10 STABILITY, PERFORMANCE EVALUATION, AND OPTIMIZATION

Sean P. Meyn

Abstract: The theme of this chapter is stability and performance approximation for MDPs on an infinite state space. The main results are centered around stochastic Lyapunov functions for verifying stability and bounding performance. An operator-theoretic framework is used to reduce the analytic arguments to the level of the finite state-space case.

10.1 INTRODUCTION

10.1.1 *Models on a general state space*

This chapter focuses on stability of Markov chain models. Our main interest is the various relationships between stability; the existence of Lyapunov functions; performance evaluation; and existence of optimal policies for controlled Markov chains. We also consider two classes of algorithms for constructing policies, the policy and value iteration algorithms, since they provide excellent examples of the application of Lyapunov function techniques for ψ-irreducible Markov chains on an uncountable state space.

Considering the importance of these topics, it is not surprising that considerable research has been done in each of these directions. In this chapter we do not attempt a survey of all existing literature, or present the most comprehensive results. In particular, only the average cost optimality criterion is treated, and the assumptions we impose imply that the average cost is independent of the starting point of the process. By restricting attention in this way we hope that we can make the methodology more transparent.

One sees in several chapters in this volume that the generalization from finite state spaces to countable state spaces can lead to considerable technicalities. In particular, invariant distributions may not exist, and the cost functions of interest may not take on finite values. It would be reasonable to assume that the move from countable state spaces, to MDPs on a general state space should be at least as difficult. This assumption is probably valid if one desires a completely general theory.

However, the MDPs that we typically come across in practice exhibit structure which simplifies analysis, sometimes bringing us to the level of difficulty found in the countable, or even the finite state space case. For example, all of the specific models to be considered in this chapter, and most in this volume, have some degree of spatial homogeneity. The processes found in most applications will also exhibit some level of continuity in the sense that from similar starting points, and similar control sequences, two realizations of the model will have similar statistical descriptions. We do not require strong continuity conditions such as the strong Feller property, although this assumption is sometimes useful to establish existence and uniqueness of solutions to the various static optimization problems that arise in the analysis of controlled Markov chains. An assumption of ψ-irreducibility, to be described and developed below, allows one to lift much of the stability theory in the discrete state space setting to models on a completely general, non-countable state space. This is an exceptionally mild assumption on the model and, without this assumption, the theory of MDPs on a general state space is currently extremely weak.

10.1.2 An operator-theoretic framework

When $\boldsymbol{x}$ is ψ-irreducible it is possible to enlarge the state space to construct an *atom* $\boldsymbol{\theta} \in \mathbb{X}$ which is reachable from any initial condition (i.e. $\mathbb{P}_x\{\tau_\theta < \infty\} > 0$, $x \in \mathbb{X}$). When the atom is *recurrent*, that is, $\mathbb{P}_\theta\{\tau_\theta < \infty\} = 1$, then an invariant measure (see (10.9)) is given by

$$\mu\{Y\} = \mathbb{E}_\theta\Big[\sum_{t=1}^{\tau_\theta} \mathbf{1}_Y(x_t)\Big], \qquad Y \in \mathbb{F}, \tag{10.1}$$

where τ_θ is the first return time to $\boldsymbol{\theta}$ (see (10.6)), and $\mathbf{1}_Y$ is the indicator function of the set Y. This construction, and related results may be found in [45, 39]. In words, the quantity $\mu\{Y\}$ expresses the mean number of times that the chain visits the set Y before returning to $\boldsymbol{\theta}$. This expression assumes that τ_θ is almost surely finite. If the *mean* return time $\mathbb{E}_\theta[\tau_\theta]$ is finite then in fact the measure μ is finite, and it can then be normalized to give an invariant probability measure. Finiteness of the mean return time to some desirable state is the standard stability condition used for Markov chains, and for MDPs in which one is interested in the average cost optimality criterion.

Unfortunately, the split chain construction is cumbersome when developing a theory for controlled Markov chains. The sample path interpretation given in (10.1) is appealing, but it will be more convenient to work within an operator-theoretic framework, following [45]. To motivate this, suppose first that we remain in the previous setting with an uncontrolled Markov chain, and suppose that do have an state $\boldsymbol{\theta} \in \mathbb{X}$ satisfying $\mathbb{P}_\theta\{\tau_\theta < \infty\} = 1$. Denote by s the function which is equal to one at $\boldsymbol{\theta}$, and zero elsewhere: That is, $s = \mathbf{1}_\theta$. We let ν denote the probability measure on $\mathbb{X}$ given by $\nu(Y) = p(Y \mid \boldsymbol{\theta})$, $Y \in \mathbb{F}$, and define the 'outer product' of s and ν by

$$s \otimes \nu\,(x, Y) := s(x)\nu(Y).$$

For example, in the finite state space case the measure ν can be interpreted as a row vector, the function s as a column vector, and $s \otimes \nu$ is the standard (outer) product of these two vectors. Hence $s \otimes \nu$ is an $N \times N$ matrix, where N is the number of states.

In general, the *kernel* $s \otimes \nu$ may be viewed as a rank-one operator which maps L_∞ to itself, where L_∞ is the set of bounded, measurable functions on $\mathbb{X}$. Several other bounded linear operators on L_∞ will be developed in this chapter. The most basic are the n-step transition kernels, defined for $n \geq 1$ by

$$P^n f\,(x) := \int_{\mathbb{X}} f(y) p^n(dy|x), \qquad f \in L_\infty,$$

where $p^n(\,\cdot \mid \cdot\,)$ is the n-step state transition function for the chain. We set $P = P^1$. We can then write, in operator-theoretic notation,

$$\mathbb{P}_\theta\{\tau_\theta \geq n, x_n \in Y\} = \nu(P - s \otimes \nu)^{n-1}\mathbf{1}_Y, \qquad n \geq 1,$$

and hence the invariant measure μ given in (10.1) is expressed in this notation as

$$\mu(Y) = \sum_{n=1}^{\infty} \nu(P - s \otimes \nu)^{n-1}\mathbf{1}_Y, \qquad Y \in \mathbb{F}. \tag{10.2}$$

It is this algebraic description of μ that will be generalized and exploited in this chapter.

How can we mimic this algebraic structure without constructing an atom $\boldsymbol{\theta}$? First, we require a function $s\colon \mathbb{X} \to \mathbb{R}_+$ and a probability measure ν on $\mathbb{F}$ satisfying the *minorization condition*,

$$p(Y \mid x) \geq s(x)\nu(Y), \qquad x \in \mathbb{X},\ Y \in \mathbb{F}.$$

In operator theoretic notation this is written $P \geq s \otimes \nu$, and in the countable state space case this means that the transition matrix P dominates an outer product of two vectors with non-negative entries.

Unfortunately, this 'one step' minorization assumption excludes a large class of models, even the simple linear models to be considered as examples below. One can however move to the resolvent kernel defined by

$$K = (1 - \beta) \sum_{t=0}^{\infty} \beta^t P^t, \tag{10.3}$$

where $\beta \in]0, 1[$ is some fixed constant. For a ψ-irreducible chain the required minorization always holds for the resolvent K [39, Theorem 5.2.3]. The move to the resolvent is useful since almost any object of interest can be mapped between the resolvent chain, and the original Markov chain. In particular, the invariant measures for P and K coincide (see [39, Theorem 10.4.3], or consider the resolvent equation in (10.11) below).

Much of the analysis then will involve the *potential kernels*, defined via

$$G := \sum_{t=0}^{\infty} K^t. \tag{10.4}$$

$$H := \sum_{t=0}^{\infty} (K - s \otimes \nu)^t. \tag{10.5}$$

In Theorem 10.1 below we demonstrate invariance of the σ-finite measure μ defined by,

$$\mu(Y) = \int_{\mathbb{X}} \nu(dx) H(x, Y), \qquad Y \in \mathbb{F},$$

provided the chain is *recurrent*. The invariant measure can be written in the compact form $\mu = \nu H$.

The measure μ will be *finite*, rather than just σ-finite, provided appropriate stability conditions are satisfied. The most natural stability assumption is equivalent to the existence of a Lyapunov function, whose form is very similar to the Poisson equation found in the average cost optimality equation. The development of these connections is one of the main themes of this chapter.

10.1.3 Overview

We conclude with an outline of the topics to follow. In the next section we review a bit of the general theory of ψ-irreducible chains, and develop some stochastic Lyapunov theory for such chains following [39, Chapters 11-14]. Following this, in Section 10.3 we develop in some detail the computation of the average cost through the Poisson equation, and the construction of bounds on the average cost. All of these results are developed for time homogeneous chains without control.

In Section 10.4 this stability theory is applied to the analysis of the average cost optimality equation (ACOE). We explore the consequences of this equation, and derive criteria for the existence of a solution.

Section 10.5 concerns two recursive algorithms for generating solutions to the ACOE: value iteration and policy iteration. It is shown that (i) either algorithm generates stabilizing stationary policies; (ii) for any of these policies, the algorithms generate uniform bounds on steady state performance. However, such results hold only if the algorithms are properly initialized.

Convergence is established for the policy iteration algorithm: Under suitable conditions, and when properly initialized, the algorithm converges to a solution of the ACOE.

Section 10.6 illustrates the theory with a detailed application to linear models, and to network scheduling.

This chapter is concluded with a discussion of some extensions and open problems.