0} f_n^\epsilon(i) \geq 0\,. \tag{5.21}$$

for all $(n, i) \in \mathbb{N} \times \mathbb{S}$.

From Lemmas 5.4.2 and 5.4.3, we have

$$\overline{\varphi}_n(i) - V(i) \;\leq\; \mathbb{E}_i^{v^*}[\varphi_0(X_n) - V(X_n)]\,,$$

from which we obtain

$$\begin{aligned}
\overline{\varphi}_n(i) \;&\leq\; V(i) + \mathbb{E}_i^{v^*}[\varphi_0(X_n)] \\
&\leq\; V(i) + \mathbb{E}_i^{v^*}[\|\varphi_0\|_V V(X_n)] \\
&\leq\; V(i) + \|\varphi_0\|_V \left(\frac{\varrho^* + \theta_2}{\theta_1} + \alpha^n V(i)\right)\,.
\end{aligned}$$

Combining this inequality with (5.21) yields

$$\begin{aligned}
(1-\alpha^n)(V(i) - \frac{\varrho^* + \theta_2}{\theta_1}) \;&\leq\; \overline{\varphi}_n(i) \\
&\leq\; V(i) + \|\varphi_0\|_V \left(\frac{\varrho^* + \theta_2}{\theta_1} + \alpha^n V(i)\right)\,.
\end{aligned} \tag{5.22}$$

From (5.22), every ω-limit point of $\overline{\Phi}_n[\varphi_0]$ lies in the set

$$G(\varphi_0) \;:=\; \left\{h : \mathbb{S} \to \mathbb{R}, -\frac{\varrho^* + \theta_2}{\theta_1} \leq h - V \leq \|\varphi_0\|_V \frac{\varrho^* + \theta_2}{\theta_1}\right\},$$

and $G(\varphi_0) \subset \mathcal{G}_{-C}$. The ω-limit set is invariant under $\overline{\Phi}_n$, and by Theorem 5.3.1 the only invariant subsets of $\mathcal{G}_{-C}$ are also subsets of $\mathcal{E}$. Thus (5.8) holds, and the rest of the result follows from Lemma 5.4.1. □

Chapter 6

LQG System with Sensor Scheduling and Intermittent Observations

6.1 Introduction

In this chapter, we define a discrete-time linear control system with multiple available sensors that communicate with the controller via an imperfect network channel. New results based on this model will be presented in Chapters 7–8.

Since our model combines elements of two fields of research, we first review the existing work before explicitly defining the system. Technological advances in various areas have led to a number of control applications with distributed, networked sensors, from communications networks [40] to structural health monitoring [12] and even to wearable computing [59]. In such systems, network capacity can cause data packets to be lost, and energy constraints can limits how many or which sensors can transmit observations in each time step. This has led to considerable research into finding optimal scheduling of sensors, as well as into handling randomness in the observation of linear systems.

The field of sensor scheduling, also known as sensor querying, is very

rich, dating back to the 1960s with the seminal work of Meier, et al [42]. In recent years, however, new applications for efficient and robust sensor networks has led to a resurgence of research in optimal sensor scheduling, and has also converged with research on partially-observed Markov decision processes. The problem was developed under the classical MDP optimization framework in [58], which demonstrated that the dynamic programming equations and optimality conditions can be recast in terms of the error covariance via the same separation principle established by [42]. [31] considered a controller that randomly chooses a sensor at each time step, and derived upper and lower bounds on the error covariance. This approach, continued in [45] and others, introduces randomness that allows stochastic approaches to the analysis of convergence and stability. Other research efforts seek computationally feasible methods of calculating optimal or near-optimal control strategies, such as [36], or focused on particular system structures to facilitate analysis [35, 41].

In [56], Sinopoli et al studied a discrete linear system with a single sensor subject to intermittent observations, modeling lost observations as a Bernoulli process with a fixed loss rate λ. The authors show that there is a critical loss rate $\lambda_c \in (0, 1)$ such that the error covariance is sure to remain bounded when $\lambda < \lambda_c$, and sure to diverge for some initial condition when $\lambda \geq \lambda_c$. The framework of [56] has been extended to include more details of the random error covariance behavior [46], weak convergence of the error covariance [37], and extension to more general transmission loss models [49].

In this work, we combine the areas of sensor scheduling and control

with intermittent observations. Our approach to the optimal sensor scheduling problem is inspired by [58], but the intermittent observations introduce another layer of randomness. As in [58], the linear control problem reduces to a Kalman filter and optimal feedback, each computed via a discrete algebraic Riccati equation. However, as in [56], the error covariance is itself stochastic, and we therefore consider the stability of the expected value over time. Optimal policies for an LQG system with two sensors, one of which has perfect transmission while the other is subject to random observation losses, are derived in [32]. Our framework is much more general, considering multiple sensors with different loss rates, combined with a dynamic congestion model that enables complex network behavior. Some limited results for sensor scheduling with intermittency are shown in [44], but the authors do not consider optimal scheduling and control. A special case of our system generalizes the result of [56] to multiple sensors each with a unique loss rate, and show that there is a multi-dimensional critical surface rather than a single critical loss rate.

The following sections describe the detailed system model and some additional notation. We then consider the concept of stability for linear systems with noise, and make a simple assumption on the stabilizability of the system. The Kalman filter is introduced as an optimal estimator, regardless of scheduling scheme or lost observations, and we conclude with some important properties of the stochastic covariance update operator.

6.2 Plant, Observation, and Network Model

We consider a linear quadratic Gaussian (LQG) system

$$\begin{aligned} X_{t+1} &= AX_t + BU_t + DW_t\,, \quad t \geq 0 \\ X_0 &\sim \mathcal{N}(\bar{x}_0, \Sigma_0)\,, \end{aligned} \tag{6.1}$$

where $X_t \in \mathbb{R}^{N_x}$ is the system state, $U_t \in \mathbb{R}^{N_u}$ is the control, and $\{W_t\}$ is the noise process. We assume that each $W_t \sim \mathcal{N}(0, \Sigma_w)$ is i.i.d. and independent of X_0 and that (A, B) is stabilizable. The system is observed via a finite number of sensors scheduled or queried by the controller at each time step. The queried sensor attempts to send information to the controller through the network; depending on the state of the network, the information may be received or lost. This behavior is modeled as

$$Y_t = \gamma_t C_{Q_{t-1}} X_t + F_{Q_{t-1}} W_t\,, \quad t \geq 1, \tag{6.2}$$

with $Y_t \in \mathbb{R}^{N_y}$. The query process $\{Q_t\}$ takes values in the finite set of allowable sensor queries $\mathbb{Q}$, and $\{\gamma_t\}$ is a Bernoulli process indicating if the data is lost in the network: each observation is either received ($\gamma_t = 1$) or lost ($\gamma = 0$). For any allowable query $q \in \mathbb{Q}$, we assume that $\det(F_q F_q^T) \neq 0$ and (primarily to simplify the analysis) that $DF_q^T = 0$. Also without loss of generality, we assume that $\text{rank}(B) = N_u$; if not, we restrict control actions to the row space of B.

The network congestion is modeled as a random process S_t, also controlled by Q_t, taking values on a finite set $\mathbb{S}$ of network states:

$$\mathbb{P}\left(S_{t+1} = s' \mid S_t = s, Q_t = q\right) = p_q(s, s')\,, \quad s, s' \in \mathbb{S}\,, t \geq 0, \tag{6.3}$$

with a known initial state $S_0 = s_0 \in \mathbb{S}$. The observed information is either lost or received according to

$$\mathbb{P}(\gamma_t = 0) = \lambda(S_t, Q_t), \qquad \mathbb{P}(\gamma_t = 1) = (1 - \lambda(S_t, Q_t)), \tag{6.4}$$

where the loss rate $\lambda : \mathbb{S} \times \mathbb{Q} \to [0, 1]$. The network state S_t is assumed to be known by the controller at every time step and, though not necessary for most of the analysis, we assume that the chain $\{S_t\}$ is irreducible and aperiodic.

At each time t, the controller makes a decision $v_t = \{U_t, Q_t\}$, the system state evolves as in (6.1), and the network state transitions according to (6.3). Then the observation at $t + 1$ is either lost or received, determined by (6.2) and (6.4). The decision v_t must be non-anticipative, i.e., should depend only on the history $\mathcal{F}_t$ of observations up to time t defined by

$$\mathcal{F}_t \;:=\; \sigma(S_0, \bar{x}_0, \Sigma_0, S_1, Y_1, \gamma_1, \ldots, S_t, Y_t, \gamma_t).$$

The complete sequence of decisions $v = \{v_t; t \geq 0\}$ is called a policy, and we call the set of admissible policies $\mathcal{V}$.

For an initial condition (S_0, X_0) and a policy $v \in \mathcal{V}$, let $\mathbb{P}^v$ be the unique probability measure on the trajectory space, and $\mathbb{E}^v$ the corresponding expectation operator. When necessary, the explicit dependence on (the law of) X_0 will be denoted as $\mathbb{P}^v_{(S_0,X_0)}$ and $\mathbb{E}^v_{(S_0,X_0)}$.

Let $\mathcal{M}_0^+ \subset \mathbb{R}^{N_x \times N_x}$ be the closed cone of $N_x \times N_x$ symmetric, positive semi-definite matrices. We also define $\mathcal{M}^+ \subset \mathcal{M}_0^+$, the set of $N_x \times N_x$ symmetric, positive definite matrices. For $\Sigma_1, \Sigma_2 \in \mathbb{R}^{N_x \times N_x}$, we say $\Sigma_1 \geq \Sigma_2$ or

$\Sigma_1 > \Sigma_2$ when $\Sigma_1 - \Sigma_2 \in \mathcal{M}_0^+$ or $\Sigma_1 - \Sigma_2 \in \mathcal{M}^+$, respectively. Note that the zero matrix $0 \in \mathcal{M}_0^+$ is the $N_x \times N_x$ matrix with all zero entries, and is the unique "smallest" element of $\mathcal{M}_0^+$, in that

$$\{\Sigma \in \mathcal{M}_0^+ : \Sigma \leq \Sigma' \text{ for all } \Sigma' \in \mathcal{M}_0^+\} = \{0\}.$$

For a square matrix G, let $\sigma(G)$ be the set of eigenvalues of G, and let $\sigma_{min}(G)$ and $\sigma_{max}(G)$ be the eigenvalues with the smallest and largest magnitude, respectively. The trace of a matrix acts as a norm on the cone of positive semidefinite symmetric matrices, and for a matrix $\Sigma \in \mathcal{M}_0^+$, $\text{tr}(\Sigma) = \sum \sigma(\Sigma)$.

6.2.1 Kalman Filtering

Since the system state cannot be observed directly, feedback controls are based on an estimate of the state process. Standard linear estimation theory tells us that the expected value of the state $\hat{X}_t := \mathbb{E}[X_t|\mathcal{F}_t]$ is a sufficient statistic, and can be dynamically calculated via the Kalman filter:

$$\hat{X}_{t+1} = A\hat{X}_t + BU_t + \hat{K}_{Q_t,\gamma_{t+1}}(\hat{\Pi}_t)\big(Y_{t+1} - C_{Q_t}(A\hat{X}_t + BU_t)\big)\,,\ \hat{X}_0 = \bar{x}_0. \quad (6.5)$$

where $\hat{\Pi}$ is the error covariance

$$\hat{\Pi}_t = \text{cov}(X_t - \hat{X}_t) = \mathbb{E}[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T].$$

The Kalman gain $\hat{K}_{q,\gamma}$ is given by

$$\hat{K}_{q,\gamma}(\hat{\Pi}) \ := \ \Xi(\hat{\Pi})\gamma C_q^T\big(\gamma^2 C_q \Xi(\hat{\Pi}) C_q^T + F_q F_q^T\big)^{-1},$$

$$\Xi(\hat{\Pi}) \;:=\; DD^T + A\hat{\Pi}A^T,$$

and the error covariance evolves on $\mathcal{M}_0^+$ as

$$\hat{\Pi}_{t+1} \;=\; \Xi(\hat{\Pi}_t) - \hat{K}_{Q_t,\gamma_{t+1}}(\hat{\Pi}_t)C_{Q_t}\Xi(\hat{\Pi}_t), \qquad \hat{\Pi}_0 = \Sigma_0. \tag{6.6}$$

Note that when an observation is lost ($\gamma_t = 0$), $\hat{K}_{q,\gamma_t} = 0$ and the observer (6.5) simply evolves without any correction factor, and the evolution of $\hat{\Pi}_t$ does not depend on the state control U_t.

6.3 Stability

A well-known necessary condition for stability is that (A, B) is stabilizable and $(A,\overline{C})$ is detectable, where $\overline{C} = [C_1 \mid C_2 \mid \cdots \mid C_{|\mathbb{Q}|}]$. In the absence of intermittency it has been shown in [58] that these conditions are also sufficient. However, with intermittency these conditions are clearly not sufficient. Moreover, algebraic sufficient conditions for stability with intermittent observations do not seem possible, even for a system without sensor scheduling [56].

Suppose that a particular query process $\{Q_t^s\}$ and estimation scheme are known that result in a bounded trajectory of the error covariance matrix. Then it is clear, by the optimality of the Kalman filter, that $\{Q_t^s\}$ together with the Kalman filter estimator will also keep the error covariance bounded. Moreover, since (A, B) is stabilizable then a feedback controller can be designed so that the variance of X stays bounded. Note that there is not strict separation principle in this case, but the partial separation result in [58] makes this possible.

As a result, in this work we will assume that the estimation is stabilizable under some scheduling policy, then investigate the optimal control problem under quadratic running cost.

Assumption 6.3.1. There exists a query process $Q^s = \{Q^s_t : t \geq 0\}$ and a system state estimator for which the error covariance remains bounded. That is, for some initial (x_0, Σ_0)

$$\sup_{t>0} \mathbb{E}^{Q^s}_{x_0,\Sigma_0}\big[\mathrm{tr}(\hat{\Pi}_t)\big] \;<\; \infty\,. \tag{6.7}$$

Without loss of generality, the estimator is the Kalman filter.

Remark 6.3.2. If (A, D) is controllable then (6.7) holds for some (x_0, Σ_0) if and only if the same holds for any initial condition under some policy. Therefore it suffices that (6.7) holds with $(x_0, \Sigma_0) = (0, 0)$. There is also a dichotomy: Unless Assumption 6.3.1 holds, then $\sup_{t>0} \mathbb{E}^{Q}_{x,\Sigma}[\hat{\Pi}_t] = \infty$ for all initial points (x, Σ) and all admissible policies Q. Therefore Assumption 6.3.1 is a necessary condition for long-term average control problem to be well posed.

Remark 6.3.3. If follows by the results of Chapter 7 that, provided (A, D) is controllable, then Assumption 6.3.1 is equivalent to the following seemingly weaker condition: There exists a constant $M > 0$ such that for every $n \in \mathbb{N}$ it holds that

$$\max_{t=1,\dots,n} \mathbb{E}^{Q_n}_{0,0}\big[\mathrm{tr}(\hat{\Pi}_t)\big] \;<\; M$$

for some admissible policy Q_n. Indeed, this condition is all that is required for Lemma 7.4.1 on which the rest of the results are based.

An algebraic characterization of Assumption 6.3.1 based on the parameters of the problem does not seem possible, though results for sensor scheduling without intermittency [58] and intermittent observations with a single sensor [56] suggest an important necessary condition for Assumption 6.3.1. Let $\mathbb{Q} = \{q_1, \ldots, q_{N_q}\}$ and define $\overline{C} := [C_{q_1}^T | \cdots | C_{q_{N_q}}^T]^T$. Then Assumption 6.3.1 holds only if $(\overline{C}, A)$ is detectable. Moreover, as we show later in Corollary 7.6.3, if $(\overline{C}, A)$ is detectable then Assumption 6.3.1 holds for an open set of the parameters λ, and therefore this assumption is generally non-vacuous.

This enables us to derive a wealth of interesting results: (a) Stabilizability leads necessarily to geometric stability; (b) The value iteration algorithm, linking the finite horizon control problem and the infinite horizon ergodic control problem, converges; (c) We extend the seminal result of Sinopoli [56], who showed that there is a stability threshold for the intermittency loss rate, to the sensor scheduling problem with multiple, sensor-dependent loss rates; (d) The analysis and results also facilitate various extensions: in the case of unknown sensor-dependent loss rates, a simple adaptive scheme can be coupled with the estimation that stabilizes the system. Also, when the loss rates depend on the dynamic network congestion (6.3), and adaptive identification scheme as in [5] can be devised which again renders the system stable.

6.3.1 Concavity and Continuity

Recall that a function $f : \mathcal{M}_0^+ \to \mathbb{R}$ is concave if for $\Sigma_1, \Sigma_2 \in \mathcal{M}_0^+$,

$$f((1-\beta)\Sigma_1 + \beta\Sigma_2) \geq (1-\beta)f(\Sigma_1) + \beta f(\Sigma_2), \quad \text{for all } \beta \in [0,1]. \tag{6.8}$$

Concavity for functions on $f : \mathcal{M}_0^+ \to \mathcal{M}_0^+$ is defined in precisely the same way, but replacing the inequality in (6.8) with the ordering defined in Section 6.2. We will sightly abuse the terminology by calling a function f on $\mathbb{S} \times \mathcal{M}_0^+$ concave/continuous/monotone if $f(s, \cdot)$ is concave/continuous/monotone for all $s \in \mathbb{S}$.

For a sensor query $q \in \mathbb{Q}$, we define a function $\mathcal{T}_q : \mathcal{M}_0^+ \to \mathcal{M}_0^+$ by

$$\mathcal{T}_q(\hat{\Pi}) \;:=\; \Xi(\hat{\Pi}) - \hat{K}_{q,1}(\hat{\Pi}) C_q \Xi(\hat{\Pi})$$

and an operator $\tilde{\mathcal{T}}_q$ on functions $f : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}$,

$$\begin{aligned} \tilde{\mathcal{T}}_q f(s, \hat{\Pi}) \;&:=\; \sum_{s' \in \mathbf{S}} p_q(s, s')((1 - \lambda(s, q)) f(s', \mathcal{T}_q(\hat{\Pi})) + \lambda(s, q) f(s', \Xi(\hat{\Pi})) \\ &=\; \mathbb{E}^q \left[f(S_{t+1}, \hat{\Pi}_{t+1}) \;\middle|\; S_t = s \,, \hat{\Pi}_t = \hat{\Pi} \right] . \end{aligned}$$

Lemma 6.3.4. *$\tilde{\mathcal{T}}_q$ preserves concavity and monotonicity for non-decreasing functions.*

Proof. $\Xi(\hat{\Pi})$ is linear in $\hat{\Pi}$, so also concave and non-decreasing. Concavity of $\mathcal{T}_q$ is a standard result (see, e.g., [31, Lemma 1]), as is the fact that $\Sigma \geq \Sigma'$ implies $\mathcal{T}_q(\Sigma) \geq \mathcal{T}_q(\Sigma')$ (e.g, [31, Lemma 2]). Since $\tilde{\mathcal{T}}_q f(s, \hat{\Pi})$ is a convex combination of $f(s', \Xi(\hat{\Pi}))$ and the various possible $f(s', \mathcal{T}_q(\hat{\Pi}))$ functions, if $f : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}$ is concave and non-decreasing in its second argument, so is $\tilde{\mathcal{T}}_q f$. □

Trace is concave and non-decreasing, so $\tilde{\mathcal{T}}_q \operatorname{tr}(\cdot)$ is also. Hence, for any

constant $M > 1$, query $q \in \mathbb{Q}$, and $\hat{\Pi} \in \mathcal{M}_0^+$,

$$\tilde{\mathcal{T}}_q \operatorname{tr}(\hat{\Pi}) \;\geq\; \left(1 - \frac{1}{M}\right) \tilde{\mathcal{T}}_q \operatorname{tr}(0) + \frac{1}{M} \tilde{\mathcal{T}}_q \operatorname{tr}(M\hat{\Pi}).$$

Rearranging and iterating for a sequence of sensor queries $\{q_0, \ldots, q_k\}$ yields

$$\tilde{\mathcal{T}}_k \circ \cdots \circ \tilde{\mathcal{T}}_0 \operatorname{tr}(M\hat{\Pi}) \;\leq\; M\tilde{\mathcal{T}}_k \circ \cdots \circ \tilde{\mathcal{T}}_0 \operatorname{tr}(\hat{\Pi}). \tag{6.9}$$

Let $v_s = \{(Q_t, U_t)\}$ be the stable policy from Assumption 6.3.1, and recall that $\hat{\Pi}_t$ does not depend on the state control $\{U_t\}$. Hence under any admissible policy of the form $\tilde{v} = \{(Q_t, \tilde{U}_t)\}$ for any $\Sigma_0 \in \mathcal{M}_0^+$, (6.9) gives us the following useful bound:

$$\mathbb{E}^{\tilde{v}}[\operatorname{tr}(\hat{\Pi}_t)] \;\leq\; \max\left\{c_1, \frac{c_1}{c_0}\operatorname{tr}(\Sigma_0)\right\} \;\leq\; c_1 + \frac{c_1}{c_0}\operatorname{tr}(\Sigma_0) \tag{6.10}$$

for all $t \geq 0$.

We also get the following straightforward result:

Lemma 6.3.5. *$\tilde{\mathcal{T}}_q$ preserves continuity and lower semi-continuity.*

Proof. Both Ξ and $\mathcal{T}_q$ are continuous by inspection, and so $\tilde{\mathcal{T}}_q$ is a convex combination of continuous functions. Hence $\tilde{\mathcal{T}}_q f$ is continuous when f is continuous. If g is lower semi-continuous, there exists an increasing sequence of continuous functions $f_n \to g$. Each $\tilde{\mathcal{T}} f_n$ is continuous and the increasing sequence $\tilde{\mathcal{T}} f_n \to \tilde{\mathcal{T}} g$, so g is lower semi-continuous. □

Chapter 7

LQG System: Optimal Control

7.1 Introduction

We now formulate and address the optimal control problem for the linear quadratic Gaussian (LQG) system defined in the previous chapter. As with the general Markov decision process (MDP) model from Chapter 2, we introduce a running cost function on the set of states and controls. In this case however, the cost function is assumed to be quadratic in the system state x and control u. (This quadratic cost assumption is the "Q" in LQG.)

Utilizing a partial separation principle and the optimal estimate derived in Section 6.2.1, in this chapter we derive the optimal feedback controller and recast the optimal control problems in terms of the state error covariance and network state only. We show optimality conditions and prove the existence of optimal controls and value functions first for the finite horizon, then for the discounted cost optimization using a receding horizon technique, and lastly for the average cost optimal control problem using a vanishing discount approach. Throughout, the concavity- and continuity-preserving properties of the operator $\tilde{\mathcal{T}}_q$ greatly facilitate the analysis. Finally, we show two results for the reduced model without network state dynamics that generalize the result

from [56] to sensor-dependent observation loss rates.

7.2 Overview of Optimal Control Problems

Much of the following development follows standard patterns; see, for example, [10, 11]. The running cost is made up of a non-negative network cost r_S and a quadratic plant cost r_P:

$$r_S(s,q) + r_P(x,u) = r_S(s,q) + x^T R x + u^T M u\,,$$

where $R, M \in \mathcal{M}^+$. To help with later analysis, we choose one network state to be the network zero state $0 \in \mathbb{S}$

$$0 \in \arg\min_{s\in\mathbb{S}} \Big(\min_{q\in\mathbb{Q}} r_S(s,q)\Big)\,,$$

and without loss of generality assume $\min_{q\in\mathbb{Q}} r_S(0,q) = 1$. We are interested in finding admissible policies in $\mathcal{V}$ that minimize the average cost,

$$J^v := \limsup_{T\to\infty} \frac{1}{T}\,\mathbb{E}^v\left[\sum_{t=0}^{T-1}\big(r_S(S_t,Q_t) + r_P(X_t,U_t)\big)\right].$$

To approach this problem, we will also consider, for $\alpha \in (0,1)$, the α-discounted finite horizon cost

$$J^v_{\alpha,N} := \mathbb{E}^v\left[\sum_{t=0}^{N-1}\alpha^t\big(r_S(S_t,Q_t) + r_P(X_t,U_t)\big) + \alpha^N X_N^T \Pi_{fin} X_N\right], \tag{7.1}$$

where $\Pi_{fin} \in \mathcal{M}_0^+$ is a terminal cost, and the α-discounted cost,

$$J^v_{\alpha} := \mathbb{E}^v\left[\sum_{t=0}^{\infty}\alpha^t\big(r_S(S_t,Q_t) + r_P(X_t,U_t)\big)\right].$$

In each of these problems and throughout the analysis, we assume that $S_0 = s \in \mathbb{S}$ and $X_0 \sim \mathcal{N}(\bar{x}_0, \Sigma_0)$ unless otherwise specified.

Unsurprisingly, in the following sections the α-discounted finite horizon problem will lead to results for the α-discounted problem, which will in turn lead to results for the average cost problem.

7.3 Optimal Control for the Finite Horizon Problem

The optimal control for the finite horizon problem is well understood; details of the following derivations can be found in, for example, [11, Sec. 5.2]. For the finite horizon α-discounted problem with any particular sequence of N sensor queries, the optimal control policy can be derived directly from (7.1), and is given by the linear feedback

$$U_{\alpha,t} = -K_{\alpha,t}\, \mathbb{E}[X_t \mid \mathcal{F}_t]\,, \tag{7.2}$$

with the feedback gain determined using backward recursion:

$$\begin{aligned} K_{\alpha,t} &= \alpha(M + \alpha B^T \Pi_{\alpha,t+1} B)^{-1} B^T \Pi_{\alpha,t+1} A\,, \\ \Pi_{\alpha,t} &= R + \alpha A^T \Pi_{\alpha,t+1} A - \alpha A^T \Pi_{\alpha,t+1} B K_{\alpha,t}\,, \end{aligned} \tag{7.3}$$

with $\Pi_{\alpha,N} = \Pi_{fin}$. However, to facilitate extension to the infinite horizon case, we note that since the system is stabilizable, there exists a unique matrix $\Pi^*_\alpha \in \mathcal{M}^+$ that solves the algebraic Riccati equation

$$\Pi^*_\alpha = R + \alpha A^T \Pi^*_\alpha A - \alpha^2 A^T \Pi^*_\alpha B (M + \alpha B^T \Pi^*_\alpha B)^{-1} B^T \Pi^*_\alpha A\,. \tag{7.4}$$

By setting $\Pi_{fin} = \Pi^*_\alpha$, the backward recursion in (7.3) is t-invariant and, as noted in Section 6.2.1, the expected value of the state can be dynamically

calculated via the Kalman filter estimate $\hat{X}$. So we can define the optimal stationary linear feedback as

$$\begin{aligned} U_t^{\alpha *} &= -K_\alpha^* \hat{X}_t \,, \\ K_\alpha^* &= (M + \alpha B^T \Pi_\alpha^* B)^{-1} \alpha B^T \Pi_\alpha^* A \,. \end{aligned} \tag{7.5}$$

The following result recasts the finite horizon optimal control problem in terms of the error covariance rather than the system state and control.

Theorem 7.3.1. *Let $v^* = \{U_t^{\alpha *}, Q_t^{\alpha *}\}$, where $U_t^{\alpha *}$ is the linear feedback defined in (7.5) and $\{Q_t^{\alpha *}\}$ is a selector from the minimizer in the N-step dynamic programming equation*

$$f_t^{(N)}(s, \hat{\Pi}) = \min_q \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}) + \alpha \tilde{\mathcal{T}}_q f_{t+1}^{(N)}(s, \hat{\Pi}) \right\} \tag{7.6}$$

for $t = 0, \ldots, N-1$ with $f_N^{(N)} = 0$ and $\tilde{\Pi}_\alpha := R - \Pi_\alpha^* + \alpha A^T \Pi_\alpha^* A$.

Then v^* is optimal in that with $\Pi_{fin} = \Pi_\alpha^*$,

$$\begin{aligned} J_{\alpha,N}^{v^*} &= \inf_{v \in \mathcal{V}} J_{\alpha,N}^v \\ &= f_0^{(N)}(s_0, \Sigma_0) + \bar{x}_0^T \Pi_\alpha^* \bar{x}_0 + \mathrm{tr}(\tilde{\Pi}_\alpha \Sigma_0) + \sum_{k=1}^{N} \alpha^k \mathrm{tr}(\Pi_\alpha^* D D^T) \,. \end{aligned} \tag{7.7}$$

Proof. Using the same approach as in [58], we note that the linear feedback (7.2) is optimal relative to $J_{\alpha,N}^v$. That is, for any admissible query sequence $\{Q_t : t \geq 0\}$ and $\tilde{\mathcal{U}}$ the corresponding set of admissible state control policies,

$$\inf_{\tilde{U} \in \tilde{\mathcal{U}}} J_{\alpha,N}^{\tilde{U},Q} = J_{\alpha,N}^{U_\alpha^*,Q} \,.$$

A straightforward calculation gives

$$\begin{aligned}
\mathbb{E}^{U^{\alpha*},Q}[\alpha X_{t+1}^T \Pi_\alpha^* X_{t+1}] &= \mathbb{E}^{U^{\alpha*},Q}[\alpha \hat{X}_{t+1}^T \Pi_\alpha^* \hat{X}_{t+1}] \\
&\quad + \mathbb{E}^{U^{\alpha*},Q}[\alpha (X_{t+1} - \hat{X}_{t+1})^T \Pi_\alpha^* (X_{t+1} - \hat{X}_{t+1})] \\
&= \alpha \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_{t+1}^T \Pi_\alpha^* \hat{X}_{t+1}] + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* \hat{\Pi}_{t+1})] \\
&= \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T (\Pi_\alpha^* - R - K_\alpha^{*T} M K_\alpha^*) \hat{X}_t] \\
&\quad + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* (\Xi(\hat{\Pi}_t) - \hat{\Pi}_{t+1})) + \mathrm{tr}(\Pi_\alpha^* \hat{\Pi}_{t+1})] \\
&= \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T (\Pi_\alpha^* - R - K_\alpha^{*T} M K_\alpha^*) \hat{X}_t] \\
&\quad + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* D D^T) + \mathrm{tr}(\Pi_\alpha^* A \hat{\Pi}_t A^T)].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}^{U^{\alpha*},Q}[r(X_t, U_t^{\alpha*})] &= \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T (R + K_\alpha^{*T} M K_\alpha^*) \hat{X}_t] \\
&\quad + \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(R \hat{\Pi}_t)] \,. \quad (7.8)
\end{aligned}$$

So for $t = 0, \ldots, N-1$,

$$\begin{aligned}
&\mathbb{E}^{U^{\alpha*},Q}[r_P(X_t, U_t^{\alpha*})] + \alpha \mathbb{E}^{U^{\alpha*},Q}[X_{t+1}^T \Pi_\alpha^* X_{t+1}] \\
&\quad = \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T \Pi_\alpha^* \hat{X}_t] + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* D D^T) \\
&\qquad + \mathrm{tr}(\Pi_\alpha^* A \hat{\Pi}_t A^T)] + \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(R \hat{\Pi}_t)] \\
&\quad = \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T \Pi_\alpha^* \hat{X}_t] + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* D D^T)] \\
&\qquad + \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* \hat{\Pi}_t) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_t)]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}^{U^{\alpha*},Q}[\hat{X}_t^T \Pi_\alpha^* \hat{X}_t] + \mathbb{E}^{U^{\alpha*},Q}[(X_t - \hat{X}_t)^T \Pi_\alpha^* (X_t - \hat{X}_t)] \\
&\qquad + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* D D^T)] + \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_t)] \\
&= \mathbb{E}^{U^{\alpha*},Q}[X_t^T \Pi_\alpha^* X_t] + \alpha \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\Pi_\alpha^* D D^T)] + \mathbb{E}^{U^{\alpha*},Q}[\mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_t)] \,.
\end{aligned}
$$

Then iterating backwards yields

$$
\begin{aligned}
J_{\alpha,N}^{U_\alpha^*,Q} &= \bar{x}_0^T \Pi_\alpha^* \bar{x}_0 + \sum_{k=1}^{N} \alpha^k \mathrm{tr}(\Pi_\alpha^* D D^T) \\
&\qquad + \mathbb{E}^{U^{\alpha*},Q}\left[\sum_{t=0}^{N-1} \alpha^t \big(r_S(S_t, Q_t) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_t) \big) \right], \quad (7.9)
\end{aligned}
$$

where the first two terms are clearly independent of the scheduling policy. If we define $f_t^{(N)}$ as the cost-to-go function for

$$
\mathbb{E}^{U^{\alpha*},Q}\left[\sum_{t=0}^{N-1} \alpha^t \big(r_S(S_t, Q_t) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_t) \big) \right],
$$

then the optimal scheduling policy $\{Q_t^{\alpha*}\}$ can be found via (7.6) by dynamic programming. □

7.4 Optimal Control for the α-Discounted Problem

Before proceeding to results about the infinite horizon optimization, we show an essential application of the bound in (6.10):

Lemma 7.4.1. *There exists a positive constant M_s such that with the query process $Q^s = \{Q_t^s : t \geq 0\}$ from Assumption 6.3.1, for any $N > 0$ and $\alpha \in (0,1)$*

$$
J_{\alpha,N}^{v^*} \leq J_{\alpha,N}^{U_\alpha^*,Q^s} \leq M_s \left(\|\bar{x}_0\|^2 + \frac{1}{1-\alpha} + \frac{\mathrm{tr}(\Sigma_0)}{1-\alpha} \right). \quad (7.10)
$$

Proof. Let $\bar{r}_S = \max_{\mathbb{S}\times\mathbb{Q}} r_S$, and from (7.9),

$$\begin{aligned} J_{\alpha,N}^{U_\alpha^*,Q^s} &\leq \sigma_{max}(\Pi_\alpha^*)\|\bar{x}_0\|^2 + \sum_{k=1}^{\infty}\alpha^k \mathrm{tr}(\Pi_\alpha^* DD^T) \\ &\quad + \mathbb{E}^{U^{\alpha *},Q^s}\left[\sum_{t=0}^{\infty}\alpha^t\big(r_S(S_t,Q_t) + \mathrm{tr}(\tilde{\Pi}_\alpha\hat{\Pi}_t)\big)\right] \\ &\leq \sigma_{max}(\Pi_\alpha^*)\|\bar{x}_0\|^2 + \frac{1}{1-\alpha}\mathrm{tr}(\Pi_\alpha^* DD^T) \\ &\quad + \frac{1}{1-\alpha}\bar{r}_S + \frac{1}{1-\alpha}\sigma_{max}(\tilde{\Pi}_\alpha)\left(c_1 + \frac{c_1}{c_0}\mathrm{tr}(\Sigma_0)\right). \end{aligned}$$

Define

$$M_s := \max\left\{\sigma_{max}(\Pi_\alpha^*), (\mathrm{tr}(\Pi_\alpha^* DD^T) + \bar{r}_S + c_1\sigma_{max}(\tilde{\Pi}_\alpha)), \frac{c_1}{c_0}\sigma_{max}(\tilde{\Pi}_\alpha)\right\},$$

and recalling that v^* is the policy that minimizes $J_{\alpha,N}^v$, the result follows. □

Once again, we can recast the optimal control problem in terms of the error covariance rather than the state and control processes. In the infinite horizon case, this leads to a modified discounted optimality equation.

Theorem 7.4.2. *For $\alpha \in (0,1)$, there exists a unique lower semicontinuous function $f_\alpha^* : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{R}_+$ that satisfies*

$$f_\alpha^*(s,\hat{\Pi}) = \min_q\{r_S(s,q) + \mathrm{tr}(\tilde{\Pi}_\alpha\hat{\Pi}) + \alpha\tilde{\mathcal{T}}_q f_\alpha^*(s,\hat{\Pi})\}, \tag{7.11}$$

with $\tilde{\Pi}_\alpha := R - \Pi_\alpha^* + \alpha A^T\Pi_\alpha^* A$. If $q_\alpha^* : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{Q}$ is a selector of the minimizer in (7.11), then the policy given by $v^* = (q_\alpha^*(S_t,\hat{\Pi}_t), U_t^{\alpha *})$ for $t \geq 0$ is optimal in the sense that $J_\alpha^{v^*} = \inf_{v\in\mathcal{V}} J_\alpha^v$, and

$$J_\alpha^{v^*} = f_\alpha^*(s_0,\Sigma_0) + \bar{x}_0^T\Pi_\alpha^*\bar{x}_0 + \mathrm{tr}(\tilde{\Pi}_\alpha\Sigma_0) + \frac{\alpha}{1-\alpha}\mathrm{tr}(\Pi_\alpha^* DD^T). \tag{7.12}$$

Further, the querying component of any optimal stationary Markov policy is an a.e. selector of the minimizer in (7.11).

Proof. First, note that from (7.6), thanks to the choice of $\Pi_{fin} = \Pi^*_\alpha$,

$$f_0^{(N+1)}(s,\hat{\Pi}) \;=\; \min_q \{r_S(s,q) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}) + \alpha \tilde{\mathcal{T}}_q f_0^{(N)}(s,\hat{\Pi})\}, \tag{7.13}$$

with $f_0^{(0)} = 0$. Let v_N^* be an optimal policy for the N-step optimization from Theorem 7.3.1, and let $\bar{v}_N = \{U_t^{\alpha *}, Q_t\}$ be the optimal feedback policy (7.2) with the scheduling policy from Assumption 6.3.1. From (7.10) with (7.7), $\{f_0^{(N)}\}$ are bounded pointwise in $\mathbb{S} \times \mathcal{M}_0^+$. Since they are also monotonically increasing in N, $f_0^{(N)} \uparrow f_\alpha^*$ for some lower semicontinuous $f_\alpha^* : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}_+$. Taking monotone limits in (7.13) implies (7.11), and similarly in (7.7) yields (7.12).

Consider the structure of (7.13). Trace is non-decreasing and concave, and the minimum of concave, non-decreasing functions is also concave and non-decreasing. $\tilde{\mathcal{T}}_q$ preserves concavity for non-decreasing functions, so for any $s \in \mathbb{S}$, initializing (7.13) with a non-decreasing and concave function (e.g., $f_0^{(0)} = 0$) guarantees that $f_\alpha^*(s,\cdot)$ is non-decreasing and concave.

Let q_α^* be the selector from the minimizer in (7.11), and using (7.8),

$$J_{\alpha N}^{v_N^*} \geq \mathbb{E}^{q_\alpha^*}\left[\sum_{t=0}^{N-1} \alpha^t r_P(X_t, U_t^{\alpha *})\right] \geq \sigma_{max}(R) \sum_{t=0}^{N-1} \alpha^t \mathbb{E}^{q_\alpha^*}\big[\mathrm{tr}(\hat{\Pi}_t)\big].$$

Since we know that $\lim_{N\to\infty} J_{\alpha N}^{v_N^*} < \infty$, it follows that $\alpha^t \mathbb{E}^{q_\alpha^*}[\mathrm{tr}(\hat{\Pi}_t)] \to 0$ as $t \to \infty$. Then the structure of $J_\alpha^{v^*}$ in (7.12) with the estimate in (7.10) imply

$\alpha^t \mathbb{E}^{q^*_\alpha}[f^*_\alpha(S_t, \hat{\Pi}_t)] \to 0$ as $t \to \infty$. Iterating (7.11) with the selector q^*_α yields

$$f^*_\alpha(s_0, \Sigma_0) \;=\; \mathbb{E}^{q^*_\alpha}_{(s_0,X_0)} \left[\sum_{k=0}^{t-1} \alpha^k \big(r_S(S_k, Q_k) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_k) \big) \right] + \alpha^t \mathbb{E}^{q^*_\alpha}_{(s_0,X_0)} \left[f^*_\alpha(S_t, \hat{\Pi}_t) \right],$$

and letting $t \to \infty$ leaves

$$f^*_\alpha(s_0, \Sigma_0) \;=\; \mathbb{E}^{q^*_\alpha}_{(s_0,X_0)} \left[\sum_{k=0}^{\infty} \alpha^k \big(r_S(S_k, Q_k) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_k) \big) \right].$$

Finally, for any other $v \in \mathcal{V}$ with $J^v_\alpha < \infty$, iterating (7.11) with v yields

$$f^*_\alpha(s_0, \Sigma_0) \;\leq\; \mathbb{E}^{v}_{(s_0,X_0)} \left[\sum_{k=0}^{\infty} \alpha^k \big(r_S(S_k, Q_k) + \mathrm{tr}(\tilde{\Pi}_\alpha \hat{\Pi}_k) \big) \right], \tag{7.14}$$

and so the structure of (7.12) implies q^*_α is optimal and f^*_α is unique. Any optimal policy v can equivalently utilize the optimal feedback $U^{\alpha *}_t$, so consider a stationary Markov policy $v = \{U^{\alpha *}_t, Q_t\}$. If Q_t is not an a.e. selector of the minimizer in (7.11), then the inequality in (7.14) is strict and v cannot be optimal. □

7.5 Optimal Control for the Average Cost Problem

Now we can proceed to the average cost problem. We adopt vanishing discount approach, using uniform properties of the discounted-cost value functions proved in the following sections.

Let $\mathcal{M}^+_\epsilon := \{\hat{\Pi} \in \mathcal{M}^+ : \sigma_{min}(\hat{\Pi}) > \epsilon\}$, and for a constant $c > 0$, define a closed ball $\mathcal{B}_c \subset \mathcal{M}^+_0$ as $\mathcal{B}_c := \{\Sigma \in \mathcal{M}^+_0 : \mathrm{tr}(\Sigma) \leq c\}$.

Lemma 7.5.1. *With (A, D) controllable, there exists an $\epsilon > 0$ such that for any query sequence $\{q_0, \ldots, q_{N_x-1}\} \in \mathbb{Q}^{N_x}$,*

$$\mathbb{P}^{\{q_0,\ldots,q_{N_x-1}\}}(\hat{\Pi}_{N_x} \in \mathcal{M}_\epsilon^+ \mid \Sigma_0 = 0) = 1\,.$$

Proof. Adapting the result from [58, Lemma 3.5], we can rewrite (6.6) as

$$\hat{\Pi}_{t+1} \;=\; \Xi(\hat{\Pi}_t) - \hat{K}_{Q_t,\gamma_{t+1}}(\hat{\Pi}_t)\big(C_{Q_t}\Xi(\hat{\Pi})C_{Q_t}^T + F_{Q_t}F_{Q_t}^T\big)\hat{K}_{Q_t,\gamma_{t+1}}^T(\hat{\Pi}_t).$$

Consider an update when $\gamma_{t+1} = 1$: $F_q F_q^T$ is positive definite for any $q \in \mathbb{Q}$, which means $z \in \ker(\hat{\Pi}_{t+1})$ only if $z \in \ker(\hat{K}_{Q_t,1}(\hat{\Pi}))$ and $z \in \ker(\Xi(\hat{\Pi}_t))$. However, from the definition of $\hat{K}_{q,1}$, $\ker(\hat{K}_{Q_t,1}(\hat{\Pi}_t)) \subset \ker(\Xi(\hat{\Pi}_t))$, and therefore $\ker(\hat{\Pi}_{t+1}) = \ker(\Xi(\hat{\Pi}_t))$. On the other hand when $\gamma_{t+1} = 0$, $\hat{\Pi}_{t+1} = \Xi(\hat{\Pi}_t)$, so whether the observation is lost or received,

$$\ker(\hat{\Pi}_{t+1}) = \ker(\Xi(\hat{\Pi}_t)) = \ker(\hat{\Pi}_t A^T) \cap \ker(D^T).$$

Hence, along any fixed N_x-step query sequence $\{q_0, \ldots, q_{N-1}\}$, if $\hat{\Pi}_t = 0$,

$$\ker(\hat{\Pi}_{t+N_x}) \;=\; \ker(D^T) \cap \ker(D^T A^T) \cap \cdots \cap \ker\big(D^T (A^T)^{N_x-1}\big)\,.$$

Since (A, D) is controllable, $\ker(\hat{\Pi}_{t+N_x}) = \{0\}$, so whether observations are lost or received, the process noise drives the error covariance into the interior of $\mathcal{M}_0^+$. Since there are only finitely many possible N_x-step query sequences and finitely many network states, we can choose ϵ to be the minimal eigenvalue of $\hat{\Pi}_{t+N_x}$ over the possible query and state sequence combinations. □

Note that in the proof of Theorem 7.4.2 we showed that $f^*_\alpha(s,\cdot)$ is non-decreasing, so $\inf_{\Sigma\in\mathcal{M}_0^+} f^*_\alpha(s,\Sigma) = f^*_\alpha(s,0)$. We define

$$\bar{f}_\alpha := f^*_\alpha - f^*_\alpha(0,0)\,,$$

and for a set $\mathcal{B}\in\mathcal{M}_0^+$,

$$\begin{aligned}\operatorname*{span}_{\mathcal{B}}(f^*_\alpha(s,\cdot)) &:= \sup_{\Sigma\in\mathcal{B}} f^*_\alpha(s,\Sigma) - \inf_{\Sigma\in\mathcal{B}} f^*_\alpha(s,\Sigma)\,,\\ \operatorname*{span}_{\mathbb{S}\times\mathcal{B}}(f^*_\alpha) &:= \sup_{s\in\mathbb{S},\Sigma\in\mathcal{B}} f^*_\alpha(s,\Sigma) - \inf_{s\in\mathbb{S},\Sigma\in\mathcal{B}} f^*_\alpha(s,\Sigma).\end{aligned}$$

Lemma 7.5.2. *The differential discounted value function $\bar{f}_\alpha$ is locally bounded, uniformly in $\alpha\in(0,1)$, and $\{\bar{f}_\alpha : \alpha\in(0,1)\}$ is locally Lipschitz equicontinuous on compact subsets of $\mathcal{M}_0^+$.*

Proof. Choose a constant $\bar{c}$ such that $\bar{c}\geq M_s$, the constant from Lemma 7.4.1, and $\mathbb{P}(\hat{\Pi}_{N_x}\in\mathcal{B}_{\bar{c}}|\hat{\Pi}_0=0)=1$ (which is possible because there are only finitely many state/query/γ sequences of length N_x). Fix an $s\in\mathbb{S}$, and with ϵ from Lemma 7.5.1, let $\Sigma^*_\alpha\in\mathcal{B}_{\bar{c}}$ such that

$$f^*_\alpha(s,\Sigma^*_\alpha)\geq \sup_{\mathcal{B}_{\bar{c}}} f^*_\alpha(s,\cdot) - \epsilon.$$

For an α-optimal policy q^*_α we have

$$\begin{aligned}f^*_\alpha(s,0) &= \mathbb{E}^{q^*_\alpha}_{s,0}\left[\sum_{t=0}^{N_x-1}\alpha^t(r_S(S_t,Q_t)+\operatorname{tr}(\Pi^*_\alpha\hat{\Pi}_t)+\alpha^{N_x}f^*_\alpha(S_{N_x},\hat{\Pi}_{N_x})\right]\\ &\geq \alpha^{N_x}\mathbb{E}^{q^*_\alpha}_{s,0}\left[f^*_\alpha(S_{N_x},\hat{\Pi}_{N_x})\right].\end{aligned}$$

Thus,

$$\begin{aligned}
\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) &\le f^*_\alpha(s,\Sigma^*_\alpha) - f^*_\alpha(s,0) + \epsilon \\
&\le f^*_\alpha(s,\Sigma^*_\alpha) - \alpha^{N_x}\mathbb{E}^{q^*_\alpha}_{s,0}\left[f^*_\alpha(S_{N_x},\hat{\Pi}_{N_x})\right] + \epsilon \\
&= (1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) \\
&\qquad + \alpha^{N_x}\mathbb{E}^{q^*_\alpha}_{s,0}\left[f^*_\alpha(s,\Sigma^*_\alpha) - f^*_\alpha(S_{N_x},\hat{\Pi}_{N_x})\right] + \epsilon \\
&\le (1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) + \alpha^{N_x}\left(\sup_{\mathcal{B}_{\bar{c}}} f^*_\alpha(s,\cdot) - f^*_\alpha(s,\epsilon I)\right) + \epsilon \\
&\le (1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) + \alpha^{N_x}\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) \\
&\qquad - \alpha^{N_x}\left(f^*_\alpha(s,\epsilon I) - f^*_\alpha(s,0)\right) + \epsilon \\
&\le (1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) + \alpha^{N_x}\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) \\
&\qquad - \alpha^{N_x}\frac{\epsilon}{\bar{c}}\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) + \epsilon \\
&\le (1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) + \alpha^{N_x}(1-\epsilon/\bar{c})\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) + \epsilon\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\operatorname*{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot)) &\le \frac{(1-\alpha^{N_x})f^*_\alpha(s,\Sigma^*_\alpha) + \epsilon}{1-\alpha^{N_x}(1-\epsilon/\bar{c})} \\
&\le \frac{(1+\alpha+\alpha^2+\cdots+\alpha^{N_x-1})(1-\alpha)f^*_\alpha(s,\Sigma^*_\alpha) + \epsilon}{\epsilon/\bar{c}} \\
&\le \frac{N_x\bar{c}}{\epsilon}(1-\alpha)f^*_\alpha(s,\Sigma^*_\alpha) + \bar{c}\,.
\end{aligned}$$

Since, by (7.10) and (7.12), $(1-\alpha)f^*_\alpha$ is bounded uniformly in α, the same is true of $\operatorname{span}_{\mathcal{B}_{\bar{c}}}(f^*_\alpha(s,\cdot))$, and since there are only finitely many states,

$\text{span}_{\mathbb{S}\times\mathcal{B}_{\bar{c}}}(f^*_\alpha)$ is also bounded uniformly in α. Define

$$m_1 := \max_{s\in\mathbb{S}} \operatorname*{span}_{\mathcal{B}_{\bar{c}}} f^*_\alpha(s,\cdot)\,.$$

Now consider $\Sigma \in \mathcal{M}_0^+$ such that $\text{tr}(\Sigma) \geq \bar{c}$. Clearly, $\Sigma' := \left(\frac{\bar{c}}{\text{tr}(\Sigma)}\Sigma\right) \in \mathcal{B}_{\bar{c}}$. Using the concavity of f^*_α we obtain

$$\begin{aligned} f^*_\alpha(s,\Sigma') &= f^*_\alpha\left(s, \tfrac{\bar{c}}{\text{tr}(\Sigma)}\Sigma + \big(1-\tfrac{\bar{c}}{\text{tr}(\Sigma)}\big)0\right) \\ &\geq \tfrac{\bar{c}}{\text{tr}(\Sigma)} f^*_\alpha(s,\Sigma) + \big(1-\tfrac{\bar{c}}{\text{tr}(\Sigma)}\big) f^*_\alpha(s,0)\,, \end{aligned}$$

and therefore, we have

$$f^*_\alpha(s,\Sigma) - f^*_\alpha(s,0) \;\leq\; \frac{\text{tr}(\Sigma)}{\bar{c}}\left(f^*_\alpha(s,\Sigma') - f^*_\alpha(s,0)\right)\,.$$

Hence, for any $\Sigma \in \mathcal{M}_0^+$,

$$f^*_\alpha(s,\Sigma) - f^*_\alpha(s,0) \;\leq\; \operatorname*{span}_{\mathcal{B}_{\bar{c}}} f^*_\alpha(s,\cdot)\left(1 + \frac{\text{tr}(\Sigma)}{\bar{c}}\right).$$

Let $m_0 := \max_{s\in\mathbb{S}}(f^*_\alpha(s,0) - f^*_\alpha(0,0))$. Then

$$\begin{aligned} \bar{f}_\alpha(s,\Sigma) &= f^*_\alpha(s,\Sigma) - f^*_\alpha(0,0) \\ &\leq f^*_\alpha(s,\Sigma) - f^*_\alpha(s,0) + f^*_\alpha(s,0) - f^*_\alpha(0,0) \\ &\leq \frac{m_1}{\bar{c}}\text{tr}(\Sigma) + (m_1 + m_0)\,. \end{aligned} \tag{7.15}$$

The function $\bar{f}_\alpha$ inherits concavity from f^*_α, so the bound in (7.15) implies Lipschitz equicontinuity of $\{\bar{f}\}$ on bounded subsets of $\mathbb{S}\times\mathcal{M}_\epsilon^+$ [48, Theorem 10.6]. Fix an initial $(s,\Sigma) \in \mathbb{S}\times\mathcal{M}_0^+$, and let $\mathbf{q} = \{q_0,\ldots,q_{N_x}\}$ be

the first $N_x + 1$ queries from the α-discounted optimal control; i.e., selectors from the minimizer in (7.11). For $k = 0, \ldots, N_x$, define $\tilde{\mathcal{T}}_{\mathbf{q}_k} = \tilde{\mathcal{T}}_{q_k} \circ \cdots \circ \tilde{\mathcal{T}}_{q_0}$, and let $\Sigma' \in \mathcal{M}_0^+$. Iterative applications of (7.11) yield

$$f_\alpha^*(s, \Sigma') - f_\alpha^*(s, \Sigma) \;\leq\; \operatorname{tr}(\tilde{\Pi}_\alpha^*(\Sigma' - \Sigma)) + \sum_{k=1}^{N_x - 1} \alpha^k \tilde{\mathcal{T}}_{\mathbf{q}_k} \operatorname{tr}(\tilde{\Pi}_\alpha^*(\Sigma' - \Sigma)) \\ + \alpha^{N_x} \big(\tilde{\mathcal{T}}_{\mathbf{q}_{N_x}} f_\alpha^*(s, \Sigma') - \tilde{\mathcal{T}}_{\mathbf{q}_{N_x}} f_\alpha^*(s, \Sigma) \big) . \quad (7.16)$$

Each $\tilde{\mathcal{T}}_{\mathbf{q}_k}$ preserves continuity in $\mathcal{M}_0^+$, and the order-preserving property of $\tilde{\mathcal{T}}_q$ guarantees that for any $\Sigma' \in \mathcal{M}_0^+$, $\hat{\Pi}_{N_x} \in \mathcal{M}_\epsilon^+$ with probability 1 for the constant ϵ from Lemma 7.5.1. $\bar{f}_\alpha(s, \cdot)$ is equicontinuous on bounded subsets of $\mathcal{M}_\epsilon^+$, so (7.16) implies $\bar{f}_\alpha(s, \cdot)$ must be equicontinuous on bounded subsets of $\mathcal{M}_0^+$. Again noting that there are finitely many states and query combinations, we can take the maximal Lipschitz constant for a particular compact set in $\mathcal{M}_0^+$. □

Theorem 7.5.3. *There exists a continuous function $f^* : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}_+$ and a constant ϱ^* that satisfy*

$$f^*(s, \hat{\Pi}) + \varrho^* = \min_q \left\{ r_S(s, q) + \operatorname{tr}(\tilde{\Pi}^* \hat{\Pi}) + \tilde{\mathcal{T}}_q f^*(s, \hat{\Pi}) \right\}, \quad (7.17)$$

with $\tilde{\Pi}^* := R - \Pi^* + A^T \Pi^* A$ and $\Pi^* \in \mathcal{M}^+$ the unique solution of the algebraic Riccati equation

$$\Pi^* = R + A^T \Pi^* A - A^T \Pi^* B (M + B^T \Pi^* B)^{-1} B^T \Pi^* A . \quad (7.18)$$

If $q^* : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{Q}$ is a selector of the minimizer in (7.17), then the policy

given by $v^* = (U_t^*, q^*(S_t, \hat{\Pi}_t)$ *for* $t \geq 0$ *with*

$$\begin{aligned} U_t^* &:= -K^* \hat{X}_t \,, \\ K^* &:= (M + B^T \Pi^* B)^{-1} B^T \Pi^* A \,, \end{aligned} \tag{7.19}$$

is optimal in the sense that $J^{v^*} = \inf_{v \in \mathcal{V}} J^v$*, and*

$$J^{v^*} = \varrho^* + \mathrm{tr}(\Pi^* D D^T) \,.$$

Further, the querying component of any optimal stationary Markov policy is an a.e. selector of the minimizer in (7.17)*.*

Proof. Since the system is stabilizable, the Riccati equation (7.4) converges as $\alpha \to 1$ to (7.18) which has a unique solution $\Pi^* \in \mathcal{M}^+$. The feedback given by (7.19) is then optimal for any given querying sequence, and we only need consider optimal sensor scheduling.

The collection $\{\bar{f}_\alpha\}$ is locally Lipschitz equicontinuous and bounded, so (repeatedly dropping to subsequences as needed) along some sequence $\alpha_k \to 1$, each $\bar{f}_{\alpha_k}(s, \cdot)$ converges to some continuous function $\bar{h}(s, \cdot)$ and $(1-\alpha) f_\alpha^*(s, 0)$ converges to a positive constant $\varrho(s)$.

Letting $f^*(s, \hat{\Pi}) = \bar{h}(s, \hat{\Pi}) + \varrho(s) - \varrho(0)$ and $\varrho^* = \varrho(0)$, we get

$$f_{\alpha_k}^*(s, \hat{\Pi}) \xrightarrow[k \to \infty]{} f^*(s, \hat{\Pi}) + \varrho^*,$$

and taking limits in (7.11) yields (7.17). With q^* a selector of the minimizer in (7.17), since the network running cost is bounded above and using (7.15), there exist constants M_0, M_1 with $M_1 > 0$ such that for all $(s, \Sigma) \in \mathbb{S} \times \mathcal{M}_0^+$,

$$\tilde{\mathcal{T}}_{q^*} f^*(s, \Sigma) - f^*(s, \Sigma) \;=\; -r_S(s, q^*) - \mathrm{tr}(\tilde{\Pi}^* \Sigma) + \varrho^*$$

$$\leq \; -M_1 f^*(s,\Sigma) + M_0\,. \tag{7.20}$$

Note that if f^* solves (7.17), so does $f^* + c$ for any constant c, and that (7.20) still holds with $M_0 \to (M_0 + M_1 c)$. Therefore, without loss of generality, we assume that

$$\min_{\mathbb{S}\times\mathcal{M}_0^+} f^* = 1.$$

The bound in (7.20) implies the geometric drift condition [43, (V4)], so the chain is geometrically ergodic and $\sup_{t\geq 0} \mathbb{E}^{q^*}_{s_0,X_0}[\mathrm{tr}(\hat{\Pi}_t)] < \infty$ for all (s_0, X_0). With K^* from (7.19), $(A - BK^*)$ is stable, so from the closed-loop state dynamics

$$X_{t+1} = (A - BK^*)X_t + BK^*(X_t - \hat{X}_t) + DW_t\,,$$

the system is stable under (q^*, U^*).

To show optimality, let $\{Q_t\}$ be any admissible querying sequence. Iterating (7.17),

$$\varrho^* + \frac{f^*(s_0,\Sigma_0) - \mathbb{E}^{Q_t}_{s_0,X_0}[f^*(S_N,\hat{\Pi}_N)]}{N} \\ \leq \; \frac{1}{N}\mathbb{E}^{Q_t}_{s_0,X_0}\left[\sum_{t=0}^{N-1} r_s(S_t,Q_t) + \mathrm{tr}(\tilde{\Pi}^*\hat{\Pi}_t)\right], \tag{7.21}$$

with equality if $Q_t = q^*$. Since the covariance $\hat{\Pi}_t$ is stable, using (7.15) we have,

$$\frac{\mathbb{E}^{Q_t}_{s_0,X_0}[f^*(S_N,\hat{\Pi}_N)]}{N} \xrightarrow[N\to\infty]{} 0\,,$$

so taking limits on both sides of (7.21) yields

$$\varrho^* \;\leq\; \limsup_{N\to\infty} \frac{1}{N}\mathbb{E}^{Q_t}_{s_0,X_0}\left[\sum_{t=0}^{N-1} r_s(S_t,Q_t) + \mathrm{tr}(\tilde{\Pi}^*\hat{\Pi}_t)\right], \qquad \mathbb{P}^{q^*}_{s_0,X_0}\text{– a.s.}$$

Indeed, for any policy $v \in \mathcal{V}$ such that the limit supremum of the r.h.s. of (7.21) is finite, we have $\frac{1}{N_k}\mathbb{E}^v_{s_0,X_0}[f^*(s_0,\hat{\Pi}_{N_k})] \to 0$ along some subsequence $N_k \to \infty$, and so

$$\liminf_{n\to\infty} \frac{\mathbb{E}^v_{s_0,X_0}[f^*(S_n,\hat{\Pi}_n)]}{n} = 0 \qquad \mathbb{P}^v_{s_0,X_0}\text{- a.s..}$$

Combining the above, for any $v \in \mathcal{V}$,

$$\varrho^* \;\leq\; \limsup_{N\to\infty} \frac{1}{N}\,\mathbb{E}^v_{s_0,X_0}\left[\sum_{t=0}^{N-1} r_s(S_t,Q_t) + \mathrm{tr}(\tilde{\Pi}^*\hat{\Pi}_t)\right], \tag{7.22}$$

and q^* is optimal. As in the discounted case, any policy with a query process that is not an a.e. selector of the minimizer in (7.17) induces a strict inequality in (7.22), and therefore such a policy cannot be optimal. □

Remark 7.5.4. It is worth noting that f^* is concave and non-decreasing in $\mathcal{M}_0^+$, and that using (7.15) and the definition of f^*, there exist constants $m_1^* > 0$ and $m_0^* \in \mathbb{R}$ such that

$$f^*(s,\Sigma) \leq m_1^*\,\mathrm{tr}(\Sigma) + m_0^*\,. \tag{7.23}$$

Furthermore, directly from (7.17),

$$f^*(s,\Sigma) \geq \sigma_{min}(\tilde{\Pi}^*)\mathrm{tr}(\Sigma) - \varrho^*\,,$$

so f^* must be strictly increasing in Σ.

Remark 7.5.5. For computational purposes, the complete cone $\mathcal{M}_0^+$ is clearly impractical. However, the following result shows that we can approximate the

process on a bounded subset $\mathcal{B}_R = \mathbb{S} \times \{\Sigma \in \mathcal{M}_0^+ : \mathrm{tr}(\Sigma) \leq R\}$ for $R > 0$. The truncated, approximate value function restricted to $\mathcal{B}_R$ solves

$$f^R(s,\Sigma) + \varrho^R = \min_{q \in \mathbb{Q}} \{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^* \Sigma) + \tilde{\mathcal{T}}_q^R f^R(s,\Sigma) \},$$

for $(s,\Sigma) \in \mathcal{B}_R$. We extend f^R on $\mathcal{B}_R^c$ with a known function that is the same order as the true value function, namely, $\mathrm{tr}(\tilde{\Pi}^* \cdot)$.

Let q^R be a measurable selector of the minimizer on $\mathcal{B}_R$ and any fixed, stable control on $\mathcal{B}_R^c$. f^R again satisfies the geometric drift condition, so the process under q^R is stable. It can be shown that as $R \to \infty$, $\varrho^R \to \varrho^*$, and so the truncated system is a good approximation of the complete system.

7.6 A Special Case: Sensor-Dependent Loss Rates

We now turn our attention to a special case of the previous results, with a single network state. In this case, the network cost is simply a function of the query process $\{Q_t\}$, taking values in the finite set of allowable sensor queries $\mathbb{Q} = \{q_1, \ldots, q_{N_q}\}$. The loss rate depends only on the query, as

$$\mathbb{P}(\gamma = 1) = (1 - \lambda_q), \qquad \mathbb{P}(\gamma = 0) = \lambda_q, \tag{7.24}$$

where the loss rate $\lambda = [\lambda_1, \ldots, \lambda_{N_q}]^T$ is vector in $[0,1]^{N_q}$. For two vectors $\lambda, \phi \in \mathbb{R}^N$, we say $\lambda \leq \phi$ if $\lambda_i \leq \phi_i$ for each $i \in \{1, \ldots, N\}$, and $\lambda < \phi$ if $\lambda_i < \phi_i$ for each $i \in \{1, \ldots, N\}$.

We are interested in characterizing the set of loss rates $\Lambda_s \subset [0,1]^{N_q}$ for which the system is stabilizable. Our formulation generalizes the problem in

[56], which analyzes the system (6.1)–(6.2) without sensor scheduling ($C_q = C$) and with a uniform loss rate ($\lambda_q = \bar{\lambda}$). The authors prove that there is a critical loss rate $\lambda_c \in (0, 1)$ such that the system is stabilizable if and only if $\bar{\lambda} < \lambda_c$ (i.e., $\Lambda_s = [0, \lambda_c)$. Here, we generalize that result, showing that when selecting different sensors induces different loss rates, there is a critical surface $\mathcal{W} \subset [0, 1]^{N_q}$. The system is stabilizable if and only if the vector $\lambda < \lambda' \in \mathcal{W}$. We also present a numerical example illustrating the critical surface.

7.6.1 Main Results

Recalling the discussion around Assumption 6.3.1, $\Lambda_s = \varnothing$ unless (A, B) is stabilizable and $(\overline{C}, A)$ is detectable. Hence, without loss of generality, we assume (A, B) is stabilizable and $(\overline{C}, A)$ is detectable and therefore, by the results in [58], $0 \in \Lambda_s$.

Theorem 7.6.1. *If the system* (6.1)–(6.2) *with* (7.24) *is stabilizable for a loss rate* $\lambda \in [0, 1]^{N_q}$*, then it is also stabilizable for any other loss rate* $\lambda \leq \lambda'$*. In other words, the set* Λ_s *is order-convex with respect to the natural ordering of positive vectors in* $\mathbb{R}^{N_q}$*.*

Proof. In order to distinguish between operations with different loss rates, we will indicate the corresponding rate in a superscript, as in

$$\tilde{\mathcal{T}}_q^{\lambda} f(\Sigma) = (1 - \lambda_q) f(\mathcal{T}_q(\Sigma)) + \lambda_q f(\Xi(\Sigma)).$$

Suppose that the system (6.1)–(6.2) with (7.24) is stabilizable for an loss rate $\lambda' \in [0, 1]^{N_q}$, and let $\{Q_t\}$ be a stabilizing query sequence. Let $\lambda \in [0, 1]^{N_q}$

such that $\lambda \leq \lambda'$. For a non-decreasing function $f : \mathcal{M}_0^+ \to \mathbb{R}$ and any $q \in \mathbb{Q}$,

$$\tilde{\mathcal{T}}_q^{\lambda} f(\Sigma) - \tilde{\mathcal{T}}_q^{\lambda'} f(\Sigma) = (\lambda'_q - \lambda_q)(f(\mathcal{T}_q(\Sigma)) - f(\Xi(\Sigma))) \leq 0.$$

Applying to $\mathrm{tr}(\cdot)$, which is non-decreasing in $\mathcal{M}_0^+$, we get

$$\tilde{\mathcal{T}}_q^{\lambda} \mathrm{tr}(\Sigma) - \tilde{\mathcal{T}}_q^{\lambda'} \mathrm{tr}(\Sigma) = -(\lambda'_q - \lambda_q)\mathrm{tr}(\hat{K}_{q,1}(\Sigma) C_q \Xi(\Sigma)) \leq 0 \tag{7.25}$$

because $\hat{K}_{q,1}(\Sigma) C_q \Xi(\Sigma) \in \mathcal{M}_0^+$. Iterating (7.25) with the stabilizing query sequence yields

$$\mathbb{E}_{\Sigma}^{Q_t,\lambda} \|X_t - \hat{X}_t\|^2 \leq \mathbb{E}_{\Sigma}^{Q_t,\lambda'} \|X_t - \hat{X}_t\|^2, \quad \text{for all } t \geq 0,$$

and stability with λ follows. □

Moreover, a lower loss rate leads to a smaller error covariance at every time step. Another important result is the following:

Theorem 7.6.2. *If the system* (6.1)–(6.2) *with* (7.24) *is stabilizable for a loss rate* $\lambda \in [0,1]^{N_q}$*, there exists an open neighborhood* $\mathcal{B} \subset [0,1]^{N_q}$ *around* λ *such that the system is stabilizable for* $\lambda' \in \mathcal{B}$*.*

Proof. Let $\lambda \in [0,1]^{N_q}$ and assume the system is stabilizable for λ. Also let f^* and q^* be the solution and selector from the minimizer of (7.17). Let $\lambda' > \lambda$ such that

$$\lambda'_q - \lambda_q < \frac{m_1 \, \sigma_{min}(\tilde{\Pi}^*)}{\bar{c} \, \sigma_{max}(A^T A)}.$$

Then, using the bound (7.20),

$$\tilde{\mathcal{T}}_q^{\lambda'} f^*(\Sigma) - f^*(\Sigma) \;\leq\; (\lambda'_q - \lambda_q) f^*(\Xi(\Sigma)) - \mathrm{tr}(\tilde{\Pi}^* \Sigma) + \varrho^* - r_S(q)$$

$$
\begin{aligned}
&\leq (\lambda'_q - \lambda_q)\left(\frac{\bar{c}}{m_1}\mathrm{tr}(\Xi(\Sigma)) + m_1 + m_0\right) \\
&\quad - \mathrm{tr}(\tilde{\Pi}^*\Sigma) + \varrho^* - r_S(q) \\
&\leq \left((\lambda'_q - \lambda_q)\frac{\bar{c}}{m_1}\sigma_{max}(A^TA) - \sigma_{min}(\tilde{\Pi}^*)\right)\mathrm{tr}(\Sigma) \\
&\quad + (\lambda'_q - \lambda_q)\left(\mathrm{tr}(DD^T) + m_1 + m_0\right) + \varrho^* - r_S(q) \\
&\leq -\delta\, f^*(\Sigma) + \bar{M}
\end{aligned}
$$

for some $\delta > 0$ and $\bar{M} \in \mathbb{R}$. Hence the chain is still geometrically ergodic (and therefore stabilizable) under $\lambda' \in [0,1]^{N_q}$ such that

$$(\lambda'_q - \lambda_q)^+ < \frac{m_1\,\sigma_{min}(\tilde{\Pi}^*)}{\bar{c}\,\sigma_{max}(A^TA)}. \qquad \square$$

An immediate corollary of Theorems 7.6.1–7.6.2 is the following.

Corollary 7.6.3. *Suppose that (A,B) is stabilizable and $(\overline{C},A)$ is detectable. Then, there exists a critical surface $\mathcal{W}$ in $(0,1]^{N_q}$ such that the system is stabilizable with loss rate λ if and only if $\lambda < \lambda' \in \mathcal{W}$. More precisely, there exists a function $\mathcal{F}:\ \mathbb{R}^{N_q-1} \to [0,1]$ which is nonincreasing in each argument such that the system is stabilizable with loss rate λ if and only if $\lambda_{N_q} < \mathcal{F}(\lambda_1,\ldots,\lambda_{N_q-1})$. In other words, Λ_s is the epigraph of $\mathcal{F}$.*

Proof. As shown in [58], under the hypotheses of the corollary, the system is stabilizable with $\lambda = 0$. The result then follows by Theorems 7.6.1 and 7.6.2 . $\square$

We call the set of sensor queries $\mathbb{Q} = \{q_1,\ldots,q_{N_q}\}$ *non-redundant* if the system is not detectable with any proper subset of the sensor queries. That is,

the system using only $\mathbb{Q} \setminus \{q_i\}$ for any $i = 1, \ldots, N_q$ is not stabilizable for any admissible query sequence. When $\mathbb{Q}$ is non-redundant and $\mathbf{q}$ is a stabilizing stationary Markov policy, the set of states where any particular query q_i is chosen,

$$\mathbf{S}_{q_i} = \{\Sigma \in \mathcal{M}_0^+ : \mathbf{q}(\Sigma) = q_i\},$$

satisfies $\mu_{\mathbf{q}}(\mathbf{S}_{q_i}) > 0$ for each $q_i \in \mathbb{Q}$. Furthermore, there must be a subset $\widehat{\mathbf{S}}_{q_i} \subset \mathbf{S}_{q_i}$ with $\mu_{\mathbf{q}}(\widehat{\mathbf{S}}_{q_i}) > 0$ such that $T_{q_i}(\widehat{\Sigma}) < \Xi(\widehat{\Sigma})$ for all $\widehat{\Sigma} \in \widehat{\mathbf{S}}_{q_i}$; if not, then a different sensor could be queried instead of q_i and the system would still be stable.

Theorem 7.6.4. *Suppose that the set of sensors is non-redundant and that $\lambda, \lambda' \in \Lambda_s$ such that $\lambda \leq \lambda'$ and $\lambda \neq \lambda'$. Then $\varrho_\lambda^* < \varrho_{\lambda'}^*$.*

Proof. Without loss of generality, let $\lambda, \lambda' \in \Lambda_s$ such that $\lambda_1 < \lambda_1'$ and $\lambda_i = \lambda_i'$ for $i = 2, \ldots N_q$. For the system with loss rate λ (respectively, λ'), let f_λ^* ($f_{\lambda'}^*$) be the solution of the ACOE, and let q^λ ($q^{\lambda'}$) be a selector of the corresponding minimizer. Define the set

$$\mathbf{S}_1^{\lambda'} = \{\Sigma \in \mathcal{M}_0^+ : q^{\lambda'}(\Sigma) = q_1, T_{q_i}(\Sigma) < \Xi(\Sigma)\},$$

which from the preceding discussion satisfies $\mu_{q^{\lambda'}}(\mathbf{S}_1^{\lambda'}) > 0$. Because $f_{\lambda'}^*$ is strictly increasing, for any query $q \in \mathbb{Q}$ we have

$$\tilde{\mathcal{T}}_q^{\lambda'} f_{\lambda'}^*(\Sigma) - \tilde{\mathcal{T}}_q^{\lambda} f_{\lambda'}^*(\Sigma) = (\lambda_q - \lambda_q') \left(f_{\lambda'}^*(T_q(\Sigma)) - f_{\lambda'}^*(\Xi(\Sigma))\right) \geq 0,$$

with strict inequality when $\Sigma \in \mathbf{S}_1^{\lambda'}$ and $q = q_1$. Define the non-negative function $g_q(\Sigma) := \tilde{\mathcal{T}}_q^{\lambda'} f_{\lambda'}^*(\Sigma) - \tilde{\mathcal{T}}_q^{\lambda} f_{\lambda'}^*(\Sigma)$. Then, for any $\Sigma \in \mathcal{M}_0^+$,

$$\begin{aligned}
\varrho_{\lambda'}^* &= r_S(q^{\lambda'}(\Sigma)) + \operatorname{tr}(\tilde{\Pi}^*\Sigma) + \tilde{\mathcal{T}}_{q^{\lambda'}}^{\lambda'} f_{\lambda'}^*(\Sigma) - f_{\lambda'}^*(\Sigma) \\
&= r_S(q^{\lambda'}(\Sigma)) + \operatorname{tr}(\tilde{\Pi}^*\Sigma) + g_{q_1}(\Sigma)\mathbb{I}_{q^{\lambda'}(\Sigma)=q_1} + \tilde{\mathcal{T}}_{q^{\lambda'}}^{\lambda} f_{\lambda'}^*(\Sigma) - f_{\lambda'}^*(\Sigma) \\
&= \frac{1}{T}\mathbb{E}_{\Sigma}^{\lambda,q^{\lambda'}}\left[\sum_{t=0}^{T-1} r_S(Q_t) + \operatorname{tr}(\tilde{\Pi}^*\hat{\Pi}_t)\right] + \frac{1}{T}\mathbb{E}_{\Sigma}^{\lambda,q^{\lambda'}}\left[\sum_{t=0}^{T-1} g_{Q_t}(\Sigma)\mathbb{I}_{Q_t=q_1}\right] \\
&\qquad + \frac{1}{T}\mathbb{E}_{\Sigma}^{\lambda,q^{\lambda'}}\left[f_{\lambda'}^*(\hat{\Pi}_T) - f_{\lambda'}^*(\hat{\Pi}_0)\right].
\end{aligned}$$

For all T large enough, the second term must be strictly positive because the process must query sensor q_1 with non-zero average frequency. Taking limits as $T \to \infty$, the third term approaches 0 and we are left with

$$\varrho_{\lambda'}^* > J_{\lambda}^{q^{\lambda'}},$$

where $J_{\lambda}^{q^{\lambda'}}$ is the average cost for the system with loss rate λ and using policy $q^{\lambda'}$. Since $q^{\lambda'}$ suboptimal, it follows that $\varrho_{\lambda}^* \leq J_{\lambda}^{q^{\lambda'}} < \varrho_{\lambda'}^*$. □

Noting that the average cost $\varrho_{\lambda}^* \to \infty$ as the system becomes less stable, the set $\Lambda(\kappa) := \{\lambda : \varrho_{\lambda}^* < \kappa\}$ is a ray-connected neighborhood of 0 for all $\kappa > 0$. Clearly, $\bigcup_{\kappa>0} \Lambda(\kappa) = \Lambda_s$.

Remark 7.6.5. Note that similar results could be shown for the more general case with network states dictating loss rates. However, the analysis is much more involved, and may require additional assumptions on the structure of the network state transition probabilities. We present the simpler version here to facilitate the analysis and the comparison to the previous works.

Remark 7.6.6. Suppose that the loss rates depend only on the query, as in (7.24), but are unknown. Then the implications of Theorem 7.6.2 are remarkable. Since stability is shown to be an open property, if one can find an estimator sequence $\hat{\lambda}_t \to \lambda$ a.s., then the system will retain stability and the long-term average performance would be the same as the if the rates were known beforehand. Since the channel is Bernoulli, recursive estimation of the loss rates leading to a.s. convergence to the true value is rather straightforward. For example, a maximum likelihood estimator can be employed, as in [29].

7.6.2 Diagonal Structures

Consider two independent one-dimensional systems

$$\begin{aligned} x_{k+1}^{(i)} &= a_i x_k^{(i)} + w_k^{(i)} \\ y_k &= x_k^{(i)} + f_i v_k^{(i)}, \end{aligned} \tag{7.26}$$

where $\{w_k^{(i)}, v_k^{(i)}, k \in \mathbb{N},\ i = 1, 2\}$ are i.i.d. Gaussian random variables. Note that we can always scale the system so that $c_i = 1$ and $w_k^{(i)}$ has unit variance, so the above representation is without loss of generality. Without loss of generality we focus on the estimation problem. It is well known that the Kalman filter with intermittent observations is stable for each subsystem separately if and only if $\lambda_i < a^{-2}$ [56].

We concentrate on the case where $a_1 = a_2 = a$ and assume that $a > 1$; otherwise the problem is trivial. Suppose that the intermittency rate is of the form (λ, λ) with $\lambda \in [0, a^{-2})$. Let ξ_1 and ξ_2 be the estimation error variances

of $x^{(1)}$ and $x^{(2)}$, respectively, and define $\xi := (\xi_1, \xi_2)$. Note that

$$\mathcal{T}_1(\xi) = \left(\frac{f_1^2\,(1 + a^2\xi_1)}{1 + a^2\xi_1 + f_1^2}, 1 + a^2\xi_2 \right),$$

and the analogous expression holds for $\mathcal{T}_2$. We have the bound

$$\frac{f_i^2\,(1 + a^2\zeta)}{1 + a^2\zeta + f_i^2} \;\le\; \max\,(f_1^2, f_2^2) \qquad \forall \zeta \in \mathbb{R}_+\,, \quad i = 1, 2\,.$$

For $\epsilon > 0$, let $\mathcal{V}_\epsilon \colon \mathbb{R}_+^2 \to \mathbb{R}_+$ be defined as follows:

$$\mathcal{V}_\epsilon(\xi) \;:=\; \begin{cases} \epsilon\,\xi_1 + (1-\epsilon)\,\xi_2\,, & \text{if } \xi_1 \ge \xi_2 \\ (1-\epsilon)\,\xi_1 + \epsilon\,\xi_2\,, & \text{otherwise.} \end{cases}$$

Let ϵ be small enough that

$$\epsilon_0 \;:=\; \big(\tfrac{\epsilon}{1-\epsilon} + \lambda\big)a^2 \;<\; 1\,, \tag{7.27}$$

and suppose $m_0 := \max\,(f_1^2, f_2^2) \le \xi_2 \le \xi_1$. Then we have

$$\begin{aligned} \widetilde{\mathcal{T}}_1\mathcal{V}_\epsilon(\xi) - \mathcal{V}_\epsilon(\xi) \;&=\; (1-\lambda)\mathcal{V}_\epsilon\big(\mathcal{T}_1(\xi)\big) + \lambda\mathcal{V}_\epsilon\big(1 + a^2\xi_1, 1 + a^2\xi_2\big) - \mathcal{V}_\epsilon(\xi) \\ &\le\; (1-\lambda)(1-\epsilon)\,m_0 + (1-\lambda)\epsilon\,(1 + a^2\xi_2) + \lambda\,\epsilon\,(1 + a^2\xi_1) \\ &\qquad\qquad + \lambda\,(1-\epsilon)\,(1 + a^2\xi_2) - \epsilon\,\xi_1 - (1-\epsilon)\,\xi_2 \\ &\le\; C_0 + \Big((1-\lambda)\tfrac{\epsilon\,a^2}{1-\epsilon} + \lambda\,a^2 - 1\Big)\mathcal{V}_\epsilon(\xi) \\ &\le\; C_0 - (1-\epsilon_0)\mathcal{V}_\epsilon(\xi)\,, \end{aligned}$$

where C_0 is a constant depending on ϵ, λ, and m_0. On the other hand, if $\xi_1 \ge \xi_2$, and $\xi_2 < m_0$, then $\mathcal{V}_\epsilon\big(\mathcal{T}_1(\xi)\big)$ is bounded and we obtain

$$\widetilde{\mathcal{T}}_1\mathcal{V}_\epsilon(\xi) - \mathcal{V}_\epsilon(\xi) \;\le\; C_0' + (\lambda a^2 - 1)\mathcal{V}_\epsilon(\xi)$$

for some constant C_0'. Therefore, by symmetry, we obtain

$$\min_{q=1,2} \widetilde{\mathcal{T}}_q \mathcal{V}_\epsilon(\xi) - \mathcal{V}_\epsilon(\xi) \;\le\; C_0'' - (1-\epsilon_0)\mathcal{V}_\epsilon(\xi) \qquad \forall \xi \in \mathbb{R}^2_+$$

for some constant C_0''. Since $(1-\epsilon_0) > 0$ by (7.27), geometric ergodicity follows.

The same technique applies for a diagonal system as in (7.26) of any order, and thus we have proved the following.

Theorem 7.6.7. *Consider a system in diagonal form as in* (7.26)*, with* $a_i = a > 1$*,* $i = 1, \dots, N_q$*. Then* $\Lambda_s = [0, 1/a^2)^{N_q}$*.*

7.6.3 Numerical Example

Our example is a one-dimensional unstable linear system with two available sensors:

$$\begin{array}{ccc} A = [2] & B = [1] & DD^T = [0.05] \\ C_1 = [0.1] & C_2 = [1] & FF^T = [0.02] \\ R = [0.01] & Q = [0] & r_S(\cdot) = 1 \end{array}$$

The first sensor has a much lower gain than the second, so is more vulnerable to the observation noise. With this structure, optimal policies either dictate that one sensor is queried continuously, or that one sensor is queried until the error covariance exceeds a threshold value, at which point the other sensor is queried.

Using a relative value iteration algorithm, the optimal policy was calculated for values of $(\lambda_1, \lambda_2) \in (0,1)^2$. Figure 7.1 shows the calculated threshold value for each λ pair where the system was stabilizable. The dark region

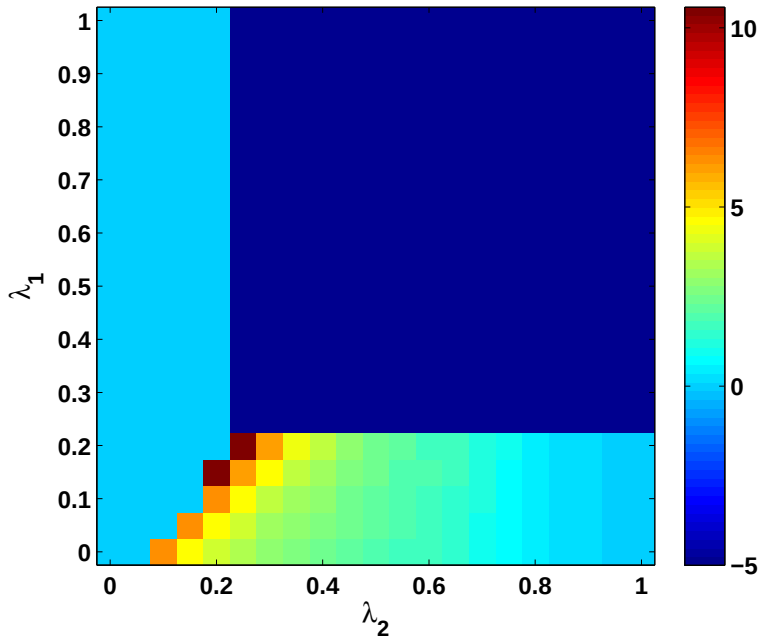

Figure 7.1: Critical surface and threshold values for λ_1 and λ_2

$(0.25 \leq \lambda_1 \leq 1,\ 0.25 \leq \lambda_2 \leq 1)$ corresponds to the loss rates that are too high to admit a stabilizing solution; the critical surface described in Corollary 7.6.3 is the border of the dark region. On other side of the critical surface $(0 \leq \lambda_1 < 0.2,\ 0 \leq \lambda_2 < 0.2)$ the color of the graph indicates the threshold value corresponding to the optimal policy. For the left portion of the graph, sensor 2 is used exclusively. However, when $\lambda_1 < 0.2$, as λ_2 increases sensor 1 becomes more desirable, and the optimal policy begins to select sensor 1 when the error covariance becomes large. In the lower right region $(0 \leq \lambda_1 < 0.2,$ $\lambda_2 \to 1)$, the high loss rate of sensor 2 drives the optimal policy to use sensor 1 exclusively.

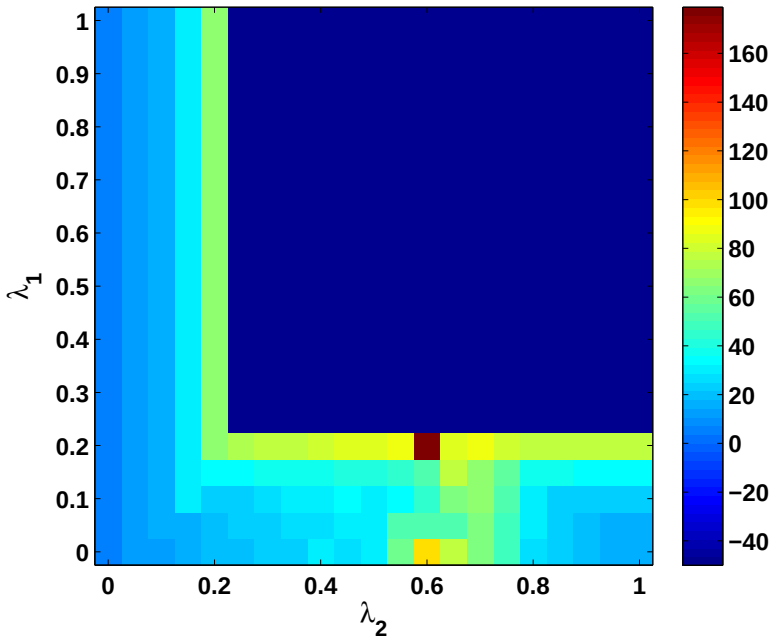

Figure 7.2: Number of iterations for the relative value iteration to converge.

Also of interest, Figure 7.2 shows how many iterations were needed by the relative value iteration to converge. The more colorful region in the lower middle indicates an area where the algorithm required significantly more iterations than elsewhere. For these λ-values, the expected average costs of using either the loss-prone stronger sensor or the reliable weaker sensor were nearly the same. Hence the difference between policies was small, and the algorithm took longer to determine the optimal policy choice.

Chapter 8

LQG System: Value Iteration

8.1 Introduction

We now investigate the convergence of the value iteration and relative value iteration algorithms for the linear quadratic Gaussian (LQG) system. Though more is known about the structure of the value function than in the countable state space (e.g., concavity, monotonicity), it is still an infeasible problem to calculate the value function and optimal policy directly. Whereas in the countable state space version we were forced to impose structural assumptions on the state space, the evolution of the error covariance $\hat{\Pi}$ on the set of positive semi-definite matrices has a natural structure that allows results without additional assumptions. The structure of the LQG system in fact guarantees that the cost function and optimal average cost satisfy the near-monotone condition, and that the cost function and value function satisfy an inequality of the form 5.2.1.

Here, we will use these properties to prove results of the same form as for the countable state space, ensuring that the value iteration converges for any bounded initialization, and therefore relative value iteration does also. We use the same notation as in Chapter 5, but in this context have systems

evolving on $\mathbb{S} \times \mathcal{M}_0^+$. Hence, we must consider ordering and continuity which were irrelevant previously. Based on the analysis of Chapter 7, we seek a concave, non-decreasing function $f^* : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}_+$ and a constant ϱ^* that solve the modified ACOE (7.17):

$$f^*(s, \hat{\Pi}) + \varrho^* \;=\; \min_q \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^* \hat{\Pi}) + \tilde{\mathcal{T}}_q f^*(s, \hat{\Pi}) \right\}.$$

The relative value iteration (RVI) and value iteration (VI) algorithms provide a sequence of functions and associated constants that, as we will show, approach f^* and ϱ. Respectively, the RVI and VI are given by

$$\varphi_{n+1}(s, \hat{\Pi}) \;=\; \min_{q \in \mathbb{Q}} \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^* \hat{\Pi}) + \tilde{\mathcal{T}}_q \varphi_n(s, \hat{\Pi}) \right\} - \varphi_n(0,0)\,, \tag{8.1}$$

$$\bar{\varphi}_{n+1}(s, \hat{\Pi}) \;=\; \min_{q \in \mathbb{Q}} \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^* \hat{\Pi}) + \tilde{\mathcal{T}}_q \bar{\varphi}_n(s, \hat{\Pi}) \right\} - \varrho^*\,, \tag{8.2}$$

where both algorithms are initialized with a function $\varphi_0 \in C_+(\mathbb{S} \times \mathcal{M}_0^+)$.

8.2 Additional Notation and Remarks

One of the useful characteristics of the linear system with quadratic costs is that the differential value function f^* has the same type of growth as the one step cost. Recalling the transformation under the optimal feedback control, the cost function

$$r(s, q, \Sigma) := r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^* \Sigma)$$

yields equivalent solutions to the optimal average cost problem. Then, since f^* is bounded above by an affine function of trace, as in (7.23), there exist

positive constants θ_1 and θ_2 such that

$$\min_{q\in\mathbb{Q}} r(s,q,\Sigma) \geq \theta_1 f^*(s,\Sigma) - \theta_2\,.$$

Without loss of generality we can assume $\theta_1 < 1$ to facilitate some later estimates.

If the cost function r_S is replaced with $r_S + c$ for some $c \in \mathbb{R}$, the resulting average cost will simply be $\varrho^* + c$ and the optimal policy will be unchanged. Hence, without loss of generality we will assume $\min_{\mathbb{S}\times\mathbb{U}} r_S = 1$. To simplify analysis, we will also occasionally use

$$\bar{r}(s,q,\hat{\Pi}) := r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^*\hat{\Pi}) - \varrho^*,$$

and for a Markov policy $\bar{q} : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{Q}$,

$$\bar{r}^{\bar{q}}(s,\hat{\Pi}) := r_S(s,\bar{q}(s,\Sigma)) + \mathrm{tr}(\tilde{\Pi}^*\hat{\Pi}) - \varrho^*.$$

For a function $f : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{R}$, define

$$\begin{aligned}\|f\|_{f^*} &:= \sup_{(s,\Sigma)\in\mathbb{S}\times\mathcal{M}_0^+} \frac{|f(s,\Sigma)|}{f^*(s,\Sigma)}\,,\\ \mathcal{O}(f^*) &:= \{f : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{R} : \|f\|_{f^*} < \infty\,, f \geq 0\}\,.\end{aligned}$$

We also define

$$\widehat{C}(\mathbb{S}\times\mathcal{M}_0^+) := \{h : \mathbb{S}\times\mathcal{M}_0^+ \to \mathbb{R}_+ : h(s,\cdot) \text{ is concave and non-decreasing}\}.$$

8.3 Main Results

The VI and RVI can be treated as discrete time dynamical systems, and we define the associated semi-cascades

$$\overline{\Phi}_n[\varphi_0] \;:=\; \{\varphi_0, \overline{\varphi}_1, \overline{\varphi}_2, \dots\}\,,$$

$$\Phi_n[\varphi_0] \;:=\; \{\varphi_0, \varphi_1, \varphi_2, \dots\}.$$

We also let $\mathcal{E} = \{f^* + c : c \in \mathbb{R}\}$ denote the set of solutions of the ACOE (7.17), and define for $c \in \mathbb{R}$ the set

$$\mathcal{G}_c := \left\{h : C_+(\mathbb{S} \times \mathcal{M}_0^+) \cap \widehat{C}(\mathbb{S} \times \mathcal{M}_0^+) : \|h\|_{f^*} < \infty, h - f^* \geq c\right\}.$$

Theorem 8.3.1. *If $\varphi_0 \in \mathcal{G}_c$ for some $c \in \mathbb{R}$, then $\overline{\Phi}_n[\varphi_0]$ converges to $c_0 + f^* \in \mathcal{E}$ for some $c_0 \in \mathbb{R}$ such that*

$$0 \leq c_0 \leq \mu_{q^*}[\varphi_0 - f^*]. \tag{8.3}$$

Also, $\Phi_n[\varphi_0]$ converges to $f^* - f^*(0,0) + \varrho^*$.

Theorem 8.3.2. *If $\varphi_0 \in \mathcal{O}_{f^*}$, then $\overline{\Phi}_n[\varphi_0]$ converges to $c_0 + f^* \in \mathcal{E}$ for some $c_0 \in \mathbb{R}$ satisfying*

$$-\frac{\varrho^* + \theta_2}{\theta_1} \leq c_0 \leq \|\varphi_0\|_{f^*} \frac{\varrho^* + \theta_2}{\theta_1}. \tag{8.4}$$

Also, $\Phi_n[\varphi_0]$ converges to $f^* - f^*(0,0) + \varrho^*$.

8.4 Supporting Lemmas

Before proving the results, we introduce some essential intermediate results. The first is a direct consequence of Lemmas 6.3.4 and 6.3.5:

Lemma 8.4.1. *If φ_0 is continuous, concave, and non-decreasing, φ_n and $\overline{\varphi}_n$ are also continuous, concave, and non-decreasing for all $n > 0$.*

Lemma 8.4.2. *For any $n \geq 0$ and $(s, \Sigma) \in \mathbb{S} \times \mathcal{M}_0^+$,*

$$\overline{\varphi}_n(s,\Sigma) \;=\; \varphi_n(s,\Sigma) - n\varrho^* + \sum_{k=0}^{n-1} \varphi_k(0,0), \tag{8.5}$$

$$\varphi_n(s,\Sigma) - \varphi_n(0,0) \;=\; \overline{\varphi}_n(s,\Sigma) - \overline{\varphi}_n(0,0), \tag{8.6}$$

$$\varphi_n(s,\Sigma) \;=\; \overline{\varphi}_n(s,\Sigma) - \overline{\varphi}_{n-1}(0,0) + \varrho^*. \tag{8.7}$$

Proof. Note that (8.5) holds trivially for $n = 0$, and that if true for any particular $n \geq 0$, then

$$\begin{aligned}
\overline{\varphi}_{n+1}(s,\Sigma) \;&=\; \min_{q\in\mathbb{Q}} \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^*\Sigma) + \tilde{\mathcal{T}}_q \overline{\varphi}_n(s,\Sigma) \right\} - \varrho^* \\
&=\; \min_{q\in\mathbb{Q}} \left\{ r_S(s,q) + \mathrm{tr}(\tilde{\Pi}^*\Sigma) + \tilde{\mathcal{T}}_q \varphi_n(s,\Sigma) \right\} \\
&\qquad\qquad - (n+1)\varrho^* + \sum_{k=0}^{n-1} \varphi_k(0,0) \\
&=\; \varphi_{n+1}(s,\Sigma) - (n+1)\varrho^* + \sum_{k=0}^{n} \varphi_k(0,0).
\end{aligned}$$

(8.6) follows directly, and (8.7) follows because

$$\begin{aligned}
\overline{\varphi}_n(s,\Sigma) - \overline{\varphi}_{n-1}(s,\Sigma) \;&=\; \varphi_n(s,\Sigma) - \varphi_{n-1}(s,\Sigma) + \varphi_{n-1}(0,0) - \varrho^* \\
&=\; \varphi_n(s,\Sigma) - \overline{\varphi}_{n-1}(s,\Sigma) + \overline{\varphi}_{n-1}(0,0) - \varrho^*. \qquad \square
\end{aligned}$$

A direct result of (8.7) is the following:

Corollary 8.4.3. *If $\overline{\varphi}_n$ converges pointwise to a function $f : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{R}$, then φ_n converges to $f - f(0,0) + \varrho^*$.*

Let $q^* : \mathbb{S} \times \mathcal{M}_0^+ \to \mathbb{Q}$ be a measurable selector from the minimizer of (7.17), and let $\mathfrak{q} = \{\mathfrak{q}_m, m \in \mathbb{N}\}$ be a measurable selector from the minimizer in (8.2) corresponding to a solution $\overline{\varphi}$. $\mathfrak{q}$ is also a measurable selector from the minimizer in (8.1) since φ and $\overline{\varphi}$ are related by (8.5) and (8.7). At the n^{th} step of the VI, define the (nonstationary) Markov control

$$\hat{q}^n \;:=\; \{\hat{q}_m^n = \mathfrak{q}_{n-m}, m \in \mathbb{N}, m < n\}\,. \tag{8.8}$$

Recalling that the inequality (7.20) satisfies the geometric drift condition [43, (V4)], we note the following direct implication.

Lemma 8.4.4. *There exists an invariant probability measure μ_{q^*} such that $\mu_{q^*}[f^*] < \infty$ and $\mathbb{E}^{q^*}_{s_0,\Sigma_0}[f^*(S_n, \hat{\Pi}_n)] \to \mu_{q^*}[f^*]$ as $n \to \infty$.*

Iterating the VI equation (8) using the standard dynamic programming formulation yields the following form:

$$\begin{aligned}\overline{\varphi}_n(s,\Sigma) \;&=\; \inf_{U \in \mathfrak{U}} \mathbb{E}^U_{s,\Sigma}\left[\varphi_0(S_n, \hat{\Pi}_n) + \sum_{k=0}^{n-1} \bar{r}(S_k, \hat{\Pi}_k, U_k)\right] \\ &=\; \mathbb{E}^{\hat{q}^n}_{s,\Sigma}\left[\varphi_0(S_n, \hat{\Pi}_n) + \sum_{k=0}^{n-1} \bar{r}\left(S_k, \hat{\Pi}_k, \hat{q}^n(S_k, \hat{\Pi}_k)\right)\right]. \end{aligned} \tag{8.9}$$

Lemma 8.4.5. *For any $n \geq 0$, it holds that*

$$\tilde{\mathcal{T}}_{\hat{q}_1^n}(\overline{\varphi}_n - f^*) \;\leq\; \overline{\varphi}_{n+1} - f^* \;\leq\; \tilde{\mathcal{T}}_{q^*}(\overline{\varphi}_n - f^*)\,.$$

Proof. By optimality we have

$$\overline{\varphi}_{n+1}(s,\Sigma) - f^*(s,\Sigma) \;=\; r_S(s,\hat{q}) - r_S(s,q^*) + \tilde{\mathcal{T}}_{\hat{q}}\,\overline{\varphi}_n(s,\Sigma) - \tilde{\mathcal{T}}_{q^*}\,f^*(s,\Sigma)$$

$$\leq \tilde{\mathcal{T}}_{q^*}(\overline{\varphi}_n(s,\Sigma) - f^*(s,\Sigma))\,,$$

and

$$\begin{aligned}\overline{\varphi}_{n+1}(s,\Sigma) - f^*(s,\Sigma) &= r_S(s,\hat{q}_1^n) - r_S(s,q^*) + \tilde{\mathcal{T}}_{\hat{q}_1^n}\overline{\varphi}_n(s,\Sigma) - \tilde{\mathcal{T}}_{q^*} f^*(s,\Sigma) \\ &\geq \tilde{\mathcal{T}}_{\hat{q}_1^n}(\overline{\varphi}_n(s,\Sigma) - f^*(s,\Sigma))\,. \qquad \square\end{aligned}$$

Lemma 8.4.6. *There exist constants $\alpha \in (0,1)$ and $c_2 \in \mathbb{R}$ such that*

$$\mathbb{E}^{q^*}_{s,\Sigma}[f^*(S_n, \hat{\Pi}_n)] \leq c_2 + \alpha^n f^*(s,\Sigma).$$

Proof. Note that the inequality in (7.20) holds without loss of generality for $M_1 < 1$. Letting $\alpha = 1 - M_1$ and rearranging, we get

$$\tilde{\mathcal{T}}_{q^*} f^*(s,\Sigma) \leq \alpha f^*(s,\Sigma) + M_0,$$

and iterating yields

$$\begin{aligned}\mathbb{E}^{q^*}_{s,\Sigma}[f^*(S_n, \hat{\Pi}_n] &\leq \sum_{k=0}^{n-1} \alpha^k M_0 + \alpha^n f^*(s,\Sigma) \\ &\leq \frac{M_0}{1-\alpha} + \alpha^n f^*(s,\Sigma). \qquad \square\end{aligned}$$

Recall that for $R > 0$, $\mathcal{B}_R = \{\Sigma \in \mathcal{M}_0^+ : \mathrm{tr}(\Sigma) \leq R\}$, and define the following shortened notation:

$$\tau_R := \tau(\mathbb{S} \times \mathcal{B}_R), \qquad \tau_R^n := \min\{n, \tau_R\}.$$

Lemma 8.4.7. *For $(s,\Sigma) \in \mathbb{S} \times \mathcal{M}_0^+$, $n \in \mathbb{N}$, and $R > 0$,*

$$\mathbb{E}^{\hat{q}^n}_{s,\Sigma}\left[\overline{\varphi}_{(n-\tau_R)}(S_{\tau_R}, \hat{\Pi}_{\tau_R})\mathbb{I}_{\tau_R > n}\right] \xrightarrow[m\to\infty]{} 0\,.$$

Proof. Iterating (8) with $\hat{q}^n$ and using the notation $\tilde{\mathcal{T}}^{(k)}_{\hat{q}^n} = \tilde{\mathcal{T}}_{\hat{q}^n_k} \circ \cdots \circ \tilde{\mathcal{T}}_{\hat{q}^n_0}$, for any $n > 0$ and stopping time τ we get

$$\begin{aligned}
\bar{\varphi}(s,\Sigma) &= \sum_{k=0}^{\tau\wedge n-1} \tilde{\mathcal{T}}^{(k)}_{\hat{q}^n} \bar{r}^{\hat{q}^n}(s,\Sigma) + \tilde{\mathcal{T}}^{(n)}_{\hat{q}^n} \left(\mathbb{I}_{\tau\geq n}\, \varphi_0(s,\Sigma) + \mathbb{I}_{\tau<n}\, \bar{\varphi}_{n-\tau}(s,\Sigma) \right) \\
&= \mathbb{E}^{\hat{q}^n}_{s,\Sigma} \left[\sum_{k=0}^{\tau\wedge n-1} \bar{r}^{\hat{q}^n}(s,\Sigma) + \mathbb{I}_{\tau\geq n}\, \varphi_0(s,\Sigma) \right] \\
&\qquad + \mathbb{E}^{\hat{q}^n}_{s,\Sigma} \left[\mathbb{I}_{\tau<n}\, \bar{\varphi}_{n-\tau}(s,\Sigma) \right]. \qquad (8.10)
\end{aligned}$$

Letting $\tau = \tau_R$, $\mathbb{P}^{\hat{q}^n}(\tau_R \geq n) \to 1$ as $R \to \infty$. So the first term in (8.10) tends to the right-hand side of (8.9) by monotone convergence, and the result follows. □

8.5 Proofs of Main Results

Proof of Theorem 8.3.1. Using Lemma 8.4.5 and recalling that $\tilde{\mathcal{T}}_q$ is order-preserving,

$$c \leq \bar{\varphi}_n - f^* \implies c = \tilde{\mathcal{T}}_{\hat{q}}\, c \leq \tilde{\mathcal{T}}_{\hat{q}^n_1}(\bar{\varphi}_n - f^*) \leq \bar{\varphi}_{n+1} - f^*.$$

Also, with Lemma 8.4.6,

$$\begin{aligned}
c &\leq \bar{\varphi}_{n+1}(s_0,\Sigma_0) - f^*(s_0,\Sigma_0) \\
&\leq \mathbb{E}^{q^*}_{s_0,\Sigma_0}\left[\varphi_0(S_n,\hat{\Pi}_n) - f^*(S_n,\hat{\Pi}_n)\right] \\
&\leq (\|\varphi_0\|_{f^*} - 1)\mathbb{E}^{q^*}_{s_0,\Sigma_0}[f^*(S_n,\hat{\Pi}_n)] \\
&\leq (\|\varphi_0\|_{f^*} - 1)(c_2 + \alpha^n f^*(s_0,\Sigma_0))\,. \qquad (8.11)
\end{aligned}$$

Since translating φ_0 by a constant simply translates the entire orbit by the same constant, without loss of generality we will assume $c = 0$. Because the cascade remains in $\mathcal{G}_0$, $\overline{\Phi}_n[\varphi_0] - f^* \geq 0$ and $\mu_{q^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is finite. By optimality,

$$\overline{\Phi}_n[\varphi_0](s,\Sigma) \;\leq\; \mathbb{E}^{q^*}_{s,\Sigma}\left[\sum_{k=0}^{n-m-1} \bar{r}(S_k, q^*(S_k,\hat{\Pi}_k),\hat{\Pi}_k) + \overline{\Phi}_m[\varphi_0](S_{n-m},\hat{\Pi}_{n-m})\right],$$

and so with $m = n-1$, we get

$$\mu_{q^*}\left[\overline{\Phi}_n[\varphi_0]\right] \leq \mu_{q^*}\left[\overline{\Phi}_{n-1}[\varphi_0]\right].$$

The map $n \to \mu_{q^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is non-increasing and bounded below, so it must be constant on the ω-limit set of φ_0 under $\overline{\Phi}_n$, denoted $\omega(\varphi_0)$. Because (8.11) implies $\sup_{n\geq 0}\|\overline{\Phi}_n[\varphi_0]\|_{f^*} < \infty$, $\{\overline{\Phi}_n[\varphi_0]\}$ are uniformly bounded by a multiple of f^*. On compact subsets of $\mathbb{S}\times\mathcal{M}_0^+$, $\{\overline{\Phi}_n[\varphi_0]\}$ are equicontinuous and uniformly bounded and so by the Arzela-Ascoli theorem $\{\overline{\Phi}_n[\varphi_0]\}$ is precompact on compact subsets. Therefore the limit set $\omega(\varphi_0)$ is non-empty and invariant [39]. Let $h \in \omega(\varphi_0)$, and define the non-negative (by Lemma 8.4.5) function

$$g_n(s,\Sigma) \;:=\; \tilde{\mathcal{T}}_{q^*}\left(\overline{\Phi}_{n-1}[h](s,\Sigma) - f^*(s,\Sigma)\right) - \left(\overline{\Phi}_n[h](s,\Sigma) - f^*(s,\Sigma)\right).$$

Then

$$\begin{aligned}\mathbb{E}^{q^*}_{s,\Sigma}&\left[\sum_{m=0}^{n-1} g_{n-m}(S_m,\hat{\Pi}_m)\right]\\ &=\; \mathbb{E}^{q^*}_{s,\Sigma}\left[h(S_n,\hat{\Pi}_n) - f^*(S_n,\hat{\Pi}_n)\right] + f^*(s,\Sigma) - \overline{\Phi}_n[h](s,\Sigma). \quad (8.12)\end{aligned}$$

Integrating with respect to the invariant distribution μ_{v^*} yields

$$\sum_{m=0}^{n-1}\mu_{q^*}[g_{n-m}] \;=\; \mu_{q^*}\left[h - \overline{\Phi}_n[h]\right] \quad \forall n\in\mathbb{N}. \qquad (8.13)$$

Since both h and $\overline{\Phi}_n[h]$ are in $\omega(\varphi_0)$, the right-hand side of (8.13) is equal to zero and therefore $g_n(s,\Sigma) = 0$, (n,s,Σ)-almost everywhere. Using Lemma 8.4.4, (8.12) becomes

$$\lim_{n\to\infty} \overline{\Phi}_n[h](s,\Sigma) \;=\; f^*(s,\Sigma) + \mu_{q^*}\left[h - f^*\right] .$$

Therefore $\omega(\varphi_0) \subset \mathcal{E} \cap \mathcal{G}_0$, and since $\mu_{q^*}\left[f^* - h\right]$ is a constant, the limit set must be a single function. Because $\mu_{q^*}\left[\overline{\Phi}_n[\varphi_0]\right]$ is non-increasing in n, the inequality (8.3) is satisfied. Finally, by Lemma 8.4.3, $\Phi_n[\varphi_0]$ converges pointwise to $f^* - f^*(0) - \varrho^*$. □

Proof of Theorem 8.3.2. For $\epsilon > 0$, let $\overline{\varphi}^\epsilon$ be the solution of (8.2) with initial data $\varphi_0 + \epsilon f^*$, and let $\{\hat{q}^n_\epsilon : n = 0, 1, \dots\}$ be the corresponding Markov control, as in (8.8). For convenience let $\alpha = (1-\theta_1)$, $C = \frac{\varrho^* + \theta_2}{\theta_1}$, and let

$$f^\epsilon_n(s,\Sigma) \;:=\; \overline{\varphi}^\epsilon_n(s,\Sigma) - (1-\alpha^n)(f^*(s,\Sigma) - C).$$

Noting that from (7.17),

$$(\tilde{\mathcal{T}}_{\hat{q}^n_\epsilon} - I) f^*(s,\Sigma) \;\geq\; -r_S(s,\hat{q}^n_\epsilon) - \operatorname{tr}(\tilde{\Pi}^*\Sigma) + \varrho^*,$$

we have

$$\begin{aligned}
F^\epsilon_n(s,\Sigma) \;&:=\; f^\epsilon_n(s,\Sigma) - \tilde{\mathcal{T}}_{\hat{q}^n_\epsilon} f^\epsilon_{n-1}(s,\Sigma) \\
&= \bar{r}^{\hat{q}^n_\epsilon}(s,\Sigma) - \theta_1\alpha^{n-1}(f^*(s,\Sigma) - C) \\
&\qquad + (1-\alpha^{n-1})(\tilde{\mathcal{T}}_{\hat{q}^n_\epsilon} - I)(f^*(s,\Sigma) - C) \\
&\geq \bar{r}^{\hat{q}^n_\epsilon}(s,\Sigma) - \theta_1\alpha^{n-1}(f^*(s,\Sigma) - C)
\end{aligned}$$

$$+ (1-\alpha^{n-1})(-\bar{r}^{\hat{q}^n_\epsilon}(s,\Sigma) - C)$$

$$\begin{aligned}
&= \alpha^{n-1}\left(-\theta_1 f^*(s,\Sigma) + \theta_2 + r(s,\Sigma,\hat{q}^n_\epsilon(s,\Sigma) + C)\right)\\
&\geq \alpha^{n-1}\left(-\theta_1 f^*(s,\Sigma) + \theta_2 + \theta_1 f^*(s,\Sigma) - \theta_2\right)\\
&= 0 \qquad \forall (s,\Sigma) \in \mathbb{S}\times\mathcal{M}_0^+ \text{ and } n \in \mathbb{Z}_+ .
\end{aligned}$$

Note that the formulation of Dynkin's formula in Corollary 3.4.9 is in fact applicable to general state spaces. Hence, applying Dynkin's formula to f^ϵ:

$$\begin{aligned}
f^\epsilon_n(s,\Sigma) &= \mathbb{E}^{\hat{q}^n_\epsilon}_{s,\Sigma}\left[\sum_{k=0}^{\tau^n_R-1} F^\epsilon_{(n-k)}(S_k,\hat{\Pi}_k) + f^\epsilon_{(n-\tau^n_R)}(S_{\tau^n_R},\hat{\Pi}_{\tau^n_R})\right]\\
&= \mathbb{E}^{\hat{q}^n_\epsilon}_{s,\Sigma}\left[\sum_{k=0}^{\tau^n_R-1} F^\epsilon_{(n-k)}(S_k,\hat{\Pi}_k) + f^\epsilon_0(S_n,\hat{\Pi}_n)\mathbb{I}_{\{n\leq\tau_R\}}\right]\\
&\qquad + \mathbb{E}^{\hat{q}^n_\epsilon}_{s,\Sigma}\left[f^\epsilon_{(n-\tau_R)}(S_{\tau_R},\hat{\Pi}_{\tau_R})\mathbb{I}_{\{n>\tau_R\}}\right] . \qquad (8.14)
\end{aligned}$$

From Lemma 8.4.7 we have for any $(s,\Sigma)\in\mathbb{S}\times\mathcal{M}_0^+$ and $n\in\mathbb{N}$,

$$\mathbb{E}^{\hat{q}^n_\epsilon}_{s,\Sigma}\left[f^\epsilon_{(n-\tau_R)}(S_{\tau_R},\hat{\Pi}_{\tau_R})\mathbb{I}_{\tau_R>n}\right] \xrightarrow[R\to\infty]{} 0. \qquad (8.15)$$

Then letting $R\to\infty$ in (8.14), using Fatou's lemma and (8.15), we have $f^\epsilon_n(s,\Sigma)\geq 0$ for all $(s,\Sigma)\in\mathbb{S}\times\mathcal{M}_0^+$ and $n\in\mathbb{N}$. By construction, $\overline{\varphi}^\epsilon \geq \overline{\varphi}$ and $\overline{\varphi}^\epsilon$ decreases with ϵ, so each $\overline{\varphi}^\epsilon$ satisfies

$$\begin{aligned}
\overline{\varphi}^\epsilon_{n+1}(s,\Sigma) &= \min_{q\in\mathbb{Q}}\left[\bar{r}(s,\Sigma,q) + \tilde{\mathcal{T}}_q\overline{\varphi}^\epsilon_n(s,\Sigma)\right],\\
\overline{\varphi}^\epsilon_0(s,\Sigma) &= \varphi_0(s,\Sigma) + \epsilon f^*(s,\Sigma),
\end{aligned}$$

and $\overline{\varphi}^{\epsilon} \downarrow \overline{\varphi}^{0}$ for some pointwise limit $\overline{\varphi}^{0}$. Clearly $\overline{\varphi}^{0}_{0} = \varphi_0$, and so if we suppose that $\overline{\varphi}^{0}_{n} = \overline{\varphi}_n$ then

$$\begin{aligned}
\overline{\varphi}^{\epsilon}_{n+1}(s,\Sigma) - \overline{\varphi}_{n+1}(s,\Sigma) &= \min_{q\in\mathbb{Q}} \left[\bar{r}(s,\Sigma,q) + \tilde{\mathcal{T}}_q \overline{\varphi}^{\epsilon}_{n}(s,\Sigma)\right] - \overline{\varphi}_{n+1}(s,\Sigma) \\
&\leq \bar{r}^{\hat{q}_n}(s,\Sigma) + \tilde{\mathcal{T}}_{\hat{q}_n}\overline{\varphi}_n(s,\Sigma) - \overline{\varphi}_{n+1}(s,\Sigma) \\
&\qquad + \tilde{\mathcal{T}}_{\hat{q}_n}(\overline{\varphi}^{\epsilon}_{n}(s,\Sigma) - \overline{\varphi}_n(s,\Sigma)) \\
&= \tilde{\mathcal{T}}_{\hat{q}_n}(\overline{\varphi}^{\epsilon}_{n}(s,\Sigma) - \overline{\varphi}_n(s,\Sigma)) \xrightarrow[\epsilon\to\infty]{} 0\,.
\end{aligned}$$

Hence, inductively, $\bar{\varphi}_0 = \bar{\varphi}$ everywhere, and so

$$\overline{\varphi}_n(s,\Sigma) - \left(f^*(s,\Sigma) - \frac{\varrho^* + \theta_2}{\theta_1}\right) = \lim_{\epsilon\downarrow 0} f^{\epsilon}_n(s,\Sigma) \geq 0 \tag{8.16}$$

for all $\forall (s,\Sigma) \in \mathbb{S}\times\mathcal{M}^{+}_{0}$ and $n\in\mathbb{N}$. From Lemmas 8.4.5 and 8.4.6, we have

$$\begin{aligned}
\overline{\varphi}_{n+1}(s,\Sigma) - f^*(s,\Sigma) &\leq \mathbb{E}^{q^*}_{(s,\Sigma)}[\varphi_0(S_n,\hat{\Pi}_n) - f^*(S_n,\hat{\Pi}_n)] \\
&\leq (\|\varphi_0\|_{f^*} - 1)\mathbb{E}^{q^*}_{(s,\Sigma)}[f^*(S_n,\hat{\Pi}_n)] \\
&\leq (\|\varphi_0\|_{f^*} - 1)(c_2 + \alpha^n f^*(s,\Sigma)).
\end{aligned}$$

Combining this inequality with (8.16) yields

$$\begin{aligned}
(1-\alpha^n)\left(f^*(s,\Sigma) - \frac{\varrho^*+\theta_2}{\theta_1}\right) &\leq \overline{\varphi}_n(s,\Sigma) \\
&\leq f^*(s,\Sigma) + \|\varphi_0\|_{f^*}\left(\frac{\varrho^*+\theta_2}{\theta_1} + \alpha^n f^*(s,\Sigma)\right). \quad (8.17)
\end{aligned}$$

From (8.17), every ω-limit point of $\overline{\Phi}_n[\varphi_0]$ lies in the set

$$G(\varphi_0) := \left\{h : \mathbb{S}\to\mathbb{R}, -\frac{\varrho^*+\theta_2}{\theta_1} \leq h - f^* \leq \|\varphi_0\|_{f^*}\frac{\varrho^*+\theta_2}{\theta_1}\right\},$$

and $G(\varphi_0) \subset \mathcal{G}_{-C}$. The ω-limit set is invariant under $\overline{\Phi}_n$, and by Theorem 8.3.1 the only invariant subsets of $\mathcal{G}_{-C}$ are also subsets of $\mathcal{E}$. Thus (8.4) holds, and the rest of the result follows from (8.7). □

8.6 Rolling Horizon Estimates

The value iteration procedure is promising as a method to generate near-optimal policies, but stability of the generated policy is not guaranteed. One would hope that the Markov policy computed at the n^{th} stage of the value iteration is a stable Markov policy and its performance converges to the optimal performance as $n \to \infty$. This topic is commonly referred to as *rolling horizon*, and is well understood for finite state MDPs [26] but it is decidedly unexplored for nonfinite state models. Among the very few results in the literature is the study in [19] for bounded running cost and under a simultaneous Doeblin hypothesis, and the results in [33] under strong blanket stability assumptions. For the model considered here there is no blanket stability; instead, the inf-compactness of the running cost penalizes unstable behavior. Exploiting the constructive steps of the value iteration convergence proofs allows us to show that the rolling horizon policies are indeed stable.

Assume for simplicity that $\varphi_0 = 0$. Using the bounds in (8.17) and (7.23) we get

$$\begin{aligned} |\overline{\varphi}_{n+1} - \overline{\varphi}_n| &\le \frac{\varrho^* + \theta_2}{\theta_1} + \alpha^n \left(f^* - \frac{\varrho^* + \theta_2}{\theta_1} \right) \\ &\le \alpha^n m_1^* \mathrm{tr}(\cdot) + \widehat{C}\,, \end{aligned} \tag{8.18}$$

where $\widehat{C}$ is the appropriate combination of constants. Recalling the definition of $\hat{q}^n$ from (8.8),

$$\begin{aligned}\tilde{\mathcal{T}}_{\hat{q}^n}\,\overline{\varphi}_{n+1} - \overline{\varphi}_{n+1} &= \tilde{\mathcal{T}}_{\hat{q}^n}(\overline{\varphi}_{n+1} - \overline{\varphi}_n) - \mathrm{tr}(\tilde{\Pi}^*\,\cdot\,) - r_S + \varrho^* \\ &\leq \alpha^n \widehat{C}_1 \mathrm{tr}(\cdot) - \mathrm{tr}(\tilde{\Pi}^*\,\cdot\,) + \widehat{C}_0\,, \end{aligned} \tag{8.19}$$

where $\widehat{C}_0$ and $\widehat{C}_1$ are appropriate combinations of constants from (8.18) and (6.10) along with ϱ^* and the minimal value of r_S. But in (8.19), after some finite number of steps N, the second trace term will dominate the first:

$$\tilde{\mathcal{T}}_{\hat{q}^n}\,\overline{\varphi}_{n+1} - \overline{\varphi}_{n+1} \leq -\widehat{C}_2\ \mathrm{tr}(\tilde{\Pi}^*\,\cdot\,) + \widehat{C}_0\,,\ \text{for all } n > N.$$

In fact, since in (8.17) we have $\overline{\varphi}_{n+1} \leq f^*$, we can use the bound in (7.23) again to show that with appropriate constants $\widehat{C}_3 > 0$ and $\widehat{C}_4$ the chain is geometrically ergodic:

$$\tilde{\mathcal{T}}_{\hat{q}^n}\,\overline{\varphi}_{n+1} - \overline{\varphi}_{n+1} \leq -\widehat{C}_3\,\overline{\varphi}_{n+1} + \widehat{C}_4\,,\ \text{for all } n > N.$$

So the policy generated by the n^{th} stage of the value iteration algorithm is geometrically stable for n large enough.

Let ϱ_n^* be the average cost obtained under the stable policy $\hat{q}^n$, and let $r^n(s,\Sigma) = r_S(s,\hat{q}(s,\Sigma),\Sigma) + \mathrm{tr}(\tilde{\Pi}^*\Sigma)$. Following the method in [26], since $\mu_{\hat{q}^n}$ is invariant under $\hat{q}^n$ we have

$$\varrho_n^* = \mu_{\hat{q}^n}[r^n] = \mu_{\hat{q}^n}\left[\overline{\varphi}_{n+1} - \tilde{\mathcal{T}}_{\hat{q}^n}\overline{\varphi}_n + \varrho^*\right] = \mu_{\hat{q}^n}\left[\overline{\varphi}_{n+1} - \overline{\varphi}_n + \varrho^*\right]\,.$$

Therefore, as $n \to \infty$, $(\overline{\varphi}_{n+1} - \overline{\varphi}_n) \to 0$ and so $\varrho_n^* \to \varrho^*$. In fact, from the bound in (8.18) the convergence to the optimal average cost is geometric.

This result has significant implications for computational effort. The geometric convergence rate indicates that only a few iterates of the VI algorithm are needed to find a stable control that is near-optimal.

Chapter 9

Conclusion and Future Work

9.1 Overview

In this dissertation, we have produced several new results in Markov decision processes (MDPs), focusing on countable state systems and a discrete linear system with intermittent observations. Notably, we extend the concept of uniform stability for MDPs on countable state spaces, and show new sufficient conditions for the convergence of the value iteration algorithm that do not require global stability assumptions. We also analyze the optimal control and value iteration algorithm for a new class of linear quadratic Gaussian (LQG) systems with multiple sensors and query-dependent intermittent observations. This new system can be applied to various remote sensing and control applications in various fields.

9.2 MDPs on a Countable State Space

In the first area, MDPs on countable state spaces, we present a number of results on structure, recurrence, and value iteration for the average cost optimization problem.

In Chapter 3 we propose a set of assumptions that facilitate the trans-

lation of continuous diffusion process results into the area of countable state MDPs. These assumptions capture the fundamental aspects of the continuous behavior of diffusion processes applied to the countable state space. We also derive analogous discrete and countable versions of Harnack's inequality and re-frame related results for the countable state space.

Two new results for countable state MDPs are presented in Chapter 4, utilizing the structural assumptions from Chapter 3. First, a uniform recurrence result extends the uniform stability theorem of [15]. The theorem shows that, under appropriate assumptions, if the hitting time for finite set from any particular initial state (appropriately separated from the finite set) is finite under any particular policy, then the supremum over policies of hitting times from any state to any set is also finite. This result fills a gap in [15] suggested by [3, Section 3.3.2]. The second result shows useful uniform bounds on the variation and value of the discounted value function on certain finite sets. Since these sets, by construction, cover the entire space, the result can be used to show pointwise convergence in vanishing discount problems.

Similar results can be explored in the future for other types of Markov processes. All of the results for the countable state space should have analogous results in for continuous time, countable state processes and for discrete time, general state processes. Even if results can only be shown under somewhat restrictive structural assumptions, the nature of those assumptions can provide insight into the underlying structure of the various processes.

In Chapter 5, we give two new sufficient conditions for the convergence

of the value iteration for MDPs with the near-monotone property. These conditions do not require global stability, instead making assumptions only about the system under the optimal policy. The first result assumes that the value function is integrable with respect to the optimal invariant distribution. In that case, if the initial function of the value iteration dominates and is of the order of the value function, then the value iteration converges. The second result relies on a more specific assumption: if the cost function plus a constant dominates the value function then the value iteration converges with any initial function of order less than the value function (such as a constant). These results dramatically expand research into convergence of the value iteration, as previous results required blanket stability assumptions.

Future efforts should investigate the rate of convergence of the value iteration under the new conditions. It is anticipated that, as in [24], initializing the algorithm with a function of appropriate form will significantly improve the convergence rate, but the problem is open. Also, the structural relationship between the value function and cost function can be exploited in many value iteration problems. One example is the LQG value iteration algorithm in Chapter 8, where the system and cost structure imply structural properties of the value function. Future work can investigate other examples of this implied structure for countable and other state spaces.

9.3 LQG System with Sensor Scheduling and Intermittent Observations

Chapters 6–8 study a discrete time linear system with additive Gaussian white noise and quadratic costs. Additionally, the observations are randomly received or lost, where the loss rate is determined by a network state and sensors chosen by the controller. Chapter 6 describes the system in detail and shows that a modified Kalman filter provides optimal estimates of the system state despite the intermittency. We also show that the covariance update operator (which is an operator because the covariance is itself random) preserves concavity and continuity for non-decreasing functions, a result which is essential in the subsequent analysis.

The various optimal control problems (finite horizon, discounted cost, average cost) are detailed in Chapter 7, and we show that for each problem, the optimal control policy consists of a fixed feedback of the expected value of the state. The optimality conditions are then transformed into MDPs on the network state and error covariance processes, with modified cost functions depending on the trace of the error covariance. We show existence of value functions and optimality criteria for all three control problems, and note the special structure of the resulting average cost value function. We also show how a special case of the result generalizes a known result in Kalman filtering with intermittent observations. For a system with N possible sensor queries, when the observation loss rate depends only on the query we can write the loss rates as a vector in $[0, 1]^N$. We then show that there exists a critical

surface in $[0,1]^N$; loss rate vectors below the critical surface imply the system is stabilizable, while vectors above the surface lead to systems that cannot be stabilized.

We assume in Chapters 6–8 that the network state S_t is known at each time $t \geq 0$, but this is not necessarily required. One can treat the network as a partially-observed MDP being controlled simultaneously with the linear system, and estimate the network state based on knowledge of the process γ_t and the loss probabilities given by $\lambda(s,q)$. The traditional approach (e.g., see [28, Chapter 8] and [2]) is to treat the observation as an extension of the process state. In this case the extended system $(S_t, \hat{\Pi}_t, \gamma_t)$ would take values on $\mathbb{S} \times \mathcal{M}_0^+ \times \{0,1\}$, where only the second and third components are available to the controller. One can then create an equivalent completely observed model $(\Psi_t, \hat{\Pi}_t, \gamma_t)$, where Ψ_t is a process that evolves on $\mathcal{P}(\mathbb{S})$, the set of probability measures on the network state space. Provided that the loss rates and size of the network state space is known, the transition matrices can be estimated up to identifiability, via well studied algorithms (e.g., [5]), and since we are dealing with the ergodic cost, the performance of an adaptive algorithm would be the same as if the transition matrix were known.

If the loss rates corresponding to each network state are also unknown, the problem is more complicated, but still may be solvable. We augment the state space $\mathbb{S}$ to $\mathbb{S} \times \{0,1\}$. If p_{ij} is the transition matrix of $\mathbb{S}$, then the transition matrix of the new state space is given by $\tilde{p}_{(i,k),(j,1)} = \lambda_j p_{ij}$, and $\tilde{p}_{(i,k),(j,0)} = (1-\lambda_j)p_{ij}$, for $k = 0, 1$. Then we are dealing with a Markovian

identification problem, except that the transition matrix is constrained to a particular form. The problems seem tractable, although it is unclear if it has been studied in the existing literature.

Another obvious area for future work is jump linear systems, in which the state matrix and input gain are also subject to random, controller-dependent switching. Such a system is modeled as [25]:

$$X_{t+1} = A_{\theta_t} X_t + B_{\theta_t} U_t + D_{\theta_t} W_t,$$

$$Y_t = C_{\theta_t} X_t + F_{\theta_t} W_t,$$

where θ_t is a Markov chain on a (usually finite) set of states. There has been extensive research into controlling the state dynamics of jump linear systems with uncontrolled Markov chains ([23, 30], among others), but little has been done incorporating controlled chains and intermittent network channels.

9.4 General Conclusions

A recurring theme throughout much of the dissertation is the exploitation of structural similarities between the cost function and value function when considering average cost optimal control. For the countable state system, we posed the similar structure as an assumption, but for the LQG system the inherent properties of the system guaranteed structural similarity. The utility of this theme suggests that a more general framework may exist for analyzing MDPs on other spaces and with other constraints. Informally, if the cost function and value function are similar enough, the value iteration algorithm will converge. However, the "similarity" used here varies from both

being integrable, to both being bounded above by the same function, to explicitly sharing a growth rate. An interesting area for future research is to better quantify the similarity between the cost and value functions, and to extend the results to a general state space. As the areas of application increase, one might pose a quite generalized statement of how and under what conditions the structural relationship between the cost and value functions affects the optimal control and value iteration problems.

Bibliography

[1] J. Abounadi, D. Bertsekas, and V. S. Borkar. Learning algorithms for Markov decision processes with average cost. *SIAM Journal on Control and Optimization*, 40(3):681–698, 2001.

[2] A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M. Ghosh, and S. Marcus. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.

[3] A. Arapostathis, V. S. Borkar, and M. K. Ghosh. *Ergodic Control of Diffusion Processes*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2012.

[4] A. Arapostathis, V. S. Borkar, and K. Kumar. Convergence of the relative value iteration for the ergodic control problem of nondegenerate diffusions under near-monotone costs. *SIAM Journal on Control and Optimization*, 52(1):1–31, 2014.

[5] A. Arapostathis and S. I. Marcus. Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Math. Control Signals Systems*, 3(1):1–29, 1990.

[6] Y. Aviv and A. Federgruen. The value iteration method for countable state Markov decision processes. *Operations Research Letters*, 24(5):223 – 234, 1999.

[7] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[8] R. Bellman. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

[9] D. P. Bertsekas. A new value iteration method for the average cost dynamic programming problem. *SIAM journal on control and optimization*, 36(2):742–759, 1998.

[10] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume II. Athena Scientific, 3rd edition, 2005.

[11] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, 3rd edition, 2005.

[12] M. Bhuiyan, G. Wang, J. Cao, and J. Wu. Deploying wireless sensor networks with fault-tolerance for structural health monitoring. *Computers, IEEE Transactions on*, 64(2):382–395, Feb 2015.

[13] P. Billingsley. *Weak Convergence in Metric Spaces*, pages 7–79. John Wiley & Sons, Inc., 2008.

[14] V. S. Borkar. A convex analytic approach to Markov decision processes. *Probability Theory and Related Fields*, 78(4):583–602, 1988.

[15] V. S. Borkar. Uniform stability of controlled Markov processes. In T. Djaferis and I. Schick, editors, *System Theory*, volume 518 of *The Springer International Series in Engineering and Computer Science*, pages 107–120. Springer US, 2000.

[16] V. Borkar. *Topics in controlled Markov chains*. Number 240 in Pitman Research Notes in Mathematics Series. Longman Scientific & Technical, UK, 1991.

[17] L. Breyer. On a parabolic Harnack inequality for Markov chains. *Preprint available at http://www.statslab.cam.ac.uk/ laird*, 1998.

[18] N. Buerle and U. Rieder. *Markov Decision Processes with Applications to Finance*. Springer-Verlag Berlin Heidelberg, 2011.

[19] R. Cavazos-Cadena. A note on the convergence rate of the value iteration scheme in controlled Markov chains. *Systems Control Lett.*, 33(4):221–230, 1998.

[20] R. Cavazos-Cadena. Value iteration in a class of communicating Markov decision chains with the average cost criterion. *SIAM Journal on Control and Optimization*, 34(6):1848–1873, 1996.

[21] R. Cavazos-Cadena and E. Fernández-Gaucherand. Denumerable controlled Markov chains with strong average optimality criterion: Bounded & unbounded costs. *Mathematical Methods of Operations Research*, 43(3):281–300, 1996.

[22] R. Cavazos-Cadena and E. Fernández-Gaucherand. Value iteration in a class of average controlled Markov chains with unbounded costs: Necessary and sufficient conditions for pointwise convergence. *Journal of applied probability*, 33:986–1002, 1996.

[23] J. Cerri and M. Terra. Control of discrete-time Markovian jump linear systems subject to partially observed chains. In *American Control Conference (ACC), 2012*, pages 1609–1614, June 2012.

[24] R.-R. Chen and S. Meyn. Value iteration and optimization of multiclass queueing networks. *Queueing Systems*, 32(1-3):65–97, 1999.

[25] O. L. V. Costa, M. D. Fragoso, and R. P. Marques. *Discrete-time Markov jump linear systems*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005.

[26] E. Della Vecchia, S. Di Marco, and A. Jean-Marie. Illustrated review of convergence conditions of the value iteration algorithm and the rolling horizon procedure for average-cost MDPs. *Ann. Oper. Res.*, 199:193–214, 2012.

[27] S. Dreyfus. Richard Bellman on the birth of dynamic programming. *Operations Research*, 50(1):pp. 48–51, 2002.

[28] E. B. Dynkin and A. A. Yushkevich. *Controlled Markov Processes*, volume 235 of *Grundlehren der mathematischen Wissenschaften*. Springer-

Verlag New York, 1 edition, 1979. Translated by Holland, C., Danskin, J.M.

[29] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus. Analysis of an adaptive control scheme for a partially observed controlled Markov chain. *IEEE Trans. Automat. Control*, 38(6):987–993, 1993.

[30] V. Gupta, R. Murray, and B. Hassibi. On the control of jump linear Markov systems with Markov state estimation. In *American Control Conference, 2003. Proceedings of the 2003*, volume 4, pages 2893–2898 vol.4, June 2003.

[31] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, February 2006.

[32] V. Gupta, B. Hassibi, and R. M. Murray. Optimal LQG control across packet-dropping links. *Systems Control Lett.*, 56(6):439–446, 2007.

[33] O. Hernández-Lerma and J. B. Lasserre. Error bounds for rolling horizon policies in discrete-time Markov control processes. *IEEE Trans. Automat. Control*, 35(10):1118–1124, 1990.

[34] A. Hordijk, P. J. Schweitzer, and H. Tijms. The asymptotic behaviour of the minimal total expected cost for the denumerable state Markov decision model. *Journal of Applied Probability*, 12(2):pp. 298–305, 1975.

[35] M. Huber. Optimal pruning for multi-step sensor scheduling. *Automatic Control, IEEE Transactions on*, 57(5):1338–1343, May 2012.

[36] S. Joshi and S. Boyd. Sensor selection via convex optimization. *Signal Processing, IEEE Transactions on*, 57(2):451–462, Feb 2009.

[37] S. Kar, B. Sinopoli, and J. Moura. Kalman filtering with intermittent observations: Weak convergence to a stationary distribution. *Automatic Control, IEEE Transactions on*, 57(2):405–420, Feb 2012.

[38] K. S. Kumar and C. Pal. Risk-sensitive control of pure jump process on countable space with near monotone cost. *Applied Mathematics & Optimization*, 68(3):311–331, 2013.

[39] J. La Salle. *The Stability of Dynamical Systems*. Society for Industrial and Applied Mathematics, 1976.

[40] R. Lenin and S. Ramaswamy. Performance analysis of wireless sensor networks using queuing networks. *Annals of Operations Research*, 233(1):237–261, 2015.

[41] C. Li and N. Elia. Stochastic sensor scheduling via distributed convex optimization. *Automatica*, 58:173 – 182, 2015.

[42] L. Meier, J. Peschon, and R. Dressler. Optimal control of measurement subsystems. *Automatic Control, IEEE Transactions on*, 12(5):528–536, October 1967.

[43] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering. Springer, 1993.

[44] Y. Mo, E. Garone, and B. Sinopoli. On infinite-horizon sensor scheduling. *Systems Control Lett.*, 67:65–70, 2014.

[45] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. Stochastic sensor scheduling for energy constrained estimation in multi-hop wireless sensor networks. *Automatic Control, IEEE Transactions on*, 56(10):2489–2495, Oct 2011.

[46] Y. Mo and B. Sinopoli. Kalman filtering with intermittent observations: Tail distribution and critical value. *Automatic Control, IEEE Transactions on*, 57(3):677–689, March 2012.

[47] M. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley series in probability and statistics. Wiley-Interscience, 2005.

[48] R. Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics and physics. Princeton University Press, 1970.

[49] E. Rohr, D. Marelli, and M. Fu. Kalman filtering with intermittent observations: On the boundedness of the expected error covariance. *Automatic Control, IEEE Transactions on*, 59(10):2724–2738, Oct 2014.

[50] E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer Series in Statistics. Springer, 2006.

[51] L. I. Sennott. Value iteration in countable state average cost Markov decision processes with unbounded costs. *Annals of Operations Research*, 28(1):261–271, 1991.

[52] L. I. Sennott. The convergence of value iteration in average cost Markov decision chains. *Operations Research Letters*, 19(1):11 – 16, 1996.

[53] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[54] T. J. Sheskin. *Markov Chains and Decision Processes for Engineers and Managers*. CRC Press, 2010.

[55] O. Sigaud and O. Buffet. *Markov Decision Processes in Artificial Intelligence*. Wiley-IEEE Press, 2010.

[56] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry. Kalman filtering with intermittent observations. *Automatic Control, IEEE Transactions on*, 49(9):1453–1464, Sept 2004.

[57] A. Wald. *Sequential Analysis*. John Wiley & Sons, 1947.

[58] W. Wu and A. Arapostathis. Optimal sensor querying: General Markovian and LQG models with controlled observations. *Automatic Control, IEEE Transactions on*, 53(6):1392–1405, July 2008.

[59] C. Zhang, C. Li, and J. Zhang. A secure privacy-preserving data aggregation model in wearable wireless sensor networks. *Journal of Electrical and Computer Engineering*, 2015, 2015. Article ID 104286.

Vita

Johnson Carroll was born to an expatriate American family in Ipswich, England, and moved several times as a child before settling in his ancestral home of Texas. He earned undergraduate degrees in Electrical Engineering, Mathematics, and Plan II Honors, followed by an M.S. in Electrical Engineering, all at the University of Texas at Austin. He qualified for the PhD program at the University of Texas at Austin in 2006, then disrupted his studies by moving to South Africa in 2009. He is currently a Senior Lecturer in the Faculty of Engineering and the Built Environment at the University of Johannesburg, South Africa.

Johnson's research interests have developed from nonlinear and hybrid control systems to controlled stochastic processes and optimization, and have recently branched out to include engineering education and curriculum design.

Permanent address: 105 Crest View Drive
Lakeway, Texas 78734

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.