ABSTRACT

Title of dissertation: MAXIMAL PERSISTENT SURVEILLANCE AND OPTIMAL SENSOR USAGE IN DENIED ENVIRONMENTS

Eduardo Romero Arvelo, Doctor of Philosophy, 2017

Dissertation directed by: Professor Nuno Martins
Department of Electrical and Computer Engineering

Research in monitoring and surveillance has flourished in recent years. Its applications include control of illegal deforestation, search for survivors in disasters, and inspection of large infrastructures, among others. Some of the current challenges lie in establishing control policies that are suited for systems with low power and limited sensing, actuation and communication capabilities. This thesis has two main focuses: *i*) control design for persistent surveillance, where the goal is to design memoryless policies that achieve surveillance of the largest possible area, while respecting certain constraints; and *ii*) optimal sensor usage for monitoring of denied environments, inside which state observations are costly.

In the first part of the thesis, we design control policies for Markov decision processes (MDP) with the objective of generating the maximal set of recurrent states, subject to convex constraints on the set of invariant probability mass functions. We propose a design method for memoryless policies and fully observable MDPs with finite state and action spaces. Our approach relies on a finitely parametrized convex program inspired by principles of entropy maximization.

Next, we explore the problem of designing controllers for autonomous robots tasked with maximal persistent surveillance of an area in which there are forbidden regions. We model each robot as an MDP whose state comprises its position on a finite two- dimensional lattice and the direction of motion. The goal is to find the minimum number of robots and an associated time-invariant memoryless control policy that guarantees that the largest number of states are persistently surveilled without ever visiting a forbidden state.

In the second part of the thesis, we study the problem of optimal sensor usage in denied environments, inside which state observations are only available by incurring an extra cost. Observations outside the denied environment are cost free. The goal is to understand the trade-off between paying to access the sensor immediately, and waiting for a free sensor use should the system exit the denied environment. We show that the analysis of this problem simplifies by recasting it as renewal reward process, which enables us to establish conditions for which any local minimum (if it exists) is also a global minimum, thus facilitating the search for its minimizer.

Finally, we extend these results to the case when the state space is discrete and the state update is ruled by a Markov chain. We establish conditions on the initial distribution of the Markov chain that guarantee that any local minimum, if it exists, is also global. In particular, these conditions rely on constraining the radial stochastic order of an auxiliary Markov process. By analyzing these problems, we hope to provide valuable insights into the core of some of the challenges that may arise in real-world implementations.

Maximal Persistent Surveillance and Optimal Sensor Usage in Denied Environments

by

Eduardo Romero Arvelo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Nuno Martins, Chair/Advisor
Professor Gilmer Blankenship
Professor Pamela Abshire
Professor Timothy Horiuchi
Professor Nikhil Chopra

Table of Contents

List of Figures

Chapter 1: Introduction

Research in monitoring, surveillance and reconnaissance has flourished with advancements that enabled the development of autonomous agents capable of undertaking such tasks [1–4]. Applications in this field include environmental monitoring [5–7], emergency response to disasters [8, 9], risk assessment of sensitive situations [10], and inspection of large aging infrastructures [11], among others. Although technology continues to evolve, allowing for ever more powerful systems, some of the current challenges lie in developing platforms that are able to cope with limited sensing, actuation, communication and processing capabilities [12–16]. Solving these challenges are key, for example, in the context of miniature robotics, where power remains a limiting factor [17].

This thesis has two main focuses: *i*) the study of control design for persistent surveillance, where the goal is to design policies that achieve surveillance of the largest possible area, while respecting certain constraints. We focus on the design of *memoryless* policies, which require minimal on-board processing; and *ii*) the problem of sensor usage for monitoring of denied environments, inside which state observations are costly. Here, the goal is to study the trade-off between sensing and not sensing and determine the optimal policy for sensor usage.

The guiding philosophy of our approach is to theoretically analyze seemingly simple problems and, by doing so, to unearth *fundamental* concepts and design principles. With this strategy, we hope to provide valuable insights into the core of some of the challenges that may arise in real-world implementations.

1.1 Maximal Persistent Surveillance

In the first part of the thesis, we examine a problem in maximal persistent surveillance, which consists of finding a memoryless control policy for an autonomous agent whose goal is to persistently visit the *largest* possible set of locations in a given area. The concept of persistent surveillance is similar to the concept of coverage [18], but differs in that the area to be surveilled must be revisited infinitely many times. In our setup, we also impose safety constraints that dictate certain *forbidden* regions.

In our approach, we model each agent as a fully-observed Markov decision process (MDP) with finite state and control spaces. The formalism of MDPs is widely used to describe the behavior of systems whose state transitions probabilistically among different configurations over time. The impact of a control policy is felt through the actions that dictate the state transition probabilities. While the traditional framework of MDPs involve a wide variety of costs that depend linearly on the parameters that characterize the probabilistic behavior of the system, we take a different approach, as we do not impose a stage cost for visiting a state or using a control action. Rather, we are concerned with *finding a control policy that leads to the maximal set of recurrent states, subject to convex constraints on the set*

of invariant joint probability mass function (`pmf`*) on the state and control action.* Examples of constraints of interest include the expected value of a function of the state and(or) action, and lower and upper bounds on invariant `pmf`s evaluated at pre-selected state and action pairs. These constraints can be used, for example, to prohibit the system from ever visiting a certain state, or to increase the proportion of time during which a particular region is visited as the system evolves.

One of the properties of our approach is that the agents' states and actions lie in discrete sets. As we will see in Chapter 3, each state may represent the agents' position and orientation on a square lattice we wish to persistently surveil. In this setup, one can think of a vehicle surveilling a city where the streets are arranged in a grid, an at each intersection (state of the MDP), the vehicle has to determine whether to continue straight or turn (actions of the MDP).

The square-grid scenario, however, is only one of numerous others that can be implemented with the MDP formulation, and it should be regarded only as a proof of concept of our results. The square lattice may be generalized to, for example, the rigid structure of a bridge, where agents are robots that move on girders and each point where a girder meets another may be construed as a state of the MDP. The robots must decide at each girder intersection which path to follow.

Alternatively, in a scenario where the robots travel in a continuous space, the states of the MDP may be abstracted as regions of the space, which has been partitioned for surveillance. For example, in the monitoring of a building each room may be thought of as a state of MDP, and after monitoring a room the robot must decide which room to visit next. The partitioning of the space may be coarse

(such as in the building scenario) or it can be fine, effectively discretizing the space. This approach, however, should be followed with caution due to state-action pair explosion.

1.1.1 Related Literature

Control design for persistent surveillance has been previously studied and many problem formulations and approaches have been proposed. In [19, 20], the authors present a semi-heuristic control policy that minimizes the time between visitations to the same region. In [21], an algorithmic approach for persistent surveillance of a convex polygon in the plane is provided. In [22], communication constraints and sensor failures are incorporated to the problem of persistent surveillance and approximate dynamic programing is used. In [23], the authors describe a dynamic programming approach with temporal logic specifications, which can be used to cast persistent surveillance problems. Speed control for robots performing persistent monitoring on a predetermined path is addressed in [24], where linear programing techniques are employed. On the implementation front, system architectures for unmanned aerial vehicles have been designed specifically for persistent surveillance purposes, such as in [25]. In our work, we focus on memoryless policies (found via convex optimization) that achieve persistent surveillance of the largest possible set of locations without violating safety constraints. These semi-heuristic approaches, however, are not restricted to memoryless policies and do not consider safety constraints.

To our knowledge, the problem of maximizing the set of recurrent states of an MDP under convex constraints had not yet been studied. The most similar framework appears in a series of papers by Arapostathis et al., where the state probability distribution is restricted to be bounded above and below by safety vectors at all times. In [26], [27] and [28], the authors propose algorithms to find the set of distributions whose evolution under a given control policy respect the safety constraint. In [29], an augmented Markov chain is used to find the the maximal set of probability distributions whose evolution respect the safety constraint over all admissible non-stationary control policies.

1.1.2 Contributions

We solve the problem of maximizing the set of recurrent states of an MDP by casting a finitely parametrized convex program, which can be easily implemented using standard convex optimization tools. The proposed optimization program maximizes the entropy of the joint `pmf` with the constraint that the `pmf` be invariant. The choice of the entropy as the objective function is rooted on the fact that the `pmf` that maximizes entropy under convex constraints also maximizes the support. The convexity of our approach is achieved by casting the invariance equation as a linear function of the `pmf`.

We apply our method to design memoryless controllers for robots that move in a finite two-dimensional lattice with the goal of achieving persistent surveillance. The concept of persistent surveillance in this context is analogous to the concept of

recurrence in Markov chains. In this setup, we also impose safety constraints that dictate that certain regions are *forbidden*. The forbidden regions may represent areas in which robots cannot operate (such as bodies of water) or are not allowed to visit (such as restricted airspace). The goal is to deploy the minimum number of robots equipped with a control policy that guarantees persistent surveillance of the largest possible set of lattice points without ever visiting a forbidden region. The results are illustrated by numerical examples.

The work on maximal persistent surveillance has been published in [30] and [31].

1.2 Optimal Sensor Usage in Denied Environments

In the second part of the thesis, we study the problem of optimal sensor usage in denied environments. In our setup, state observations inside denied environments are only available by paying a price in the cost to be minimized, whereas outside the environment observations are free. One can imagine an agent whose goal is to monitor a desired location inside the denied environment, but disturbances may cause it to wander away. The agent can only return to the desired location once it observes its current location. In this scenario, there are two ways the agent accesses its location: by drifting outside the denied environment, where observations are readily available; or by paying a price, which can be thought of as a communication or energy cost, to receive an observation of its location.

Our goal is to understand the trade-off between: i) waiting for a free sensor

use in the future and *ii*) paying a penalty to access the sensor immediately. The agent is modeled as a discrete-time Markov process that evolves in time inside of the denied environment, and at each time step, the only control available is the decision whether or not to reset the process back to its initial distribution (and by doing so incurring an extra cost). The optimal policy needs to take into account the prospect of a cost-free reset happening in the future, should the process exit the denied environment.

1.2.1 Related Literature

Research in monitoring and surveillance of denied environments has a rich literature. On the implementation side, most of the focus is in navigation and path planning for unmanned aerial vehicles (UAV) in GPS-denied environments [32]. The solutions that have been proposed usually rely on on-board sensing capabilities, such as vision [33–35] and optical sensors [36,37]. The work in [38] proposes a scheme for collaborative navigation in multi-agent systems. In this thesis, we work on a simple problem formulation to determine the optimal decision of whether to access a costly sensor (in the absence of any other sensing capability) with the prospect of having a free sensor access in the future, should the agent escape the denied environment.

On the theoretical front, efforts in event-based control and estimation aim at establishing protocols for which actuating, computing and sensing are undertaken only when needed, rather than periodically [39–42]. The work in [43], for example, proposes a sporadic control scheme for first order linear systems. In our work, when

the agent exists the denied environment and receives a state observation can be thought of as event-based sensing. The difference is that the event under consideration is a problem constraint rather than a design parameter, which is generally the case in the event-based control and estimation literature.

A deceptively similar framework is that of control and estimation over costly or limited communication channels. These setups occur in the context of networked systems, where different components of systems are not collocated. The authors of [44] consider an estimation problem where only a limited number of observations are available over a finite horizon. The work in [45] proposes an estimation framework in which a pre-processor decides whether or not to transmit measurements to the estimator. The most similar problem setup to ours is the one in [46, 47], where the authors consider an estimation problem in which the system can be reset by paying a cost, and where the reset policy is a function of the current state of the system.

There are two main features of our work that distinguishes it from the ones mentioned in the previous paragraph: *i*) in our work, the decision of whether or not to use the sensor cannot be a function of the state, which is in general the case in networked estimation problems; and *ii*) in our setup, we consider "event-based", cost-free sensor usage, which introduces an incentive to not request a costly observation. To our knowledge, a framework that combines these two characteristics has not been studied before.

1.2.2 Contributions

We show that the analysis of this problem simplifies by recognizing that the stochastic process is in fact a renewal process, and our design parameter is the support of the stopping time associated with the renewal process. This approach enables us to establish conditions for which any local minimum, if it exists, is also global, thus facilitating the search for the minimizer. We also explore the radial stochastic order of an associated conditioned process, and establish sufficient conditions for which the previous results hold.

Furthermore, we extend these results to the case when the state space is discrete and the state update is ruled by a Markov chain. Specifically, we focus on chains whose state space is indexed by integers, and at each time step it can move, with equal probability, to a neighboring state; or stay at the current state. We establish conditions on the initial `pmf` of the Markov chain that guarantee that any local minimum, if it exists, is also global. In particular, these conditions rely on constraining the radial stochastic order of a auxiliary Markov process. The theoretical results are illustrated by numerical examples.

1.3 Thesis Organization

The thesis is organized as follows. Chapter 2 describes the theoretical approach to maximal persistent surveillance, by solving an optimization problem that aims at finding the control policy that maximizes the set of recurrent states of an MDP. We dedicate Chapter 3 to the application of these results to persistent surveillance.

In Chapter 4, we explore the problem of sensor usage in denied environments for stochastic processes in $\mathbb{R}^n$. Chapter 5 extends the results of the previous chapter to Markov Chains and focuses on establishing sufficient conditions on the initial `pmf` of the Markov Chain. Finally, conclusions and future directions are discussed in Chapter 6.

Chapter 2: Control Polcies that Maximize the Set of Recurrent States in MDPs subject to Convex Constraints

The formalism of Markov decision processes (MDPs) is widely used to describe the behavior of systems whose state transitions probabilistically among different configurations over time. The impact of a control policy is felt through the actions that dictate the state transition probabilities. The traditional framework of MDPs involve a wide variety of costs that depend linearly on the parameters that characterize the probabilistic behavior of the system. Typically, at each time step, the system incurs a cost (or reward) associated with the current state and control action, and the goal is to devise a control policy that minimizes the expected cost over a predetermined time interval. The two most commonly adopted approaches to tackle these problems are linear and dynamic programming [48, 49].

In this chapter we take a different look at MDPs, as we do not impose a stage cost for visiting a state or using a control action. Rather, we are concerned with *finding a control policy that leads to the maximal set of recurrent states, subject to convex constraints on the set of invariant joint probability mass function (*`pmf`*) on the state and control action*. Examples of constraints of interest include lower and upper bounds on invariant `pmf`s evaluated at pre-selected state and action pairs, or

on the expected value of a function of the state and(or) action. These constraints can be used, for example, to prohibit the system from ever visiting a certain state, or to increase the proportion of time during which a particular region is visited as the system evolves. The motivation for recurrence comes from applications in surveillance, where it is desirable that the system visits as many states as possible to improve information gathering.

We consider fully observable MDPs with finite state and action spaces, and limit our search to memoryless policies. We efficiently solve this problem by casting a finitely parametrized convex program, which can be easily implemented using standard convex optimization tools, such as [50]. The proposed optimization program maximizes the entropy of the joint `pmf` with the constraint that the `pmf` is invariant. The choice of the entropy as the objective function is rooted on the fact that the `pmf` that maximizes entropy under convex constraints also maximizes the support. The convexity of our approach is achieved by casting the invariance equation as a linear function of the `pmf`. We also show that, by maximizing the entropy of the joint `pmf` (rather than the entropy of the marginal `pmf` with respect to the state), we achieve the maximal set of recurrent states with the *least* number of recurrent classes.

Once a control policy is applied to an MDP, one can construct a directed graph of transitions for the resulting Markov chain. Here, the vertices of the graph are the states and an edge from i to j indicates that the transition from i to j has positive probability. In this case, the set of recurrent states is the union of the strongly connected components that are closed, each representing a recurrent class. Hence,

one could use standard algorithms [51], such as Kosaraju's or Tarjan's, to efficiently find the strongly connected components of the graph and perform the union of the ones that are closed. However, a control policy that "maximizes" the set of recurrent states cannot be easily obtained using this method because the graph of transitions may change for different candidate solutions. Furthermore, because we consider constraints on the set of invariant `pmf`s, both the maximal set of recurrent states and the corresponding control policies will, in general, depend on the actual values of the entries of the transition probability matrices.

Broadly speaking, reachability is concerned with the determination of whether a set of states can be reached from another via an appropriate control policy [52]. There are two main reasons why our formulation cannot be cast as a reachability problem. The first is that reachability is in general distinct from recurrence, particularly when a reachable state is transient. The second follows from the discussion above, where we emphasize that optimal solutions may depend on the probabilities of the transitions, and not only on whether which ones occur with nonzero probability.

To our knowledge, the problem of maximizing the set of recurrent states under convex constraints has not yet been studied. The most similar framework appears in a series of papers by Arapostathis et al., where the state probability distribution is restricted to be bounded above and below by safety vectors at all times. In [26], [27] and [28], the authors propose algorithms to find the set of distributions whose evolution under a given control policy respect the safety constraint. In [29], an augmented Markov chain is used to find the the maximal set of probability

distributions whose evolution respect the safety constraint over all admissible non-stationary control policies.

The chapter is organized as follows. Section 2.1 provides basic definitions and the problem statement. The convex program that solves the problem is presented in Section 2.2. Numerical examples are given in Section 2.3, while Section 2.4 discusses results concerning multiple recurrent classes. Conclusions are given in Section 2.5.

2.1 Preliminaries and Problem Statement

The following notation is used throughout the chapter:

$\mathbb{X}$	state space of the MDP
$\mathbb{U}$	set of control actions
X_k	state of the MDP at time k
U_k	control action at time k
$\mathbb{P}_{\mathbb{X}}$	set of all `pmf`s with support in $\mathbb{X}$
$\mathbb{P}_{\mathbb{U}}$	set of all `pmf`s with support in $\mathbb{U}$
$\mathbb{P}_{\mathbb{X}\mathbb{U}}$	set of all joint `pmf`s with support in $\mathbb{X} \times \mathbb{U}$
$\mathbb{S}_f$	support of a `pmf` f

The state and control action of the MDP at time k are given by X_k and U_k, respectively. The recursion of the MDP is given by the (conditional) `pmf` of X_{k+1} given the previous state X_k and control action U_k, and is denoted as:

$$\mathcal{Q}(x^+, x, u) \stackrel{def}{=} P(X_{k+1} = x^+ | X_k = x, U_k = u).$$

We denote any time-homogeneous control policy by

$$\mathcal{K}(u,x) \stackrel{def}{=} P(U_k = u | X_k = x), \qquad u \in \mathbb{U}, x \in \mathbb{X}$$

where $\sum_{u \in \mathbb{U}} \mathcal{K}(u,x) = 1$ for all x in $\mathbb{X}$. The set of all such policies is denoted as $\mathbb{K}$.

Assumption Throughout this chapter, we assume that the MDP $\mathcal{Q}$ is given. Hence, all quantities and sets that depend on the closed loop behavior are indexed only by the underlying control policy $\mathcal{K}$.

A `pmf` f_{XU} in $\mathbb{P}_{\mathbb{XU}}$ is said to be *invariant* under control policy $\mathcal{K}$ if it satisfies the following invariance relation:

$$f_{XU}(x^+,u^+) = \mathcal{K}(u^+,x^+) \sum_{x \in \mathbb{X}, u \in \mathbb{U}} \mathcal{Q}(x^+,x,u) f_{XU}(x,u), \tag{2.1}$$

for all x^+ in $\mathbb{X}$ and u^+ in $\mathbb{U}$. The set of invariant `pmf`s associated with control policy $\mathcal{K}$ is given by:

$$\mathbb{I}_{\mathcal{K}} \stackrel{def}{=} \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : (2.1) \text{ holds with control policy } \mathcal{K} \Big\}$$

Finally, the set of all invariant `pmf`s are given by:

$$\mathbb{I} \stackrel{def}{=} \bigcup_{\mathcal{K} \in \mathbb{K}} \mathbb{I}_{\mathcal{K}}$$

Problem 2.1.1. *Given $\mathbb{W}$, which is a convex subset of $\mathbb{P}_{\mathbb{XU}}$, find a joint* `pmf` f^*_{XU} *in $\mathbb{I} \cap \mathbb{W}$ and a corresponding control policy $\mathcal{K}^*$ such that the following inclusion holds:*

$$\mathbb{S}^{\mathbb{X}}_{f_{XU}} \subseteq \quad \mathbb{S}^{\mathbb{X}}_{f^*_{XU}}, \quad f_{XU} \in \mathbb{I} \cap \mathbb{W}; \tag{2.2}$$

where $\mathbb{S}^{\mathbb{X}}_f = \{x \in \mathbb{X} | \sum_{u \in \mathbb{U}} f(x,u) > 0\}$.

Remark Note that a `pmf` f^*_{XU} that satisfies (2.2) has maximal support among all members of the set $\mathbb{I} \cap \mathbb{W}$.

The convex set $\mathbb{W}$ can be used to cast a wide variety of constraints. For example, suppose the set $\mathbb{F}$ in $\mathbb{X}$ contains the states we wish the system never visits. This requirement can be cast in the following constraint on f_{XU}:

$$\mathbb{W} = \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : \sum_{u \in \mathbb{U}} f_{XU}(x,u) = 0, x \in \mathbb{F} \Big\}.$$

2.2 Solution via Convex Optimization

We propose the following convex program to solve Problem 2.1.1:

$$f^*_{XU} = \arg \max_{f_{XU} \in \mathbb{W}} \mathcal{H}(f_{XU}) \tag{2.3}$$

subject to:

$$\sum_{u^+ \in \mathbb{U}} f_{XU}(x^+, u^+) = \sum_{x \in \mathbb{X}, u \in \mathbb{U}} \mathcal{Q}(x^+, x, u) f_{XU}(x,u), \quad x^+ \in \mathbb{X} \tag{2.4}$$

where $\mathcal{H} : \mathbb{P}_{\mathbb{XU}} \to \Re_{\geq 0}$ is the entropy of f_{XU}, and is given by

$$\mathcal{H}(f_{XU}) = -\sum_{u \in \mathbb{U}} \sum_{x \in \mathbb{X}} f_{XU}(x,u) \ln(f_{XU}(x,u)),$$

where we adopt the standard convention that $0 \ln(0) = 0$.

The following theorem states precisely how the proposed convex program provides a solution to Problem 2.1.1.

Theorem 2.2.1. *Let $\mathbb{W}$ be given, and assume that (2.3)-(2.4) is feasible and that f^*_{XU} is the optimal solution. In addition, adopt the marginal* `pmf` $f^*_X(x) = \sum_{u \in \mathbb{U}} f^*_{XU}(x,u)$ *and let $\mathcal{G} : \mathbb{U} \times \mathbb{X} \to [0,1]$ be any function satisfying $\sum_{u \in \mathbb{U}} \mathcal{G}(u,x) = 1$ for all x in $\mathbb{X}$. The following holds:*

(a) $f^*_{XU} \in \mathbb{W} \cap \mathbb{I}$, *with* $\mathcal{K}^*$ *given by:*

$$\mathcal{K}^*(u,x) = \begin{cases} \frac{f^*_{XU}(x,u)}{f^*_X(x)}, & x \in \mathbb{S}^{\mathbb{X}}_{f^*_{XU}} \\ \mathcal{G}(u,x), & otherwise \end{cases}, (u,x) \in \mathbb{U} \times \mathbb{X} \tag{2.5}$$

(b) $\mathbb{S}^{\mathbb{X}}_{f_{XU}} \subseteq \mathbb{S}^{\mathbb{X}}_{f^*_{XU}}, \quad f_{XU} \in \mathbb{I} \cap \mathbb{W},$

where $\mathbb{S}^{\mathbb{X}}_f = \{x \in \mathbb{X} | \sum_{u \in \mathbb{U}} f(x,u) > 0\}$, $f \in \mathbb{P}_{\mathbb{XU}}$.

Remark An alternative formulation of the proposed convex program that maximizes the entropy of the marginal `pmf` with respect to the state (rather than the joint `pmf`) would also provide a solution to Problem 2.1.1. As we discuss in Section 2.4, our formulation has the advantage that the resulting maximal set of recurrent states is guaranteed to have the smallest number of recurrent classes.

To facilitate the proof of the second part of Theorem 2.2.1, we introduce the following lemma:

Lemma 2.2.2. *Let* $\mathbb{Y}$ *be a finite set and* $\mathbb{V}$ *be a convex subset of* $\mathbb{P}_{\mathbb{Y}}$*, which is the set of all* `pmf`*s with support in* $\mathbb{Y}$*. Consider the following problem:*

$$f^* = \arg\max_{f \in \mathbb{V}} \mathcal{H}(f)$$

where $\mathcal{H}(f)$ *is the entropy of* f *in* $\mathbb{P}_{\mathbb{Y}}$ *and is given by* $\mathcal{H}(f) = -\sum_{y \in \mathbb{Y}} f(y) \ln(f(y))$*, where we adopt the convention that* $0 \ln(0) = 0$*. The following holds:*

$$\mathbb{S}_f \subseteq \mathbb{S}_{f^*}, \ f \in \mathbb{V}$$

where $\mathbb{S}_f = \{y \in \mathbb{Y} | f(y) > 0\}$, $f \in \mathbb{P}_{\mathbb{Y}}$.

Proof. Select an arbitrary f in $\mathbb{V}$ and define $f_\lambda \stackrel{def}{=} \lambda f^* + (1-\lambda)f$ for $0 \leq \lambda \leq 1$. From the convexity of $\mathbb{V}$, we conclude that f_λ is in $\mathbb{V}$ for all λ in $[0,1]$. Since f^* has maximal entropy, it must be that there exists a $\bar{\lambda}$ in $[0,1)$ such that

$$\frac{d}{d\lambda}\mathcal{H}(f_\lambda) \geq 0,\ \lambda \in (\bar{\lambda}, 1). \tag{2.6}$$

Proof by contradiction: Suppose that $\mathbb{S}_f \nsubseteq \mathbb{S}_{f^*}$ and hence that there exists a y' in $\mathbb{Y}$ such that $f(y') > 0$ and $f^*(y') = 0$. We have that

$$\frac{d}{d\lambda}\Big(f_\lambda(y')\ln(f_\lambda(y'))\Big) = -f(y')\big(\ln(f_\lambda(y')) + 1\big)$$

goes to ∞, as λ approaches 1, since $\lim_{\lambda\to 1} f_\lambda(y') = 0$. This implies that there exists a $\tilde{\lambda}$ in $[0,1)$ such that

$$\frac{d}{d\lambda}\mathcal{H}(f_\lambda) < 0,\ \lambda \in (\tilde{\lambda}, 1),$$

which contradicts (2.6). □

See [53] for an alternative proof that relies on the concept of relative entropy.

Proof of Theorem 2.2.1.

(a) (Proof that $f^*_{XU} \in \mathbb{W} \cap \mathbb{I}$, with $\mathcal{K}^*$ given by (2.5)) The inclusion of f^*_{XU} in $\mathbb{W}$ follows from the assumption that (2.3)-(2.4) is feasible. To show that f^*_{XU} belongs to $\mathbb{I}$ with control policy given by (2.5), we first note that by feasibility it must satisfy the constraint in (2.4). For each pair (x^+, u^+) in $\mathbb{X}\times\mathbb{U}$, we multiply both sides of (2.4) by $\mathcal{K}^*(u^+, x^+)$:

$$\mathcal{K}^*(u^+, x^+)\sum_{u^+\in\mathbb{U}} f^*_{XU}(x^+, u^+) = \tag{2.7}$$

$$\mathcal{K}^*(u^+, x^+)\sum_{x\in\mathbb{X}, u\in\mathbb{U}} \mathcal{Q}(x^+, x, u) f^*_{XU}(x, u),\ x^+ \in \mathbb{X}$$

The right-hand side of (2.7) matches the right-hand side of the invariance relation given by (2.1), and for each x^+ in $\mathbb{S}^{\mathbb{X}}_{f^*_{XU}}$, the left-hand side is equal to $f^*_{XU}(x^+, u^+)$, and therefore the invariance relation in (2.1) is satisfied. Moreover, for values of x^+ not in $\mathbb{S}^{\mathbb{X}}_{f^*_{XU}}$, the invariance relation is trivially satisfied. Therefore, f^*_{XU} belongs to $\mathbb{I}$ with control policy given by (2.5).

(b) (Proof that $\mathbb{S}^{\mathbb{X}}_{f_{XU}} \subseteq \mathbb{S}^{\mathbb{X}}_{f^*_{XU}}$, $f_{XU} \in \mathbb{I} \cap \mathbb{W}$) Follows by noting that the set of feasible `pmf`s of the optimization program (2.3)-(2.4) is convex and applying Lemma 2.2.2.

□

Remark The implementation of the optimization program described in (2.3)-(2.4) can be done with any appropriate convex optimization solver. Throughout Chapters 2 and 3, we provide numerical examples that were cast and solved using on CVX [50], which relies on the software package SDPT3 as its default solver [54]. Each call to SDPT3 can be solved with $\mathcal{O}\big((nm)^{3.5} \log \epsilon^{-1}\big)$ iterations [55], where n is the number of states in the MDP, m is the number of control actions, and ϵ is the accepted duality gap. It is important to note that CVX relies on a successive approximation method due to the logarithmic nature of the objective function. In our experience, however, CVX was able to reliably solve all numerical problems we tested.

2.3 Simple Numerical Examples

Example 2.3.1. *We use our method to solve Problem 2.1.1 for the case when* $\mathbb{X} = \{1, \ldots, 7\}$ *and* $\mathbb{U} = \{1, 2\}$*, and we adopt an MDP specified by the following*

probability transition matrices:

$$Q^1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .6 & 0 & .4 \end{bmatrix}, \quad Q^2 = \begin{bmatrix} .4 & .6 & 0 & 0 & 0 & 0 & 0 \\ .3 & .7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & .3 & 0 & 0 & 0 & .7 \\ 0 & 0 & 0 & 0 & 0 & .5 & .5 \end{bmatrix}$$

where we use the notation $Q^u_{ij} \stackrel{def}{=} \mathcal{Q}(i,j,u)$. *The directed graph of possible transitions of the MDP is depicted in Fig. 2.1.*

We proceed to solving Example 2.3.1, subject to four distinct convex constraints on the invariant `pmf`s and discuss the resulting changes on the control policy and, in some cases, also the variation on the maximal set of recurrent states[1]. For clarity of exposition, all constraints are imposed on the invariant probability of state 4.

Notation: In what follows, $F^*_{ij} \stackrel{def}{=} f^*_{XU}(i,j)$, $K^*_{ij} \stackrel{def}{=} \mathcal{K}^*(i,j)$, and $G \in [0,1]^{2\times 7}$ is any matrix whose columns sum up to 1.

1. The following is the solution of Example 2.3.1, with

$$\mathbb{W} = \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : \sum_{u\in\mathbb{U}} f_{XU}(4,u) \geq .\mathbf{2} \Big\}.$$

[1]Our results are rounded to two decimal points, or represented as fractions where appropriate.

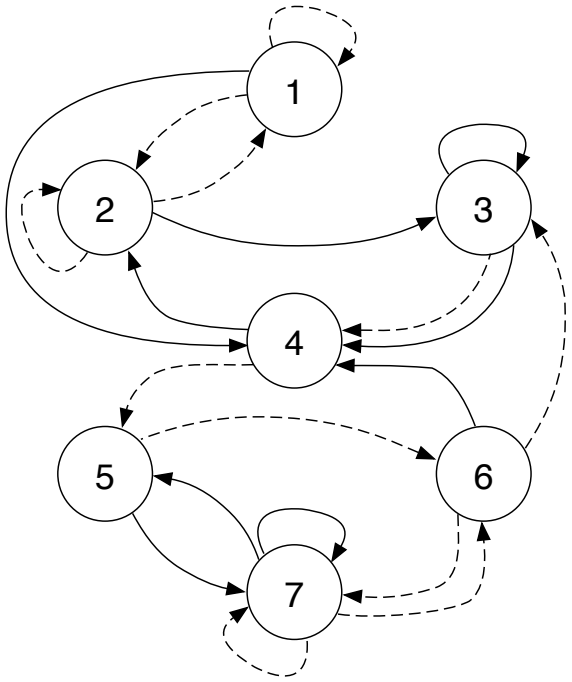


Figure 2.1: Example of an MDP with seven states and two control actions. The dashed and solid lines represent transitions with nonzero probability for control actions 1 and 2, respectively.

$$\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7\}$$

$$F^* = \begin{bmatrix} .02 & .09 & .07 & \mathbf{.11} & .05 & .08 & .07 \\ .02 & .11 & .07 & \mathbf{.09} & .09 & .05 & .08 \end{bmatrix},$$

$$K^* = \begin{bmatrix} .42 & .45 & .49 & .53 & .38 & .62 & .47 \\ .58 & .55 & .51 & .47 & .62 & .38 & .53 \end{bmatrix}$$

2. The following is the solution of Example 2.3.1, with

$$\mathbb{W} = \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : \sum_{u \in \mathbb{U}} f_{XU}(4, u) \geq \mathbf{.25} \Big\}.$$

$$\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7\}$$

$$F^* = \begin{bmatrix} .02 & .11 & .06 & \mathbf{.13} & .04 & .11 & .04 \\ .01 & .09 & .09 & \mathbf{.12} & .10 & .02 & .06 \end{bmatrix},$$

$$K^* = \begin{bmatrix} .70 & .57 & .40 & .53 & .28 & .84 & .38 \\ .30 & .43 & .60 & .47 & .72 & .16 & .62 \end{bmatrix}$$

Remark While the maximal set of recurrent states remained unchanged with respect to the previous case, the control policy had to change to accommodate the different constraint.

3. The following is the solution of Example 2.3.1, with

$$\mathbb{W} = \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : \sum_{u \in \mathbb{U}} f_{XU}(4,u) \geq {}^{\mathbf{1}}\!/_{\mathbf{3}} \Big\}.$$

$$\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} = \{2,\ 3,\ 4,\ 5,\ 6\}$$

$$F^* = \begin{bmatrix} 0 & {}^{1}\!/_{6} & 0 & {}^{\mathbf{1}}\!/_{\mathbf{6}} & 0 & {}^{1}\!/_{6} & 0 \\ 0 & 0 & {}^{1}\!/_{6} & {}^{\mathbf{1}}\!/_{\mathbf{6}} & {}^{1}\!/_{6} & 0 & 0 \end{bmatrix},$$

$$K^* = \begin{bmatrix} G_{1,1} & 1 & 0 & .5 & 0 & 1 & G_{1,7} \\ G_{2,1} & 0 & 1 & .5 & 1 & 0 & G_{2,7} \end{bmatrix}$$

Remark In this case, states 1 and 7 disappear from the maximal set of recurrent states. The corresponding closed-loop transition graph is depicted in Fig. 2.2. Note that the following control policy would also lead to an invariant `pmf` that satisfies the constraint, however the resulting set of recurrent states

would not be maximal.

$$K = \begin{bmatrix} G_{1,1} & 1 & 0 & 1 & G_{1,5} & G_{1,6} & G_{1,7} \\ G_{2,1} & 0 & 1 & 0 & G_{2,5} & G_{2,6} & G_{2,7} \end{bmatrix}$$

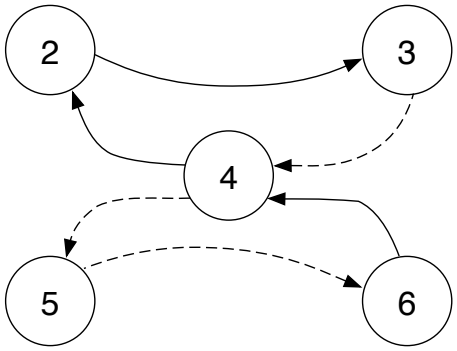


Figure 2.2: Closed loop configuration for case (3).

4. The following is the solution of Example 2.3.1, with

$$\mathbb{W} = \Big\{ f_{XU} \in \mathbb{P}_{\mathbb{XU}} : \sum_{u \in \mathbb{U}} f_{XU}(4, u) = \mathbf{0} \Big\}.$$

$$\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} = \{1,\ 2,\ 5,\ 7\}$$

$$F^* = \begin{bmatrix} 0 & 0 & 0 & \mathbf{0} & 0.19 & 0 & .32 \\ .16 & .33 & 0 & \mathbf{0} & 0 & 0 & 0 \end{bmatrix},$$

$$K^* = \begin{bmatrix} 0 & 0 & G_{1,3} & G_{1,4} & 1 & G_{1,7} & 1 \\ 1 & 1 & G_{2,3} & G_{2,4} & 0 & G_{2,7} & 0 \end{bmatrix}$$

Remark The maximal set of recurrent states contains two recurrent classes in this case. The transitions graph for the closed-loop transition graph is depicted in Fig. 2.3.

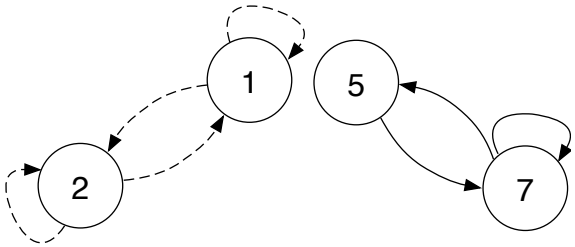


Figure 2.3: Closed loop configuration for case (4).

2.4 Minimum Number of Recurrent Classes

For a given control policy, the resulting Markov chain may contain multiple recurrent classes, and in this case the invariant `pmf` is not unique. In this section we justify the choice of maximizing the entropy of the joint invariant `pmf` rather than the entropy of the marginal `pmf` with respect to the state. Consider now the following optimization program:

$$f_{XU}^{*} = \arg \max_{f_{XU} \in \mathbb{W}} \mathcal{H}(\sum_{u \in \mathbb{U}} f_{XU}(\cdot, u)) \tag{2.8}$$

subject to:

$$\sum_{u^{+} \in \mathbb{U}} f_{XU}(x^{+}, u^{+}) = \sum_{x \in \mathbb{X}, u \in \mathbb{U}} \mathcal{Q}(x^{+}, x, u) f_{XU}(x, u), \quad x^{+} \in \mathbb{X}, \tag{2.9}$$

and note that its associated control policy, as suggested by Theorem 2.2.1, would also solve Problem 2.1.1 with the added benefit that the entropy of the marginal `pmf` is computationally cheaper. The advantage, however, of using the joint `pmf` is that it guarantees the minimum number of recurrent classes, as this section will show.

Let f_{XU} be an invariant `pmf` associated with control policy $\mathcal{K}$, and let $\eta_{f_{XU}}$ denote the number of distinct recurrent classes associated with f_{XU}. There must

exist invariant pmfs f^i_{XU}, $i = 1, ..., \eta_{f_{XU}}$, with the property that $\mathbb{S}^{\mathbb{X}}_{f^i_{XU}}$, $i = 1, ..., \eta_{f_{XU}}$, are pairwise disjoint, such that:

$$f_{XU} = \sum_{i=1}^{\eta_{f_{XU}}} \lambda_i f^i_{XU},$$

where $\lambda_i > 0$ and $\sum_{i=1}^{\eta_{f_{XU}}} \lambda_i = 1$.

Let the set of feasible pmfs with maximal support be given by:

$$\mathbb{IW}^{\max} \stackrel{def}{=} \Big\{ f_{XU} \in \mathbb{I} \cap \mathbb{W} : \mathbb{S}^{\mathbb{X}}_{\bar{f}_{XU}} \subseteq \ \mathbb{S}^{\mathbb{X}}_{f_{XU}}, \bar{f}_{XU} \in \mathbb{I} \cap \mathbb{W} \Big\} \tag{2.10}$$

Corollary 2.4.1. *Let f^*_{XU} be the solution to (2.3)-(2.4). The following holds:*

$$\eta_{f^*_{XU}} \leq \eta_{f_{XU}}, \quad f_{XU} \in \mathbb{IW}^{max}.$$

Proof. Suppose there exists a pmf $\hat{f}_{XU}$ in $\mathbb{IW}^{\max}$ such that $\eta_{\hat{f}_{XU}} < \eta_{f^*_{XU}}$. Since the set of recurrent classes associated with $\hat{f}_{XU}$ is smaller than the one associated with f^*_{XU}, there must exist a pair (x, u) in $\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} \times \mathbb{U}$ for which $\hat{f}_{XU}(x, u) > 0$ and $f^*_{XU}(x, u) = 0$, which contradicts Lemma 2.2.2. □

Example 2.4.2. *This simple example shows the difference between having an objective function as in (2.3) and (2.8). Let $\mathbb{X} = \{1, 2, 3, 4\}$, $\mathbb{U} = \{1, 2\}$ and $\mathbb{W} = \mathbb{P}_{\mathbb{XU}}$ (i.e., no further constraints are imposed), and consider the MDP specified by the following probability transition matrices:*

$$Q^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad Q^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where we use the notation $Q_{ij}^{u} \stackrel{def}{=} \mathcal{Q}(i,j,u)$. *The directed graph of possible transitions of the MDP is depicted in Fig. 2.4.*

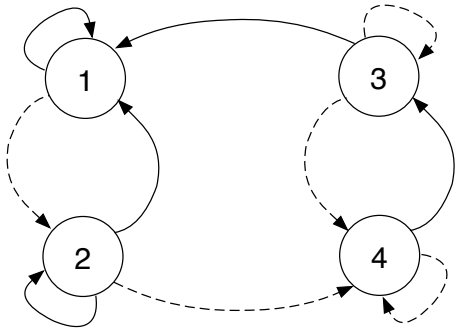


Figure 2.4: Posible transitions of the MDP in Example 2.4.2

Notation: In what follows, $F_{ij} \stackrel{def}{=} f_{XU}(i,j)$ and $K_{ij} \stackrel{def}{=} \mathcal{K}(i,j)$.

Consider the following invariant joint `pmf`:

$$F = \begin{bmatrix} .125 & 0.25 & 0 & .125 \\ .125 & 0 & .25 & .125 \end{bmatrix},$$

which can be achieved by the following control policy:

$$K = \begin{bmatrix} 0.5 & 1 & 0 & .5 \\ 0.5 & 0 & 1 & .5 \end{bmatrix},$$

The closed-loop configuration of possible transitions can be seen in Fig. 2.5. Note that $\mathbb{S}_{f_{XU}}^{\mathbb{X}} = \mathbb{X}$ and that the marginal `pmf` is uniform and has maximum entropy. Therefore the proposed `pmf` is a solution to (2.3)-(2.4), and, alongside the proposed control policy, solves Problem 2.1.1. However, the resulting Markov chain has two distinct recurrent classes.

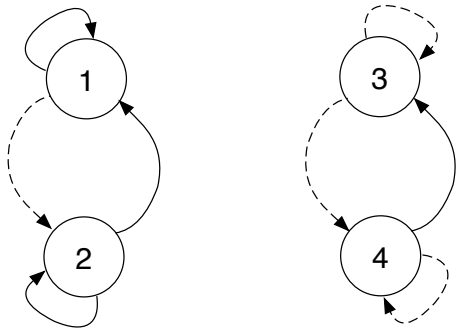


Figure 2.5: Closed-loop configuration for Example 2.4.2 with policy given in (2.12) and resulting invariant `pmf` (2.12).

Let us proceed by using the original convex program (2.3)-(2.4). We get:

$$F^* = \begin{bmatrix} 0.1268 & 0.1134 & 0.1015 & 0.1583 \\ 0.1583 & 0.1015 & 0.1134 & 0.1268 \end{bmatrix}, \tag{2.11}$$

$$K^* = \begin{bmatrix} 0.4447 & 0.5277 & 0.4723 & 0.5553 \\ 0.5553 & 0.4723 & 0.5277 & 0.4447 \end{bmatrix}, \tag{2.12}$$

which maintains the closed loop interconnection of Fig. 2.4 and results in a Markov chain with one recurrent class. Note that the marginal `pmf` does not have maximum entropy, but the entropy of the joint `pmf` is maximum.

2.5 Conclusion

This chapter addresses the design of full-state feedback memoryless policies for MDPs with finite state and action spaces. The main problem is to design policies that lead to the largest set of recurrent states, subject to convex constraints on the set of invariant `pmf`s. We described a finitely parametrized convex program that solves the problem via entropy maximization principles. Our approach has the

advantage of yielding a closed-loop Markov chain with least number of recurrent classes.

Chapter 3: Maximal Persistent Surveillance

In this chapter we apply the results we have obtained to the problem of maximal persistent surveillance. Some of the results from Chapter 2 are reproduced here for completeness, but proofs have been omitted. The results presented in this chapter have been published in [66].

The problem of maximal persistent surveillance consists of finding a memoryless control policy that results in the *largest* possible set of points in the lattice being persistently visited. The concept of persistent surveillance is similar to the concept of coverage [18], but differs in that the area to be surveilled must be revisited infinitely many times. In our setup, we also impose safety constraints that dictate certain *forbidden* regions. The forbidden regions may represent areas in which robots cannot operate (such as bodies of water) or are not allowed to visit (such as restricted airspace). The goal is to deploy the minimum number of robots equipped with a control policy that guarantees persistent surveillance of the largest possible set of lattice points without ever visiting a forbidden region.

Control design for persistent surveillance has been previously studied and many problem formulations and approaches have been proposed. In [19, 20], the authors present a semi-heuristic control policy that minimizes the time between visita-

tions to the same region. In [21], an algorithmic approach for persistent surveillance of a convex polygon in the plane is provided. In [22], communication constraints and sensor failures are incorporated in the problem of persistent surveillance and approximate dynamic programing is used. In [23], the authors describe a dynamic programming approach with temporal logic specifications, which can be used to cast persistent surveillance problems. Speed control for robots performing persistent monitoring on a predetermined path is addressed in [24], where linear programing techniques are employed. In our work, we focus on deploying the minimum number of robots, equipped with memoryless policies (found via convex optimization), tasked with persistent surveillance of the largest possible set of locations without violating safety constraints.

We model each robot as a fully-observed MDP with finite state and control spaces. This approach, which has been successfully used in the context of navigation and path planning ([67–69]), allows for the development of robust and highly scalable algorithms. Without loss of generality, we consider robots whose state is taken as its position on a finite two-dimensional lattice and direction of motion (taken from a set of four possible orientations), and limit the control space to two control actions ("forward" and "turn right"). The limitation in the control space illustrates how constrained actuation can be incorporated in our formulation, however the ideas described in this paper can be extended to more general dynamics and state/control spaces.

The remainder of this chapter is organized as follows. Section 3.1 provides notation, basic definitions and the problem statement. The convex program that

computes the maximal set of persistently surveilled states and its associated control policy is presented in Section 3.2. Section 3.3 provides details on computing the smallest deployment of robots necessary for maximal persistent surveillance. We discuss limiting behavior and use of additional constraints in Section 3.4. In Section 3.5, we discuss approaches to the problem of maximal persistent surveillance when a fixed number of robots are available. Numerical examples are given throughout the paper to illustrate concepts and the proposed methodology. Conclusions can be found in Section 3.6.

3.1 Preliminaries and Problem Statements

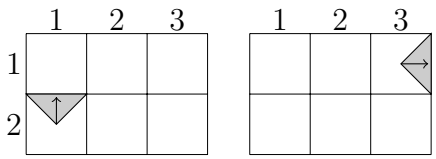
The following notation is used throughout the paper:

$\mathbb{Z} \times \mathbb{Y}$	set of lattice positions
$\mathbb{O}$	set of orientations
$\mathbb{X} \stackrel{def}{=} \mathbb{Z} \times \mathbb{Y} \times \mathbb{O}$	set of robot states
$\mathbb{F} \subset \mathbb{X}$	set of forbidden states
$\mathbb{U}$	set of control actions

The state of the robot will be graphically represented as shown in Fig. 3.1. The robot's dynamics are governed by the (conditional) probability of X_{k+1} given the current state X_k and control action U_k, and are denoted as:

$$\mathcal{Q}(x^+, x, u) \stackrel{def}{=} P(X_{k+1} = x^+ \mid X_k = x, U_k = u),$$

where $x, x^+ \in \mathbb{X}, u \in \mathbb{U}$.



$(1, 2, U)$ $(3, 1, R)$

Figure 3.1: Graphical representation of the state of the robot. In this examples, we use $\mathbb{Z} = \{1, 2, 3\}$, $\mathbb{Y} = \{1, 2\}$, and $\mathbb{O} = \{R, U, L, D\}$, where R, U, L and D represent right, up, left and down directions, respectively.

We denote any memoryless control policy by

$$\mathcal{K}(u, x) \stackrel{def}{=} P(U_k = u \mid X_k = x), \qquad u \in \mathbb{U}, x \in \mathbb{X},$$

where $\sum_{u \in \mathbb{U}} \mathcal{K}(u, x) = 1$ for all x in $\mathbb{X}$. The set of all such policies is denoted as $\mathbb{K}$. Note that the computation of a control action may be deterministic (when $\mathcal{K}(u, x) = 1$ for a given action u) or carried out in a randomized manner, in which case the policy dictates the probabilities assigned to each control action for a given state.

Assumption When multiple robots are considered, we assume that they are identical and have dynamics governed by $\mathcal{Q}$. In these situations, every robot executes the same control policy. Moreover, multiple robots are allowed to occupy the same state.

Given a control policy $\mathcal{K}$, the conditional state transition probability of the closed loop is represented as:

$$\mathcal{P}_{\mathcal{K}}(X_{k+1} = x^+ \big| X_k = x) \stackrel{def}{=} \sum_{u \in \mathbb{U}} \mathcal{Q}(x^+, x, u)\mathcal{K}(u, x).$$

We will refer to this quantity as $\mathcal{Q}_{\mathcal{K}}(x^+, x) \stackrel{def}{=} \mathcal{P}_{\mathcal{K}}(X_{k+1} = x^+ \big| X_k = x)$.

3.1.1 Recurrence and Persistent Surveillance

A state $x \in \mathbb{X}$ is *recurrent* under a control policy $\mathcal{K}$ if the probability of a robot revisiting state x is one, that is:

$$P_{\mathcal{K}}(X_k = x \text{ for some } k > 0 \;\big|\; X_0 = x) = 1. \tag{3.1}$$

We define the *set of recurrent states* $\mathbb{X}_{\mathcal{K}}^R$ under control policy $\mathcal{K}$ as follows:

$$\mathbb{X}_{\mathcal{K}}^R \stackrel{def}{=} \Big\{ x \in \mathbb{X} : (3.1) \; holds \Big\}.$$

Remark Membership in $\mathbb{X}_{\mathcal{K}}^R$ guarantees that once a state is visited, it will be revisited infinitely many times under control policy $\mathcal{K}$. It does not, however, guarantee that each state in $\mathbb{X}_{\mathcal{K}}^R$ will be visited for all initial states in $\mathbb{X}_{\mathcal{K}}^R$ because $\mathbb{X}_{\mathcal{K}}^R$ may contain multiple recurrent classes. In fact, a robot will visit a certain recurrent state x with probability one if and only if it is initialized in the same recurrent class. Moreover, note that once a robot enters a recurrent class, it will never exit under control policy $\mathcal{K}$.

We say a state x is *persistently surveilled* under control policy $\mathcal{K}$ and initial state $x_0 \in \mathbb{X}$ if it is recurrent under $\mathcal{K}$ and

$$P_{\mathcal{K}}(X_k = x \text{ for some } k > 0 \;\big|\; X_0 = x_0) = 1. \tag{3.2}$$

If a state x is persistently surveilled under control policy $\mathcal{K}$ and initial state $x_0 \in \mathbb{X}_{\mathcal{K}}^R$, then it must be that x and x_0 belong to the same recurrent class.

We define the *set of persistently surveilled states* $\mathbb{X}^{ps}_{x_0,\mathcal{K}}$ under control policy $\mathcal{K}$ and initial state $x_0 \in \mathbb{X}$ to be:

$$\mathbb{X}^{ps}_{x_0,\mathcal{K}} \stackrel{def}{=} \Big\{ x \in \mathbb{X}^R_{\mathcal{K}} : (3.2)\ holds \Big\}.$$

The set $\mathbb{X}^{ps}_{x_0,\mathcal{K}}$ is a recurrent class of the closed loop dynamics $\mathcal{Q}_{\mathcal{K}}$. Note that for every state x in $\mathbb{X}^{ps}_{x_0,\mathcal{K}}$, it holds that $\mathbb{X}^{ps}_{x,\mathcal{K}} = \mathbb{X}^{ps}_{x_0,\mathcal{K}}$. Moreover, if there exists a recurrent state for which $\mathbb{X}^{ps}_{x_0,\mathcal{K}} = \mathbb{X}^R_{\mathcal{K}}$, the set $\mathbb{X}^R_{\mathcal{K}}$ has *only one* recurrent class.

Given a set $\mathbb{F}$ of forbidden states, we define the set of states that are recurrent and for which the probability of transitioning into $\mathbb{F}$ is zero.

The *set of* $\mathbb{F}$*-safe recurrent states* $\mathbb{X}^R_{\mathcal{K},\mathbb{F}}$ under a control policy $\mathcal{K}$ is defined as:

$$\mathbb{X}^R_{\mathcal{K},\mathbb{F}} \stackrel{def}{=} \Big\{ x \in \mathbb{X}^R_{\mathcal{K}} :\ \mathcal{Q}_{\mathcal{K}}(x^+, x) = 0,\ x^+ \in \mathbb{F} \Big\}.$$

We define the **maximal** *set of* $\mathbb{F}$*-safe recurrent states* as:

$$\mathbb{X}^R_{\mathbb{F}} \stackrel{def}{=} \bigcup_{\mathcal{K} \in \mathbb{K}} \mathbb{X}^R_{\mathcal{K},\mathbb{F}}.$$

Finally, the *set of* $\mathbb{F}$*-safe persistently surveilled states* $\mathbb{X}^{ps}_{s_0,\mathcal{K},\mathbb{F}}$ under a control policy $\mathcal{K}$ and initial state $s_0 \in \mathbb{S}$ is defined as:

$$\mathbb{X}^{ps}_{s_0,\mathcal{K},\mathbb{F}} \stackrel{def}{=} \Big\{ x \in \mathbb{X}^{ps}_{s_0,\mathcal{K}} :\ \mathcal{Q}_{\mathcal{K}}(x^+, x) = 0,\ x^+ \in \mathbb{F} \Big\}.$$

Remark As before, $\mathbb{X}^{ps}_{x_0,\mathcal{K},\mathbb{F}}$ is a (safe) recurrent class of $\mathcal{Q}_{\mathcal{K}}$.

3.1.2 Problem Statement

We start by addressing the following problem:

Problem 3.1.1. *(Maximal set of $\mathbb{F}$-safe recurrent states). Given a set of forbidden states $\mathbb{F}$, determine:*

(a) $\mathbb{X}_{\mathbb{F}}^{R}$; and

(b) a control policy $\mathcal{K}^*$ such that $\mathbb{X}_{\mathcal{K}^*}^{R} = \mathbb{X}_{\mathbb{F}}^{R}$.

In light of Remark 3.1.1, note that in order to persistently surveil all possible states, we need to determine how many robots to use and in which state they should be initialized. The following problem addresses this issue.

Problem 3.1.2. *(Maximal $\mathbb{F}$-safe persistent surveillance). Given a set of forbidden states $\mathbb{F}$, determine the* minimum *number of robots r, a control policy $\hat{\mathcal{K}}$ and a set of initial states $\{x^1, ..., x^r\}$, so that*

$$\bigcup_{i=1}^{r} \mathbb{X}_{x^i,\hat{\mathcal{K}},\mathbb{F}}^{ps} = \mathbb{X}_{\mathbb{F}}^{R}. \tag{3.3}$$

Remark The following is a list of important comments on Problems 3.1.1 and 3.1.2.

- There is no $\mathcal{K}$ such that the states in $\mathbb{X} \setminus \mathbb{X}_{\mathbb{F}}^{R}$ can be $\mathbb{F}$-safe and recurrent.
- Once r robots are initialized with initial states $\{x^1, ..., x^r\}$, it is guaranteed that the largest possible set of states will be visited infinitely many times without ever visiting a forbidden state.

We will use the results from Chapter 2 to solve Problems 3.1.1 and 3.1.2.

3.2 Computing the Maximal Set of Recurrent States

Recall that $\mathbb{P}_{\mathbb{XU}}$ is the set of all `pmf`s with support in $\mathbb{X} \times \mathbb{U}$, and consider the following convex optimization program:

$$f_{XU}^* = \arg \max_{f_{XU} \in \mathbb{P}_{\mathbb{XU}}} \mathcal{H}(f_{XU}) \tag{3.4}$$

subject to:

$$\sum_{u^+ \in \mathbb{U}} f_{XU}(x^+, u^+) = \sum_{x \in \mathbb{X}, u \in \mathbb{U}} \mathcal{Q}(x^+, x, u) f_{XU}(x, u) \tag{3.5}$$

$$\sum_{u \in \mathbb{U}} f_{XU}(x, u) = 0, \quad x \in \mathbb{F} \tag{3.6}$$

where $\mathcal{H} : \mathbb{P}_{\mathbb{XU}} \to \Re_{\geq 0}$ is the entropy of f_{XU}, and is given by

$$\mathcal{H}(f_{XU}) = -\sum_{u \in \mathbb{U}} \sum_{x \in \mathbb{X}} f_{XU}(x, u) \ln \big(f_{XU}(x, u) \big)$$

where we adopt the standard convention that $0 \ln(0) = 0$.

The following proposition, which has been modified from Theorem 2.2.1, provides a solution to Problem 3.1.1.

Proposition 3.2.1. *Let $\mathbb{F}$ be given, assume that (3.4)-(3.6) is feasible, and that f_{XU}^* is the optimal solution. In addition, adopt the marginal* `pmf` *$f_X^*(x) = \sum_{u \in \mathbb{U}} f_{XU}^*(x, u)$ and let $\mathcal{G} : \mathbb{U} \times \mathbb{X} \to [0, 1]$ be any function satisfying $\sum_{u \in \mathbb{U}} \mathcal{G}(x, s) = 1$ for all x in $\mathbb{X}$. The following holds:*

(a) $\mathbb{X}_{\mathbb{F}}^R = \mathbb{S}_{f_{XU}^*}^{\mathbb{X}}$

(b) $\mathbb{X}^R_{\mathcal{K}^*} \setminus \mathbb{F} = \mathbb{X}^R_{\mathbb{F}}$ *for* $\mathcal{K}^*$ *given by:*

$$\mathcal{K}^*(u,x) = \begin{cases} \frac{f^*_{XU}(x,u)}{f^*_X(x)}, & x \in \mathbb{S}^{\mathbb{X}}_{f^*_{XU}} \\ \mathcal{G}(u,x), & \textit{otherwise} \end{cases}, \quad (u,x) \in \mathbb{U} \times \mathbb{X} \tag{3.7}$$

where $\mathbb{S}^{\mathbb{X}}_{f^*_{XU}}$ *is given by* $\mathbb{S}^{\mathbb{X}}_{f^*_{XU}} = \{x \in \mathbb{X} : \sum_{u \in \mathbb{U}} f^*_{XU}(x,u) > 0\}$.

The proof of Proposition 3.2.1 closely follows the proof of Theorem 2.2.1 and is omitted. However, it is important to highlight that constraint (3.6) enforces $\mathbb{F}$-safety.

Example 3.2.2. *Let* $\mathbb{Z} = \mathbb{Y} = \{1, ..., 5\}$, $\mathbb{O} = \{R, U, L, D\}$ *and consider a robot whose action space is given by* $\mathbb{U} = \{\text{“}forward\text{”}, \text{“}turn\ right\text{”}\}$. *Moreover, let the set of forbidden states be given by:*

$$\mathbb{F} = \big\{(z, y, \theta) \in \mathbb{X} : (z, y) \in \{(1,1), (1,5), (5,1), (5,5), (3,3)\}\big\},$$

which means the robot is prohibited from visiting the center and corner locations of the lattice.

In order to specify $\mathcal{Q}$, *we first define an auxiliary conditional pmf* $\mathcal{Q}'$ *defined on* $\mathbb{Z}' = \mathbb{Y}' = \{1, 2, 3\}$ *and* $\mathbb{O}' = \{R, U, L, D\}$. *For clarity,* $\mathcal{Q}'$ *is shown graphically in Fig. 3.2, which contains the probabilities of transitioning to states shown as dark triangles given the previous state shown as a white triangle. There is uncertainty only for transitions that occur on the edge of the lattice. Since we consider dynamics that are spatially invariant, the transition probabilities for states not shown in Fig. 3.2 can be computed by appropriate manipulation of the ones shown. Similarly,* $\mathcal{Q}$ *is constructed by appropriate expansion of* $\mathcal{Q}'$.

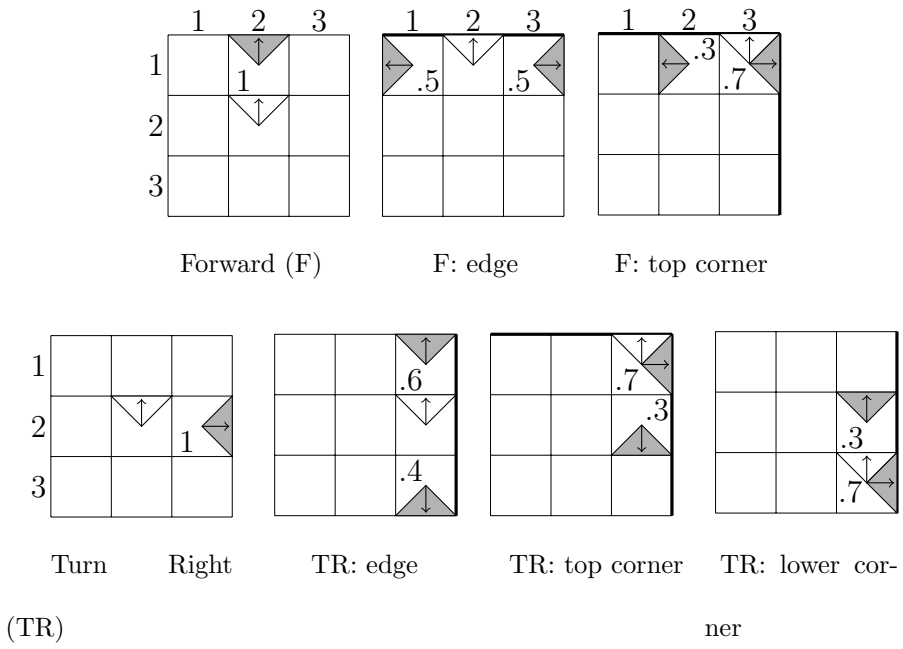


Figure 3.2: Graphical representation of some transitions in $\mathcal{Q}'$.

We use [50] to solve (3.4)-(3.6) and use Proposition 3.2.1 to compute $\mathbb{X}_{\mathbb{F}}^{R}$ and a control policy $\mathcal{K}^*$ such that $\mathbb{X}_{\mathcal{K}^*}^{R} = \mathbb{X}_{\mathbb{F}}^{R}$. The set $\mathbb{X}_{\mathbb{F}}^{R}$ can be seen in Fig. 3.3, where the areas in red represent the states in $\mathbb{F}$, and the triangles in blue represent the states in $\mathbb{X}_{\mathbb{F}}^{R}$. The control policy $\mathcal{K}^*$, computed using (3.7), has been omitted due to space constraints.

3.3 Maximal Persistent Surveillance and Robot Deployment

In this section, we provide a solution to Problem 3.1.2, which seeks the minimum number r of robots, a control policy $\hat{\mathcal{K}}$ and a set of initial states $\{x^1, ..., x^r\}$ so that $\bigcup_{i=1}^{r} \mathbb{X}_{x^i,\hat{\mathcal{K}},\mathbb{F}}^{ps} = \mathbb{X}_{\mathbb{F}}^{R}$.

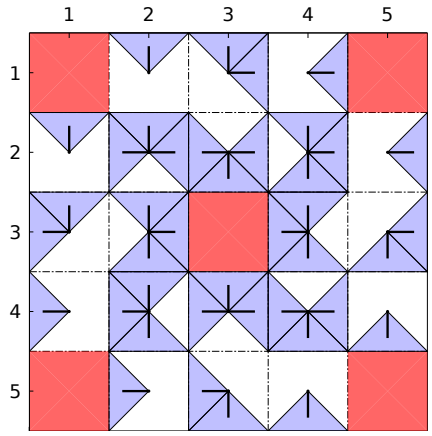


Figure 3.3: Depiction of $\mathbb{X}_{\mathbb{F}}^{R}$ in blue. The red areas represent the forbidden states.

In light of a previous remark, recall that the set of $\mathbb{F}$-safe persistently surveilled states $\mathbb{X}_{x_0,\mathcal{K},\mathbb{F}}^{ps}$ is a *recurrent class* of $\mathcal{Q}_{\mathcal{K}}$. In practice, this means that when a robot with initial state $X_0 = x_0$ applies control policy $\mathcal{K}$, it is guaranteed that:

- the robot will never leave $\mathbb{X}_{x_0,\mathcal{K},\mathbb{F}}^{ps}$;
- every state in $\mathbb{X}_{x_0,\mathcal{K},\mathbb{F}}^{ps}$ will be visited infinitely many times;
- states in $\mathbb{F}$ will never be visited.

To find all the (safe) recurrent classes in $\mathbb{X}_{\mathcal{K},\mathbb{F}}^{R}$, flood-fill-type algorithms may be used, where the graph of $\mathcal{Q}_{\mathcal{K}}$ is traversed, either in a depth-first or breath-first manner [51]. An edge from x to x^+ of the graph of $\mathcal{Q}_{\mathcal{K}}$ exists if and only if $\mathcal{Q}_{\mathcal{K}}(x^+, x) > 0$ holds.

Given $\mathbb{F}$ and a control policy $\mathcal{K}$, let $n_{\mathcal{K}}$ be the number of distinct recurrent classes of $\mathcal{Q}_{\mathcal{K}}$, and note that the following holds:

$$\bigcup_{i=1}^{n_{\mathcal{K}}} \mathbb{X}_{x^i,\mathcal{K},\mathbb{F}}^{ps} = \mathbb{X}_{\mathcal{K},\mathbb{F}}^{R},$$

where $\{x^1, ..., x^{n_\mathcal{K}}\}$ is a set of initial states, and $\left\{\mathbb{X}^{ps}_{x^i,\mathcal{K},\mathbb{F}}\right\}_{i=1}^{n_\mathcal{K}}$ are distinct recurrent classes.

We define the set of all admissible control policies whose $\mathbb{F}$-safe set of recurrent states are maximal to be:

$$\mathbb{K}^R_\mathbb{F} = \left\{\mathcal{K} \in \mathbb{K} : \mathbb{X}^R_{\mathcal{K},\mathbb{F}} = \mathbb{X}^R_\mathbb{F}\right\},$$

The following proposition shows that $n_{\mathcal{K}^*} \leq n_\mathcal{K}$ for all $\mathcal{K}$ in $\mathbb{K}^R_\mathbb{F}$.

Proposition 3.3.1. *Let $\mathbb{F}$ be given, and take $\mathcal{K}^*$ to be the control policy in (3.7). The following holds:*

$$n_{\mathcal{K}^*} \leq n_\mathcal{K}, \qquad \mathcal{K} \in \mathbb{K}^R_\mathbb{F}.$$

The proof of Proposition 3.3.1 is similar to that of Corollary 2.4.1 and is omitted.

Remark By exploring the graph of $\mathcal{Q}_{\mathcal{K}^*}$ found in Example 3.2.2, we conclude that only one robot is required to perform maximal persistent surveillance (i.e., $\mathbb{X}^R_\mathbb{F}$ contains only one recurrent class). Any state in $\mathbb{X}^R_\mathbb{F}$ may be selected as the robot's initial state.

Example 3.3.2. *Consider again the example described in Example 3.2.2, and suppose that we now change the set of forbidden states to include location* $(4, 3)$ *(i.e., let* $\mathbb{F} = \left\{(z, y, \theta) \in \mathbb{X} : (z, y) \in \{(1, 1), (1, 5), (5, 1), (5, 5), (3, 3), (4, 3)\}\right\}$*).*

Re-solving (3.4)-(3.6), applying Propositions 3.2.1 and 3.3.1, and searching the graph of the closed loop Markov chain, we conclude that at least three robots

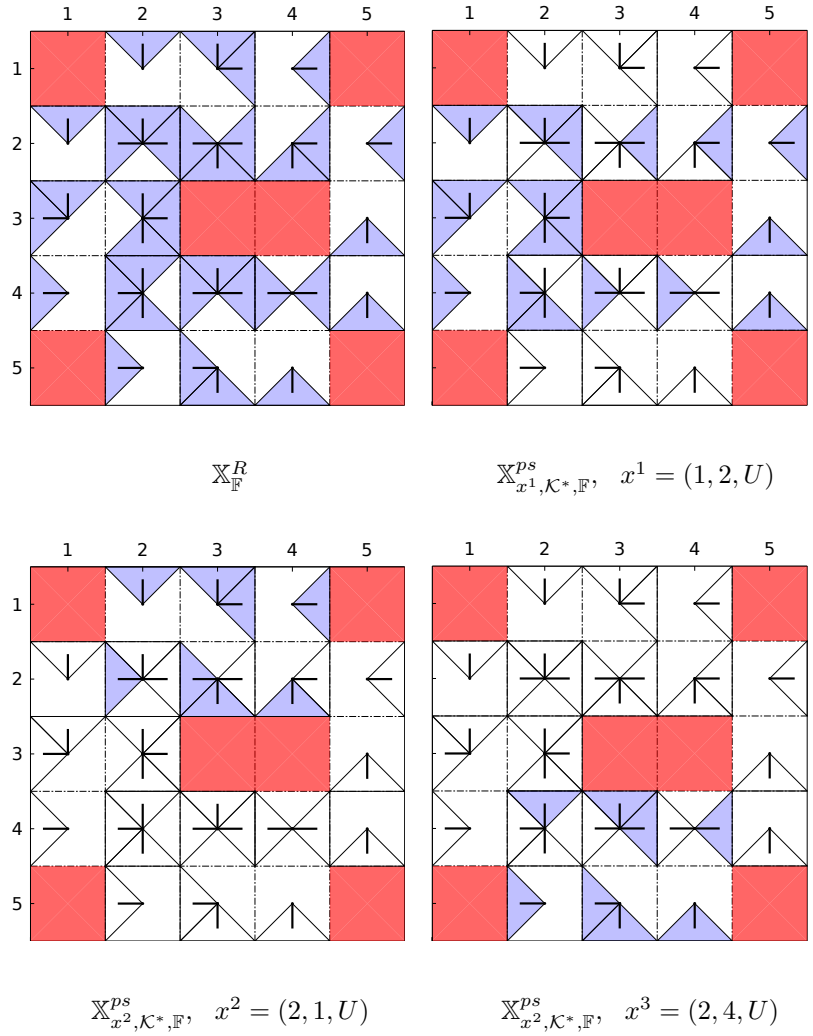


Figure 3.4: Top left: maximal set of recurrent states $\mathbb{X}_{\mathbb{F}}^{R}$ (in blue). Others: three recurrent classes whose union is $\mathbb{X}_{\mathbb{F}}^{R}$.

are required to perform maximal persistent surveillance of $\mathbb{X}_{\mathbb{F}}^{R}$ (see Fig. 3.4). Any state from each recurrent class may be used as initial states, so we can chose the set of initial states to be: $\big\{(1,2,U),(2,1,U),(2,4,U)\big\}$. Note that the set $\mathbb{X}_{\mathbb{F}}^{R}$ is now smaller (34 vs. 40 states).

3.4 Limiting Behavior and Other Constraints

We define $\mathcal{T}_{\mathcal{K}}$, the long term proportion of time the robot, under control policy $\mathcal{K}$, visits state x in $\mathbb{X}$ having started at state x_0, to be:

$$\mathcal{T}_{\mathcal{K}}(x, x_0) \stackrel{def}{=} \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} \mathcal{I}\big(X_i = x, X_0 = x_0\big),$$

where $\mathcal{I}$ is the indicator function.

3.4.1 Limiting Behavior with One Recurrent Class

Given a forbidden set $\mathbb{F}$, and let f^*_{XU} be the optimal solution to (3.4)-(3.6), and $\mathcal{K}^*$ be the control policy computed in (2.5), and suppose $\mathbb{X}^R_{\mathcal{K}^*,\mathbb{F}}$ has only one recurrent class. For any initial state x_0 in $\mathbb{X}^R_{\mathbb{F}}$, the following holds with probability one:

$$\mathcal{T}_{\mathcal{K}^*}(x, x_0) = f^*_X(x), \tag{3.8}$$

were $f^*_X(x) = \sum_{u\in\mathbb{U}} f^*_{XU}(x, u)$. Since we have not imposed aperiodicity on $\mathcal{Q}_{\mathcal{K}^*}$, we cannot state stronger convergence. However, equation (3.8) still tells us valuable information regarding the limiting behavior of the robot.

Note that the `pmf` that maximizes the entropy is "as uniform as possible." However, additional convex constraints can be added to our formulation in order to shape the distribution of the optimal `pmf` and, thus, influence the limiting behavior of the robot.

Consider the following constraint:

$$\sum_{(z,y)\in\mathbb{D},\ \theta\in\mathbb{O},\ u\in\mathbb{U}} f_{XU}((z,y,\theta),u) > \alpha, \tag{3.9}$$

where $\mathbb{D} \subset \mathbb{Z} \times \mathbb{Y}$ is a region of the lattice. The set $\mathbb{D}$ can be interpreted as a region of high interest that should be surveilled more often. Suppose the convex program (3.4)-(3.6) and (3.9) is feasible, that f^{**}_{XU} is the optimal solution and $\mathcal{K}^{**}$ is the associated control policy. The following holds for any x_0 in $\mathbb{X}^R_{\mathbb{F}}$ with probability one:

$$\sum_{(z,y)\in\mathbb{D},\ \theta\in\mathbb{O}} \mathcal{T}_{\mathcal{K}^{**}}\big((z,y,\theta),s_0\big) > \alpha.$$

Example 3.4.1. *Let $\mathbb{Z} = \mathbb{Y} = \{1,...,10\}$, $\mathbb{O} = \{R,U,L,D\}$, and consider again a robot whose action space is given by $\mathbb{U} = \{$"Forward", "Turn Right"$\}$. The dynamics $\mathcal{Q}$ are similar to what was used in Examples 3.2.2 and 3.3.2, except that we add uncertainty to the transition of states that lie in the interior of the grid (see Fig. 3.5). The probabilities for states on the edge of the grid are the same as before (see Fig.3.2).*

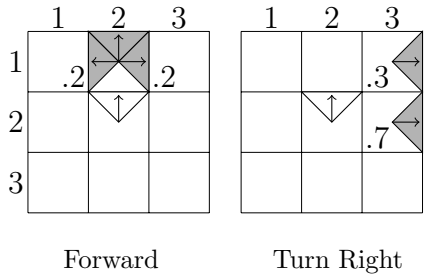


Figure 3.5: Graphical representation of some transitions in $\mathcal{Q}'$.

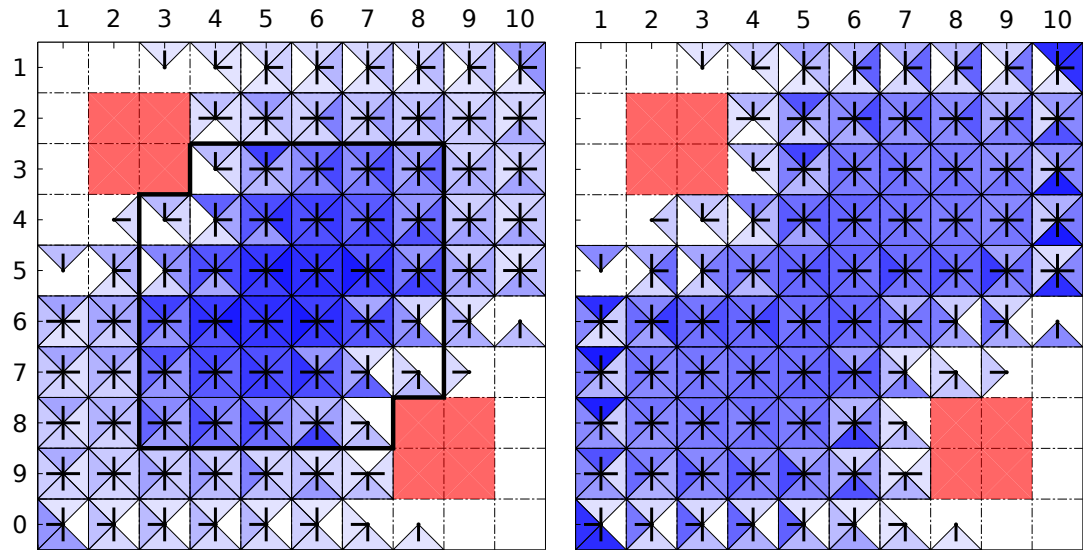


Figure 3.6: Left: Solution to Example 3.4.1; Right: Solution to Example 3.4.1 disregarding the extra constraint.

The set of forbidden states is given by:

$$\mathbb{F} = \{(z, y, \theta) \in \mathbb{X} : (z, y) \in \{(2,2),(2,3),(3,2),(3,3),(8,8),(8,9),(9,8),(9,9)\}\},$$

and consider $\mathbb{D} = \{(z, y) \in \mathbb{Z} \times \mathbb{Y} : 3 \leq z, y \leq 8\}$, *and let* $\alpha = 0.75$.

We solve (3.4)-(3.6) using [50]. In Fig. 3.6, each state that belongs in $\mathbb{S}_{\mathbb{F}}^{R}$ is shown in blue, where the darker the blue, the higher the value of f_X^*. The image on the left in Fig. 3.6 shows the solution to Example 3.4.1. The image on the right shows the solution disregarding the extra constraint. Note that the distribution is relatively uniform.

3.4.2 Limiting Behavior with Multiple Recurrent Classes

Consider again f^*_{XU} and $\mathcal{K}^*$ as before, and, without loss of generality, let $\mathbb{X}^R_{\mathcal{K}^*,\mathbb{F}}$ have two recurrent classes with initial states x^1 and x^2 $\big($i.e., $\mathbb{X}^{ps}_{x^1,\mathcal{K}^*,\mathbb{F}} \cup \mathbb{X}^{ps}_{x^2,\mathcal{K}^*,\mathbb{F}} = \mathbb{X}^R_{\mathcal{K}^*,\mathbb{F}}\big)$. For any initial state x_0 in $\mathbb{X}^{ps}_{x^1,\mathcal{K}^*,\mathbb{F}}$ $\big($equiv., $\mathbb{X}^{ps}_{x^2,\mathcal{K}^*,\mathbb{F}}\big)$, the following holds with probability one:

$$\mathcal{T}_{\mathcal{K}^*}(x, x_0) = \frac{f^*_X(x)}{\beta}, \tag{3.10}$$

where $\beta = \sum_{x \in \mathbb{X}^{ps}_{x^1,\mathcal{K}^*,\mathbb{F}}} f^*_X(x)$ $\big($equiv. $\beta = \sum_{x \in \mathbb{X}^{ps}_{x^2,\mathcal{K}^*,\mathbb{F}}} f^*_X(x)\big)$.

With equation (3.10) in mind, note that additional convex constraints may also be used to influence the limiting behavior of the robots. Moreover, by carefully selecting the number of robots allocated to each recurrent class, one can achieve a desirable limiting behavior for the ensemble of robots.

3.5 Deployment of a Fixed Number of Robots

The problem we proposed and solved in this chapter is that of finding the minimum number of robots capable of achieving maximal persistent surveillance. A natural extension is to tackle the situation when a fixed number of robots (which may be fewer than the minimum number necessary for maximal persistent surveillance) is available to undertake the task. The following formalizes the problem:

Problem 3.5.1. *Given a set of forbidden states $\mathbb{F}$ and a number of available robots*

$\bar{r}$, determine a control policy $\hat{\mathcal{K}}$ and a set of initial states $\{x^1, ..., x^{\bar{r}}\}$, so that

$$\bigcup_{i=1}^{\bar{r}} \mathbb{X}^{ps}_{x^i,\hat{\mathcal{K}},\mathbb{F}} \text{ is largest.} \tag{3.11}$$

Recall that by solving the original optimization program (3.4)- (3.6), we obtain a control policy $\mathcal{K}^*$ and a set of recurrent states with $n_{\mathcal{K}^*}$ recurrent classes. The following proposition solves Problem 3.5.1:

Proposition 3.5.2. *Without loss of generality, let $\{\bar{x}^1, ..., \bar{x}^{n_{\mathcal{K}^*}}\}$ be members of each recurrent class of the closed loop Markov chain $\mathcal{Q}_{\mathcal{K}^*}$, such that:*

$$\left|\mathbb{X}^{ps}_{\bar{x}^1,\mathcal{K}^*,\mathbb{F}}\right| \geq \cdots \geq \left|\mathbb{X}^{ps}_{\bar{x}^{n_{\mathcal{K}^*}},\mathcal{K}^*,\mathbb{F}}\right| \tag{3.12}$$

Two situations arise:

i) $\bar{r} \geq n_{\mathcal{K}^*}$

In this case the number of available robots is sufficient for maximal persistent surveillance, with control policy $\mathcal{K}^*$. The additional robots can be distributed at random amongst the recurrent classes.

ii) $\bar{r} < n_{\mathcal{K}^*}$

When the number of available robots is not sufficient for maximal persistent surveillance, choose control policy $\mathcal{K}^*$ and initial states $\{\bar{x}^1, ..., x^{\bar{r}}\}$.

Proof. Situation *i*) follows from the main contents of this chapter. It remains to show that the strategy for situation *ii*) is optimal. This can be seen by noting that by solving (3.4)- (3.6) we maximize the entropy of the joint `pmf` f_{XU}. By Lemma 2.2.2, the `pmf` that maximizes the entropy, not only has the largest support, but

also the *maximal* support. This means that any state-action pair that can be made recurrent *is* recurrent in the closed loop Markov chain $\mathcal{Q}_{\mathcal{K}^*}$. Therefore the best that can be done with fewer robots than $n_{\mathcal{K}^*}$ is to deploy them to the largest recurrent classes. □

3.6 Conclusion

In this chapter we have applied the results obtained in Chapter 2 to the problem of maximal persistent surveillance for robots whose dynamics are governed by MDPs. We dealt with safety constraints by casting a convex constraint on the invariant `pmf`, which allowed the application of our method. The simple structure of the resulting controllers makes them implementable in small robots.

Chapter 4: Sensor Usage in Denied Environments

In this chapter, we study the problem of sensor usage in denied environments, inside which observations of the state of the system are only available by paying a price in the cost to be minimized. One can imagine an agent whose goal is to stay put at a desired location inside the denied environment, but disturbances cause it to wander away. The agent can only return to the desired location once it observes its current location. In this scenario, there are two ways the agent accesses its location: by drifting outside the denied environment, where observations are readily available; or by paying a price, which can be thought of as a communication cost, to receive an observation of its location.

It should be clear that the agent may have an incentive to *not* pay to access its location with the expectation that it will sooner or later wander outside the denied environment, at which point it will receive free access to its location and be able to return to the desired location. On the other hand, recall that the goal of the agent is to stay put at the desired location and, by waiting until the eventuality of existing the denied environment, the cost paid by deviating too much from the desired location may outweigh the cost of using the sensor.

In the literature, efforts in event-based control and estimation aim at estab-

lishing protocols for which actuating, computing and sensing are undertaken only when needed, rather than periodically [39–42]. The work in [43], for example, proposes a sporadic control scheme for first order linear systems. In our work, when the agent exists the denied environment and receives a state observation can be though as an event-based sensing. The difference is that the event under consideration is a problem constraint rather than a design parameter, which is generally the case in the event-based control and estimation literature.

A deceptively similar framework is that of control and estimation over costly or limited communication channels. These setups occur in the context of networked systems, where different components of systems are not collocated. The authors of [44] consider an estimation problem where only a limited number of observations are available over a finite horizon. The work in [45] proposes an estimation framework in which a pre-processor decides whether or not to transmit measurements to the estimator. The most similar problem setup to ours is the one in [46, 47], where the authors consider an estimation problem in which the system can be reset by paying a cost, and where the reset policy is a function of the current state of the system.

There are two main features of our work that distinguishes it from the ones mentioned in the previous paragraph: *i*) in our work, the decision of whether or not to use the sensor cannot be a function of the state, which is in general the case in networked estimation problems; and *ii*) in our setup, we consider "event-based", cost-free sensor usage, which introduces an incentive to not request a costly observation. To our knowledge, a framework that combines these two characteristics has not been studied before.

Our main goal is to understand the trade-off between waiting for a free sensor use in the future and paying to immediate sensor access. In our problem formulation, we strip this problem down to the fundamentals: we consider a Markov process that evolves in time inside of the denied environment. At each time step, the only control variable is the decision whether or not to reset the process back to its initial state by paying a price (this represents the task of using the sensor *and* returning to the desired location).

We show that the analysis of this problem simplifies by recognizing that the stochastic process is in fact a renewal process, and our design parameter is the support of the stopping time associated with the renewal process. Our approach enables us to establish conditions for which any local minimum, if it exists, is also global, thus facilitating the search for the minimizer. We provide various numerical examples to illustrate these results.

Although the treatment and examples we provide focus on $\mathbb{R}^n$, some of the results in this chapter can be extended to stochastic processes in different spaces, since the cost we consider rely solely on expectations and the nature of the underlying process does not play a direct role in the cost. In Chapter 5, for example, we extend the results of this chapter to Markov chains.

This chapter is organized as follows: in Section 4.1, we formalize the problem formulation. Section 4.2 recasts the problem using as a renewal process. In Section 4.3, we establish a recursive expression for the cost we wish to minimize. Section 4.4 describes conditions on the problem that facilities the search for the optimal design parameter. In Section 4.5, we introduce the concept of radial stochastic order, which

plays an important role in determining if the conditions established in Section 4.4 are satisfied. Sections 4.6 and 4.7 provide numerical examples. Finally, we conclude in Section 4.8.

4.1 Initial Considerations

Consider a discrete-time Markov process X_k that takes values in $\mathbb{R}^n$ and a subset $\mathbb{D}$ of $\mathbb{R}^n$. The process is initialized according to a known distribution that is symmetric around 0, which we take to be, without loss of generality, the desired state. At each time step and for as long as the process remains contained in $\mathbb{D}$, a decision can be made to reset the process back to its initial distribution. This is called a *controlled* or an *active reset*. Every time the process undergoes such reset, a penalty of ρ is accrued in the cost we wish to minimize, which penalizes the deviation from 0 in the averaged infinite horizon sense. Whenever the process exits the set $\mathbb{D}$, however, it undergoes a *passive reset* where it is reinitialized to the initial distribution without penalty.

Note that, without paying the price, direct observation of the process is not available in the set $\mathbb{D}$ and the decision of resetting the process can only depend on the binary observation (and its history) of whether the process has exited the set $\mathbb{D}$. In other words, a controlled reset depends only on the number of time steps that have elapsed since the previous reset.

It should be clear that if no penalty is imposed per controlled reset, the optimal policy would be, in most cases of interest, to perform a reset at every time step. On

the other hand, if the reset penalty is too high, it may be beneficial to simply wait out until the process exits the set $\mathbb{D}$ and a penalty-free reset takes place. Although we will explore these situations, the main focus will be on cases that fall in between these two extremes.

The following notation is used throughout the chapter:

X_k	process at time k
U_k	control at time k
ρ	cost per active reset
$\mathbb{E}[Y]$	expectation of random variable Y
$\mathbb{P}(A)$	probability of event A

4.1.1 Problem Formulation

In what follows, we formalize the problem. Recall that X_k, $k = 0, 1, ...,$ is a Markov process that takes values in $\mathbb{R}^n$. Let R_k, $k = 0, 1, ...,$ be a sequence of independent identically distributed random variables, whose probability distribution is symmetric around 0. The denied environment is represented by the set $\mathbb{D}$, which is also symmetric around 0. If a reset (either active or passive) happens at time k, the value of the process at time $k+1$ is set to R_{k+1}. Recall that a passive reset happens whenever the process exists the set $\mathbb{D}$ whereas a controlled reset happens according to the binary reset variable U_k, which takes the value 1 for a controlled reset and 0 otherwise. Whenever a reset does not happen, the process evolves according to its own update function f, which may also depend on an auxiliary sequence of independent random variables.

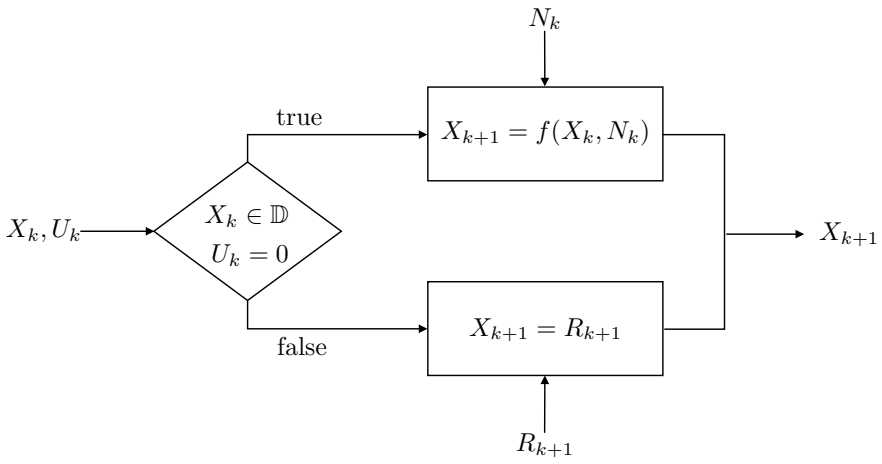


Figure 4.1: Diagram of process update rule

Let the process X_k evolve according to the following recursion:

$$X_{k+1} \stackrel{def}{=} \begin{cases} f(X_k, N_k), & U_k = 0 \text{ and } X_k \in \mathbb{D} \\ R_{k+1}, & U_k = 1 \text{ or } X_k \notin \mathbb{D} \end{cases} \tag{4.1}$$

for some real-valued function f and where N_k is an independent identically distributed sequence of random variables, which can be thought of as the noise in the system. The block diagram of the recursion rule is given in Fig. 4.1.

As previously discussed, the reset variable U_k cannot depend on the process value X_k since direct observations of the process are not available. It can, however, use knowledge of the last time the process exited the set $\mathbb{D}$. Therefore, the reset variable is determined according to a policy $\mathcal{U}$ of the following type:

$$U_k = \mathcal{U}\big(\{\mathcal{I}_j(X_j \notin \mathbb{D})\}_{j=0}^{k}\big) \tag{4.2}$$

where $\mathcal{U}$ maps the history up to time k of the indicator function $\mathcal{I}_j$ of the event $X_j \notin \mathbb{D}$ to a binary value 0 (wait) or 1 (reset).

For a given reset penalty ρ in $\mathbb{R}_+$, consider the following cost:

$$\mathcal{G}_{\mathcal{U}} \stackrel{def}{=} \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\sum_{k=1}^{N} g(X_k, U_k)\right], \tag{4.3}$$

where the reset variable U_k is determined according to the policy $\mathcal{U}$, and the state cost g is some non-negative real-valued function that has the following structure:

$$g(x,u) = \begin{cases} \rho & u = 1 \\ 0 & x \notin \mathbb{D} \\ \bar{g}(x) & \text{otherwise} \end{cases} \tag{4.4}$$

for some non-negative real-valued function $\bar{g} : \mathbb{R}^n \to \mathbb{R}_+$.

A stage cost of this type means that when a passive reset occurs ($x \notin \mathbb{D}$), no cost is incurred, whereas when a controlled reset takes place ($u = 1$), a cost of ρ is incurred. Finally when the stage is reset-free, a cost of $\bar{g}$ is incurred which depends only on the value of the process.

We are ready to state the main problem:

Problem 4.1.1. *Given the stochastic process and real-valued cost described in (4.1) and (4.3), a positive reset penalty ρ and a subset set $\mathbb{D}$ in $\mathbb{R}^n$, find the optimal reset policy $\mathcal{U}^*$ as in (4.2) such that:*

$$\mathcal{U}^* = \arg\min_{\mathcal{U}} \mathcal{G}_{\mathcal{U}} \tag{4.5}$$

4.2 Renewal Process

The first step in solving Problem 4.1.1 is to recognize that the underlining stochastic process is in fact a *renewal process*, since it renews when $U_k = 1$ or when

X_k exits the set $\mathbb{D}$. With this approach, instead of using the cost in (4.3), we consider an equivalent cost structure that uses the concept of *cycles*. The idea is simple: we rewrite the cost using the expectation of *a* cycle whose length is a random variable T, which is known as a *stopping time*, and function as the inter-arrival time of a cycle in the context of renewal reward processes. This means we can focus on only one cycle (of unknown length given by T) rather than the entirety of the whole infinite horizon. This is a known result in renewal reward process theory and is given below:

Lemma 4.2.1. *The cost in (4.3) is equivalent to the following cost:*

$$\mathcal{G}_{\mathcal{U}} = \frac{\mathbb{E}\left[\sum_{k=1}^{T} g(X_k, U_k)\right]}{\mathbb{E}[T]} \tag{4.6}$$

where the random variable T is the length of the cycle.

Proof. See [70], Theorem 3.6.1 □

From now on and with some abuse of notation, we will use the time index k to refer to time steps inside of a cycle rather than those in the complete infinite time horizon. Therefore whenever the cycle reaches time step $k = 4$, we can assume a reset has *not* happened in k= 0, 1, 2 or 3 (otherwise the cycle would have ended and a new one begun).

The probability distribution of the cycle length T will depend on the parameters of the process in (4.1) *and* on the reset policy $\mathcal{U}$, which is our design parameter. In fact the ability to shape the distribution of T through $\mathcal{U}$ is what differs Problem 4.1.1 from a standard renewal process problem.

It should now be clear that the reset policy $\mathcal{U}$ can only depend on how many time steps have elapsed without a reset having occurred. Moreover, any reset policy of interest can be parameterized by a single positive integer i, namely the number of time steps to wait without a passive reset having occurred before triggering a controlled reset. In other words, the reset policy parameterized by i at (cycle) time k is given by:

$$\mathcal{U}_i\big(\{\mathcal{I}_j(X_j \notin \mathbb{D})\}_{j=0}^{k}\big) = \begin{cases} 1, & k = i \text{ and } \sum_{j=0}^{k} \mathcal{I}_j(X_j \notin \mathbb{D}) = 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.7}$$

In light of this discussion, we can reparameterize the cost in terms of i, as follows:

$$\mathcal{G}_i = \frac{\mathbb{E}\left[\sum_{k=1}^{T_i} g(X_k, U_k)\right]}{\mathbb{E}[T_i]} \tag{4.8}$$

where U_k is determined according to the policy $\mathcal{U}_i$ in (4.7), and T_i is the appropriate stopping time for X_k under policy $\mathcal{U}_i$.

We are now ready to state an equivalent problem to Problem 4.1.1:

Problem 4.2.2. *Given the stochastic process and real-valued cost described in (4.1) and (4.8), a positive reset penalty ρ and a subset $\mathbb{D}$ in $\mathbb{R}^n$, find the optimal reset parameter i^*, such that:*

$$i^* = \arg\min_i \mathcal{G}_i \tag{4.9}$$

4.3 The Recursive Expression of $\mathcal{G}_i$

The main challenge in finding the optimal reset parameter lies in computing the expectations in $\mathcal{G}_i$, which in general require heavy computational effort, even for simple one dimensional cases with additive noise. It is therefore beneficial to explore the behavior of $\mathcal{G}_i$ as a function of the reset cost ρ and establish properties that can be exploited in the search for the optimal reset parameter. The goal of this section is to develop a recursive expression for $\mathcal{G}$. We begin by defining the following probabilities:

$$p_j \stackrel{def}{=} \mathbb{P}\Big(X_1^{j-1} \in \mathbb{D},\ X_j \notin \mathbb{D}\Big), \tag{4.10}$$

$$q_j \stackrel{def}{=} \mathbb{P}\Big(X_1^{j} \in \mathbb{D}\Big). \tag{4.11}$$

where $X_1^{j-1} \in \mathbb{D}$ is short-hand notation for the event: $X_1 \in \mathbb{D}, X_2 \in \mathbb{D}, ..., X_{j-1} \in \mathbb{D}$.

Note that p_j is the probability that the process exists the set $\mathbb{D}$ at time j but not before, while q_j is the probability that the process remains in $\mathbb{D}$ up to and including time j. These probabilities make up the probability mass function of the cycle length T_i associated with reset policy $\mathcal{U}_i$, which we can now define as follows:

$$T_i = \begin{cases} j & X_1^{j-1} \in \mathbb{D},\ X_j \notin \mathbb{D} \\ \\ i & X_1^{i} \in \mathbb{D} \end{cases} \tag{4.12}$$

Accordingly, the probability mass function of T_i is:

$$\mathbb{P}(T_i = j) = \begin{cases} p_j & j < i \\ \\ p_i + q_i & j = i. \end{cases} \tag{4.13}$$

The next major step is to construct a straightforward but very useful recursion of $\mathcal{G}_i$, but before doing so it is important to establish two useful relationships.

Lemma 4.3.1. *The following holds:*

i) $q_i = p_{i+1} + q_{i+1}$,

ii) $\mathbb{E}\big[T_{i+1}\big] = \mathbb{E}\big[T_i\big] + q_i$, *where* $\mathbb{E}\big[T_1\big] = 1$.

Proof. Item i follows from the law of total probability. For item ii, note the following:

$$\begin{aligned}
\mathbb{E}[T_{i+1}] &\stackrel{def}{=} \sum_{j=1}^{i} jp_j + (i+1)(p_{i+1} + q_{i+1}) \\
&= \sum_{j=1}^{i} jp_j + (i+1)q_i && \text{(Lemma 4.3.1, item } i\text{)} \\
&= \sum_{j=1}^{i} jp_j + iq_i + q_i \\
&= \sum_{j=1}^{i-1} jp_j + i(p_i + q_i) + q_i \\
&= \mathbb{E}[T_i] + q_i
\end{aligned}$$

□

The next lemma provides a recursive method to compute $\mathcal{G}_i$.

Lemma 4.3.2. *The cost function* $\mathcal{G}_i$ *in (4.8) can be computed recursively as follows:*

$$\mathcal{G}_{i+1} = \frac{t_i}{t_{i+1}}\mathcal{G}_i + \frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1}\Big\} \tag{4.14}$$

where $t_i \stackrel{def}{=} \mathbb{E}[T_i]$, $\eta_i \stackrel{def}{=} \mathbb{E}\big[\bar{g}(X_i) \;\big|\; X_1^i \in \mathbb{D}\big]$ *and* $\mathcal{G}_1 = \rho q_1$.

The usefulness of this lemma is two-fold. First it provides a much more efficient way of computing $\mathcal{G}_i$ for all i than the original form. The recursive expression relies on the computation of two difficult-to-compute parameters (for each i): the probabilities q_i's (from which one can compute the probabilities p_i's by applying Lemma 4.3.1, item i); and the expectations $\eta_i = \mathbb{E}\big[\bar{g}(X_i) \;\big|\; X_1^i \in \mathbb{D}\big]$. Second, it makes explicit the dependence of the evolution of $\mathcal{G}_i$ on the reset penalty ρ, which will be of great importance in the next section when we establish the main result of the chapter.

We proceed by proving the lemma:

Proof of Lemma 4.3.2. Recall that the cost $\mathcal{G}_i$ in (4.8) is given by:

$$\mathcal{G}_i = \frac{\mathbb{E}\Big[\sum_{k=1}^{T_i} g(X_k, U_k)\Big]}{\mathbb{E}[T_i]},$$

and let $v_i \stackrel{def}{=} \mathbb{E}\big[\sum_{k=1}^{T_i} g(X_k, U_k)\big]$, so that the cost can written as:

$$\mathcal{G}_i = \frac{v_i}{t_i}. \tag{4.15}$$

The expression for each v_i can be expanded by conditioning on all the events that would trigger a reset:

$$v_i = \sum_{j=1}^{i} \mathbb{E}\left[\sum_{k=1}^{T_i} g(X_k, U_k) \;\middle|\; X_1^{j-1} \in \mathbb{D},\ X_j \notin \mathbb{D}\right] p_j + \mathbb{E}\left[\sum_{k=1}^{T_i} g(X_k, U_k) \;\middle|\; X_1^{j} \in \mathbb{D}\right] q_i \tag{4.16}$$

$$= \sum_{j=1}^{i} \mathbb{E}\left[\sum_{k=1}^{j-1} \bar{g}(X_k) \;\middle|\; X_1^{j-1} \in \mathbb{D},\ X_j \notin \mathbb{D}\right] p_j + \mathbb{E}\left[\sum_{k=1}^{i-1} \bar{g}(X_k) + \rho \;\middle|\; X_1^{i} \in \mathbb{D}\right] q_i \tag{4.17}$$

where the first term is obtained by noticing that for each j a penalty-free reset takes place at time j and a cost of $\bar{g}$ is incurred up until time $j-1$ (since $g(x,u)=0$ when $x \notin \mathbb{D}$). The second term represents the only instance a controlled reset happens under policy $\mathcal{U}_i$ and a penalty of ρ is incurred (along with the stage costs $\bar{g}$ until time $i-1$).

For simplicity of notation, we define:

$$h_j^p \stackrel{def}{=} \mathbb{E}\left[\sum_{k=1}^{j-1} \bar{g}(X_k) \;\middle|\; X_1^{j-1} \in \mathbb{D},\ X_j \notin \mathbb{D}\right], \text{and} \tag{4.18}$$

$$h_i^c \stackrel{def}{=} \mathbb{E}\left[\sum_{k=1}^{i-1} \bar{g}(X_k) \;\middle|\; X_1^{i} \in \mathbb{D}\right], \tag{4.19}$$

where h_j^p is the expectation of sum of the stage costs conditioned on the event that a *passive* reset happens at time j and h_i^c is the expectation conditioned on a *controlled* reset (at time i), without accounting for the reset penalty ρ. Therefore the expression for v_i can be written as follows:

$$v_i = \sum_{j=1}^{i} h_j^p p_j + h_i^c q_i + \rho q_i. \tag{4.20}$$

We proceed by taking the difference $\mathcal{G}_{i+1} - \mathcal{G}_i$:

$$\begin{aligned}
\mathcal{G}_{i+1} - \mathcal{G}_i &= \frac{v_{i+1}}{t_{i+1}} - \frac{v_i}{t_i} \\
&= \frac{1}{t_{i+1}t_i}\Big\{v_{i+1}t_i - v_i t_{i+1}\Big\} \\
&= \frac{1}{t_{i+1}t_i}\Big\{v_{i+1}t_i - v_i(t_i + q_i)\Big\} \\
&= \frac{1}{t_{i+1}t_i}\Big\{(v_{i+1} - v_i)t_i - v_i q_i)\Big\}
\end{aligned} \tag{4.21}$$

The difference $v_{i+1} - v_i$ can be simplified as follows:

$$\begin{aligned}
v_{i+1} - v_i &= \sum_{j=1}^{i+1} h_j^p p_j + h_{i+1}^c q_{i+1} + \rho q_{i+1} - \Big(\sum_{j=1}^{i} h_j^p p_j + h_i^c q_i + \rho q_i \Big) \\
&= h_{i+1}^p p_{i+1} + h_{i+1}^c q_{i+1} + \rho q_{i+1} - h_i^c q_i - \rho q_i \\
&= h_{i+1}^p p_{i+1} + h_{i+1}^c q_{i+1} - h_i^c q_i + \rho(q_{i+1} - q_i) \\
&= h_{i+1}^p p_{i+1} + h_{i+1}^c q_{i+1} - h_i^c q_i - \rho p_{i+1} \qquad \text{(Lemma 4.3.1, item } i\text{)} \\
&= h_{i+1}^p \mathbb{P}(X_{i+1} \notin \mathbb{D} \mid X_1^i \in \mathbb{D}) q_i + h_{i+1}^c \mathbb{P}(X_{i+1} \in \mathbb{D} \mid X_1^i \in \mathbb{D}) q_i \\
&\qquad - h_i^c q_i - \rho p_{i+1} \qquad \text{(Baye's rule)} \\
&= \big(h_{i+1}^p \mathbb{P}(X_{i+1} \notin \mathbb{D} \mid X_1^i \in \mathbb{D}) + h_{i+1}^c \mathbb{P}(X_{i+1} \in \mathbb{D} \mid X_1^i \in \mathbb{D}) \big) q_i \\
&\qquad - h_i^c q_i - \rho p_{i+1}
\end{aligned}$$

The term in parenthesis can be further reduced as follows:

$$\begin{aligned}
&h_{i+1}^p \mathbb{P}(X_{i+1} \notin \mathbb{D} \mid X_1^i \in \mathbb{D}) + h_{i+1}^c \mathbb{P}(X_{i+1} \in \mathbb{D} \mid X_1^i \in \mathbb{D}) = \\
&\quad = \mathbb{E}\left[\sum_{k=1}^{i} \bar{g}(X_k) \;\Big|\; X_1^i \in \mathbb{D},\ X_{i+1} \notin \mathbb{D} \right] \mathbb{P}(X_{i+1} \notin \mathbb{D} \mid X_1^i \in \mathbb{D}) + \\
&\qquad \mathbb{E}\left[\sum_{k=1}^{i} \bar{g}(X_k) \;\Big|\; X_1^{i+1} \in \mathbb{D} \right] \mathbb{P}(X_{i+1} \in \mathbb{D} \mid X_1^i \in \mathbb{D}) = \\
&\quad = \mathbb{E}\left[\sum_{k=1}^{i} \bar{g}(X_k) \;\Big|\; X_1^i \in \mathbb{D} \right]
\end{aligned}$$

where the last step is found by the law of total expectation. Therefore $v_{i+1} - v_i$ equals:

$$\begin{aligned}
v_{i+1} - v_i &= \mathbb{E}\left[\sum_{k=1}^{i} \bar{g}(X_k) \;\Big|\; X_1^i \in \mathbb{D} \right] q_i - h_i^c q_i - \rho p_{i+1} \\
&= \left(\mathbb{E}\left[\sum_{k=1}^{i} \bar{g}(X_k) \;\Big|\; X_1^i \in \mathbb{D} \right] - \mathbb{E}\left[\sum_{k=1}^{i-1} \bar{g}(X_k) \;\Big|\; X_1^i \in \mathbb{D} \right] \right) q_i - \rho p_{i+1} \\
&= \mathbb{E}\Big[\bar{g}(X_i) \;\Big|\; X_1^i \in \mathbb{D} \Big] q_i - \rho p_{i+1}.
\end{aligned}$$

Define $\eta_i \stackrel{def}{=} \mathbb{E}\big[\bar{g}(X_i) \;\big|\; X_1^i \in \mathbb{D}\big]$ and we reach the simplest expression for $v_{i+1} - v_i$:

$$v_{i+1} - v_i = \eta_i q_i - \rho p_{i+1},$$

Continuing from Eq. (4.21) we have:

$$\begin{aligned}
\mathcal{G}_{i+1} - \mathcal{G}_i &= \frac{1}{t_{i+1}t_i}\Big\{(\eta_i q_i - \rho p_{i+1})t_i - v_i q_i)\Big\} \\
&= \frac{1}{t_{i+1}}\Big\{(\eta_i q_i - \rho p_{i+1}) - \frac{v_i}{t_i} q_i)\Big\} \\
&= \frac{1}{t_{i+1}}\Big\{(\eta_i q_i - \rho p_{i+1}) - \mathcal{G}_i q_i)\Big\} \qquad (4.22)
\end{aligned}$$

Finally, rearranging terms we arrive at the desired recursive expression:

$$\begin{aligned}
\mathcal{G}_{i+1} &= \mathcal{G}_i + \frac{1}{t_{i+1}}\Big\{(\eta_i q_i - \rho p_{i+1}) - \mathcal{G}_i q_i)\Big\} \\
&= \frac{1}{t_{i+1}}\Big\{\mathcal{G}_i t_{i+1} - \mathcal{G}_i q_i)\Big\} + \frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1}\Big\} \\
&= \frac{1}{t_{i+1}}\Big\{\mathcal{G}_i (t_{i+1} - q_i)\Big\} + \frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1}\Big\} \\
&= \frac{1}{t_{i+1}}\Big\{\mathcal{G}_i t_i\Big\} + \frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1}\Big\} \quad \text{(Lemma 4.3.1, item } ii\text{)} \\
&= \frac{t_i}{t_{i+1}}\mathcal{G}_i + \frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1}\Big\}
\end{aligned}$$

□

4.4 Finding the Optimal Reset Parameter

The search for the optimal reset parameter is theoretically simple since $\mathcal{G}_i$ maps the natural numbers to the real line. However, computing the values of $\mathcal{G}_i$, even with the aide of the recursive expression found in Lemma 4.3.2, is computationally

daunting. Therefore, it is desirable to investigate properties of $\mathcal{G}_i$ and to establish conditions that provide guarantees relating to the optimal reset parameter. In this section, we establish sufficient conditions that guarantees that if a local minimizer exists it is global. In situations like this, the optimal reset parameter can be found, as we will discuss later, by recursively computing $\mathcal{G}_i$ and comparing it with the previous value.

Theorem 4.4.1. *Consider Problem 5.1.3 with stage cost $\bar{g}$ and reset penalty ρ. Define the function function $r : \mathbb{R}^n \to \mathbb{R}$ where:*

$$r(x) \stackrel{def}{=} \bar{g}(x) - \rho \mathbb{P}(f(x, N) \notin \mathbb{D}) \tag{4.23}$$

where N is a random variable with the same distribution as N_i's used in Eq. (4.1). Further suppose that:

$$\mathbb{E}\big[r(X_i) \mid X_1^i \in \mathbb{D}\big] \quad \leq \quad \mathbb{E}\big[r(X_{i+1}) \mid X_1^{i+1} \in \mathbb{D}\big] \tag{4.24}$$

for all i.

The following holds: if there exists a parameter i such that $\mathcal{G}_i \geq \mathcal{G}_{i-1}$, then $\mathcal{G}_j \geq \mathcal{G}_{j-1}$ for all $j > i$

Theorem 4.4.1 states that, under certain conditions, once the function $\mathcal{G}_i$ starts increasing it always increases from then on. We proceed with the proof of the theorem.

Proof of Theorem 4.4.1. Recall from Eq. (4.22) the expression for the difference $\mathcal{G}_{i+1} - \mathcal{G}_i$ used in the proof of Lemma 4.3.2:

$$\mathcal{G}_{i+1} - \mathcal{G}_i = \frac{1}{t_{i+1}} \Big\{ (\eta_i q_i - \rho p_{i+1}) - \mathcal{G}_i q_i) \Big\} \tag{4.25}$$

By assumption, there exists a reset parameter i and an $\epsilon \geq 0$ such that $\mathcal{G}_i - \mathcal{G}_{i-1} = \epsilon$, thus we have:

$$\frac{1}{t_i}\Big\{\eta_{i-1}q_{i-1} - \rho p_i - \mathcal{G}_{i-1}q_{i-1})\Big\} = \epsilon$$

$$\eta_{i-1}q_{i-1} - \rho p_i - \mathcal{G}_{i-1}q_{i-1} = t_i\epsilon$$

$$\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho - \mathcal{G}_{i-1} = \frac{t_i}{q_{i-1}}\epsilon \tag{4.26}$$

Assume by contradiction that there exists a $\delta > 0$ such that $\mathcal{G}_{i+1} - \mathcal{G}_i = -\delta$ and therefore:

$$\frac{1}{t_{i+1}}\Big\{\eta_i q_i - \rho p_{i+1} - \mathcal{G}_i q_i\Big\} = -\delta$$

$$\eta_i q_i - \rho p_{i+1} - \mathcal{G}_i q_i = -t_{i+1}\delta$$

$$\eta_i - \frac{p_{i+1}}{q_i}\rho - \mathcal{G}_i = -\frac{t_{i+1}}{q_i}\delta \tag{4.27}$$

Subtracting Eq. (4.27) from Eq. (4.26):

$$\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho - \mathcal{G}_{i-1} - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho - \mathcal{G}_i\Big) = \frac{t_i}{q_{i-1}}\epsilon - \Big(-\frac{t_{i+1}}{q_i}\delta\Big)$$

$$\Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big) - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) - \mathcal{G}_{i-1} + \mathcal{G}_i = \frac{t_i}{q_{i-1}}\epsilon + \frac{t_{i+1}}{q_i}\delta$$

$$\Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big) - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) = \mathcal{G}_{i-1} - \mathcal{G}_i + \frac{t_i}{q_{i-1}}\epsilon + \frac{t_{i+1}}{q_i}\delta$$

$$\Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big) - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) = -\epsilon + \frac{t_i}{q_{i-1}}\epsilon + \frac{t_{i+1}}{q_i}\delta$$

$$\Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big) - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) = \frac{t_i - q_{i-1}}{q_{i-1}}\epsilon + \frac{t_{i+1}}{q_i}\delta$$

$$\Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big) - \Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) = \frac{t_{i-1}}{q_{i-1}}\epsilon + \frac{t_{i+1}}{q_i}\delta$$

Since both terms in the right hand side of the last expression are assumed to be

positive, we conclude that:

$$\Big(\eta_i - \frac{p_{i+1}}{q_i}\rho\Big) < \Big(\eta_{i-1} - \frac{p_i}{q_{i-1}}\rho\Big), \tag{4.28}$$

Next, we note that:

$$\begin{aligned}\frac{p_{i+1}}{q_i} &= \frac{\mathbb{P}\Big(X_1^i \in \mathbb{D},\ X_{i+1} \notin \mathbb{D}\Big)}{\mathbb{P}\Big(X_1^i \in \mathbb{D}\Big)}, \\ &= \mathbb{P}\Big(X_{i+1} \notin \mathbb{D} \ \Big|\ X_1^i \in \mathbb{D}\Big),\end{aligned}$$

and therefore:

$$\begin{aligned}\eta_i - \frac{p_{i+1}}{q_i}\rho &= \mathbb{E}\Big[\bar{g}(X_i) \ \Big|\ X_1^i \in \mathbb{D}\Big] - \rho\mathbb{P}\Big(X_{i+1} \notin \mathbb{D} \ \Big|\ X_1^i \in \mathbb{D}\Big) \\ &= \mathbb{E}\Big[\bar{g}(X_i) - \rho\mathcal{I}(X_{i+1} \notin \mathbb{D}) \ \Big|\ X_1^i \in \mathbb{D}\Big] \\ &= \mathbb{E}\Big[\bar{g}(X_i) - \rho\mathcal{I}(f(X_i, N_i) \notin \mathbb{D}) \ \Big|\ X_1^i \in \mathbb{D}\Big] \\ &= \mathbb{E}_{X_i|X_1^i \in \mathbb{D}}\Big[\mathbb{E}\Big[\bar{g}(X_i) - \rho\mathcal{I}(f(X_i, N_i) \notin \mathbb{D}) \ \Big|\ X_1^i \in \mathbb{D}, X_i\Big]\Big] \\ &= \mathbb{E}_{X_i|X_1^i \in \mathbb{D}}\Big[\mathbb{E}\Big[\bar{g}(X_i) - \rho\mathcal{I}(f(X_i, N_i) \notin \mathbb{D}) \ \Big|\ X_i\Big]\Big] \\ &= \mathbb{E}_{X_i|X_1^i \in \mathbb{D}}\Big[\bar{g}(X_i) - \rho\mathbb{P}(f(X_i, N_i) \notin \mathbb{D})\Big]\end{aligned}$$

Define the function $r(x) \stackrel{def}{=} \bar{g}(x) - \rho\mathbb{P}(f(x, N) \notin \mathbb{D})$. The inequality in (4.28) becomes:

$$\mathbb{E}_{X_i|X_1^i \in \mathbb{D}}\Big[r(X_i)\Big] < \mathbb{E}_{X_{i-1}|X_1^{i-1} \in \mathbb{D}}\Big[r(X_{i-1})\Big]$$

which is a contradiction, by the assumption in (4.24).

The same method can be recursively applied to show that $\mathcal{G}_j \geq \mathcal{G}_{j-1}$ for all $j > i$. □

A direct result of Theorem 4.4.1 related to the optimal parameter is the following:

Corollary 4.4.2. *Under the same conditions of Theorem 4.4.1, if a local minimum exists it is global.*

Proof. Follows directly from Theorem 4.4.1 □

The optimal reset parameter can be found using the scheme described in Fig.4.2.

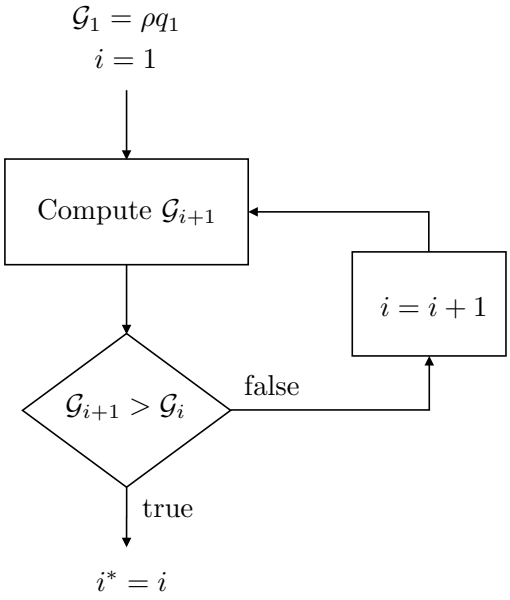


Figure 4.2: Finding the optimal reset parameter

It is also useful to establish properties concerning cases when the optimal strategy is to always perform controlled-reset:

Corollary 4.4.3. *Consider the same conditions as in Theorem 4.4.1, and suppose*

the reset penalty ρ satisfies the following condition:

$$\rho \leq \frac{q_1 \eta_1}{p_2 + q_1^2} \tag{4.29}$$

The following holds: the policy to always reset is optimal, i.e. $i^* = 1$.

Proof. We proceed by showing that if ρ satisfies inequality (4.29), then it must be that $\mathcal{G}_2 - \mathcal{G}_1 \geq 0$ and, by Theorem 4.4.1, $\mathcal{G}_{i+1} - \mathcal{G}_i \geq 0$ for all i. From inequality (4.29), we get:

$$\begin{aligned}
q_1 \eta_1 &\geq (p_2 + q_1^2)\rho \\
q_1 \eta_1 &\geq p_2 \rho + q_1 \mathcal{G}_1 \\
q_1 \eta_1 - p_2 \rho - q_1 \mathcal{G}_1 &\geq 0 \\
\frac{1}{t_2}\Big\{ q_1 \eta_1 - p_2 \rho - q_1 \mathcal{G}_1 \Big\} &\geq 0 \\
\mathcal{G}_2 - \mathcal{G}_1 &\geq 0
\end{aligned}$$

where the last line follows from Eq. (4.22), and the result follows. □

4.5 Radial Stochastic Order

Recall condition (4.24) in Theorem 4.4.1, repeated below for convenience:

$$\mathbb{E}\big[r(X_i) \mid X_1^i \in \mathbb{D}\big] \quad \leq \quad \mathbb{E}\big[r(X_{i+1}) \mid X_1^{i+1} \in \mathbb{D}\big]$$

for all i.

Checking if this condition is satisfied for a set of problem parameters may be difficult or even impractical. Therefore it is important to establish easy-to-check conditions for which the inequality in (4.24) holds for all i. One approach to

accomplish this is to establish a *stochastic order* of the conditional process $\{X_i | X_1^i \in \mathbb{D}\}$.

Stochastic orders are partial orders and come in many different flavors. At their essence, they provide a way to *compare* random variables or analyze the evolution of stochastic processes. For example, by the *usual stochastic order* in $\mathbb{R}$, the random variable X is less than the random variable Y if $\mathbb{P}(X < \tau) \geq \mathbb{P}(Y < \tau)$ for all τ, and when this holds, $\mathbb{E}[l(X)] \leq \mathbb{E}[l(Y)]$ for all nondecreasing functions $l : \mathbb{R} \to \mathbb{R}$.

While the concept of the usual stochastic order in $\mathbb{R}$ is straightforward, their vector counterparts face many challenges, especially in regards to the expectation property for nondecreasing functions, which is crucial for our purposes. Therefore, from now on we restrict ourselves to the class of problems with *radial* symmetry where the denied set $\mathbb{D}$ in $\mathbb{R}^n$ is given by:

$$\mathbb{D} = \{x \in \mathbb{R}^n \mid \|x\| < w\} \tag{4.30}$$

where $\|\cdot\|$ represents the Euclidean norm in $\mathbb{R}^n$.

Definiton 4.5.1. *A stochastic process X_k in $\mathbb{R}^n$ is said to be increasing in the radial stochastic order, written $X_k \leq_r X_{k+1}$, if*

$$\int_{\|x\| \leq t} f_{X_k}(x)dx \quad \geq \quad \int_{\|x\| \leq t} f_{X_{k+1}}(x)dx \tag{4.31}$$

for all $k \geq 1$ and all t in $\mathbb{R}_+$; where f_{X_k} is the joint probability density function (`pdf`*) of the process at time k.*

In other words, if $X_k \leq_r X_{k+1}$ for all k, the probability of X_k being within any radius t of the origin does not increase. A very useful result in the study of

stochastic orders refers to the expectation of nondecreasing functions. Here we focus on functions that are radially nondecreasing, which we define next.

Definiton 4.5.2. *A function $l : \mathbb{R}^n \to \mathbb{R}$ is radially nondecreasing function when, for all x and y with $||x|| \leq ||y||$, the following holds:*

$$l(x) \leq l(y) \tag{4.32}$$

Lemma 4.5.3. *Suppose Y_k is a vector stochastic process in $\mathbb{R}^n$ with increasing stochastic order, and let $l : \mathbb{R}^n \to \mathbb{R}$ be a radially nondecreasing function. The following holds:*

$$\mathbb{E}[l(Y_{k+1})] \geq \mathbb{E}[l(Y_k)] \tag{4.33}$$

for all i.

Proof. The proof is straigthfoward by defining a new stochastic process Z_k where $Z_k = ||Y_k||$ and a function $\tilde{l} : \mathbb{R}_+ \to \mathbb{R}$ where: $\tilde{l}(z) \stackrel{def}{=} l(y)$ for any y such that $||y|| = z$. Clearly, Z_k has increasing *usual stochastic order* and $\tilde{l}$ is nondecreasing, and the result follows by Theorem 1.2.8 in [71]. □

Corollary 4.5.4. *Condition (4.24) in Theorem 4.4.1 is satisfied for all i if the following holds:*

- *The conditioned process X_k given that $X_1^k \in \mathbb{D}$ has increasing stochastic order.*
- *The function $r : \mathbb{R}^n \to \mathbb{R}$ where:*

$$r(x) \stackrel{def}{=} \bar{g}(x) - \rho\mathbb{P}(f(x, N) \notin \mathbb{D}) \tag{4.34}$$

  is radially nondecreasing.

Proof. Follows from Lemma 4.5.3. □

4.6 Concrete Cases: the Real Line

To illustrate the results of this chapter, we analyze a case with a real-valued stochastic process with additive noise. The denied set is given by:

$$\mathbb{D} = \{x \in \mathbb{R} \ : \ -w \leq x \leq w\} \tag{4.35}$$

for some w in $\mathbb{R}_+$.

The stage cost is given by $\bar{g}(x) = x^2$, and the process has the following update rule:

$$X_{k+1} = \begin{cases} X_k + N_k, & U_k = 0 \text{ and } X_k \in \mathbb{D} \\ R_k \ , & U_k = 1 \text{ or } X_k \notin \mathbb{D} \end{cases} \tag{4.36}$$

where R_k and N_k are two independent sequences of independent and identically distributed random variables

4.6.1 Uniform noise

Suppose R_k and N_k are *uniform* with support in $[-a, a]$, for some a in $\mathbb{R}_+$. The purpose of this example is to illustrate that depending on the value of a when compared to w, the process X_k conditioned on the event $X_1^k \in \mathbb{D}$ *may not* satisfy the stochastic ordering condition of Theorem 4.4.1. For example, suppose a and w are such that $w/2 < a \leq w$. The probability density function of X_1 given $X_1 \in \mathbb{D}$ is simply the uniform density since $a \leq w$. The density function of X_2 given $X_1^2 \in \mathbb{D}$

is given below:

$$f_{X_2|X_1^2\in\mathbb{D}}(x) = \begin{cases} \frac{x+2a}{w(4a-w)} & -w \le x \le 0 \\ \frac{-x+2a}{w(4a-w)} & 0 \le x \le w \\ 0 & \text{otherwise} \end{cases} \tag{4.37}$$

The plots of the density functions in question can be seen in Fig. 4.3.

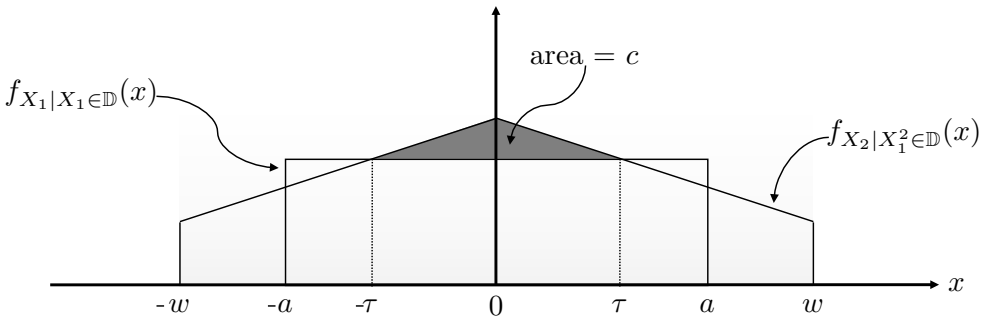


Figure 4.3: Example where the stochastic ordering condition in Theorem 4.4.1 is not satisfied.

It is immediate that:

$$\frac{2a}{w(4a-w)} = f_{X_2|X_1^2\in\mathbb{D}}(0) \quad > \quad f_{X_1|X_1\in\mathbb{D}}(0) = \frac{1}{2a} \tag{4.38}$$

for all values of a and w that satisfy $\frac{w}{2} < a \le w$, and therefore there exists some τ such that:

$$\mathbb{P}(-\tau \le X_1 \le \tau \mid X_1 \in \mathbb{D}) \quad < \quad \mathbb{P}(-\tau \le X_2 \le \tau \mid X_1^2 \in \mathbb{D}), \tag{4.39}$$

One example of such τ is depicted in Fig. 4.3, for which:

$$\mathbb{P}(-\tau \le X_2 \le \tau \mid X_1^2 \in \mathbb{D}) - \mathbb{P}(-\tau \le X_1 \le \tau \mid X_1 \in \mathbb{D}) = c > 0. \tag{4.40}$$

for some positive c highlighted in Fig. 4.3.

Therefore $X_1|X_1 \in \mathbb{D} \ >_{st} \ X_2|X_1^2 \in \mathbb{D}$ and Theorem 4.4.1 cannot be applied.

With the aid of numerical computation, it can be shown that the stochastic ordering condition of Theorem 4.4.1 is satisfied whenever:

$$a \leq \frac{w}{2} \tag{4.41}$$

Due to the nature of convolutions that involve the uniform density, it is tedious to demonstrate this analytically, though it could be done by showing that, when the condition (4.41) is satisfied, the `pdf`s of two consecutive (conditioned) random variable cross only once in the interval $[0, w]$.

The constraint that $a \leq w/2$ is in fact a realistic one, since in most cases of interest the event that the agent leaves the set $\mathbb{D}$ in one time step has very low (or zero) probability.

Suppose therefore that $a \leq w/2$ and therefore the stochastic ordering condition of Theorem 4.4.1 is satisfied. In order to satisfy the second condition of the theorem, consider a uniform random variable N with support in $[-a, a]$ and the function:

$$r(x) = x^2 - \rho \mathbb{P}(|x + N| > w) \tag{4.42}$$

The second term in r is given by:

$$\mathbb{P}\big(|x + N| > w\big) = \begin{cases} 0 & 0 \leq x < w - a \\ \\ \frac{1}{2a}x - \frac{w-a}{2a} & w - a \leq x \leq w, \end{cases} \tag{4.43}$$

and therefore:

$$r(x) = \begin{cases} x^2 & 0 \leq x < w - a \\ x^2 - \frac{\rho}{2a}x + \frac{\rho(w-a)}{2a} & w - a \leq x \leq w \end{cases} \tag{4.44}$$

Clearly, r is increasing in $[0, w - a]$. To guarantee that r is nondecreasing in the entire interval, the derivative of $r(x)$ must be nonnegative in said interval, and therefore ρ needs to satisfy:

$$\begin{aligned} 2x - \frac{\rho}{2a} &\geq 0 \\ 2x &\geq \frac{\rho}{2a} \\ \rho &\leq 4xa \end{aligned}$$

for all x. This will hold in the interval $[w - a, w]$ when:

$$\rho \leq 4a(w - a). \tag{4.45}$$

Moreover, from Corollary 4.4.3 we know that an optimal policy is to always reset (that is, $i^* = 1$ when the reset penalty satisfies:

$$\rho \leq \frac{a^2}{3}. \tag{4.46}$$

Example 4.6.1. *Consider the denied set $\mathbb{D}$ given in (4.35) where $w = 10$, that is:*

$$\mathbb{D} = \{x \in \mathbb{R} \ : \ -10 \leq x \leq 10\}$$

The process is updated according to (4.36) where the noise and reset variables are uniform with support in $[-3, 3]$ (that is, $a = 3$). Finally, $\bar{g}(x) = x^2$.

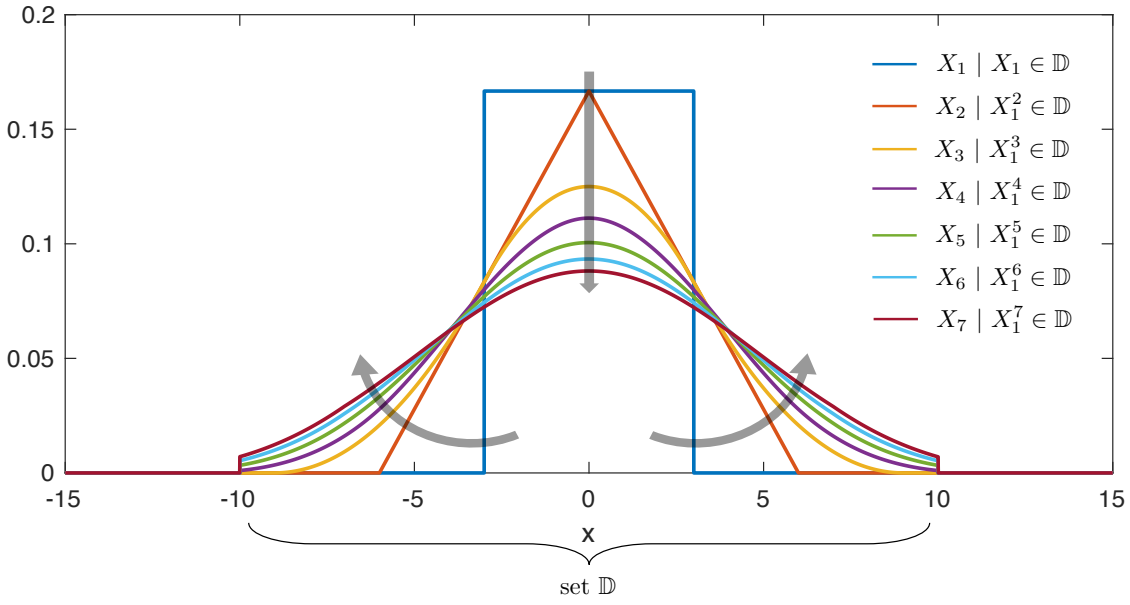


Figure 4.4: Conditioned `pdf`s for Example 4.6.1

We begin by checking the stochastic ordering condition of Theorem 4.4.1. Fig. 4.4 shows the evolution of the first seven conditioned `pdf`s of interest, where it can be checked by inspection that the stochastic order condition is satisfied.

The next step is to consider the values of ρ for which the function r given in (4.44) is nondecreasing. By the inequality in (4.45), ρ must be less than or equal to 84. Fig. 4.5 shows the plots of the function r for three values of ρ: 70, 84, and 100. As expected when $\rho = 100$, the function does not satisfy the nondecreasing condition.

Finally, Fig. 4.6 shows the cost $\mathcal{G}_i$ for four values of ρ for which the function r satisfies the nondecreasing requirement. As expected, when $\rho = 2 \leq 3 = a^2/3$ the optimal reset parameter is $i^* = 1$.

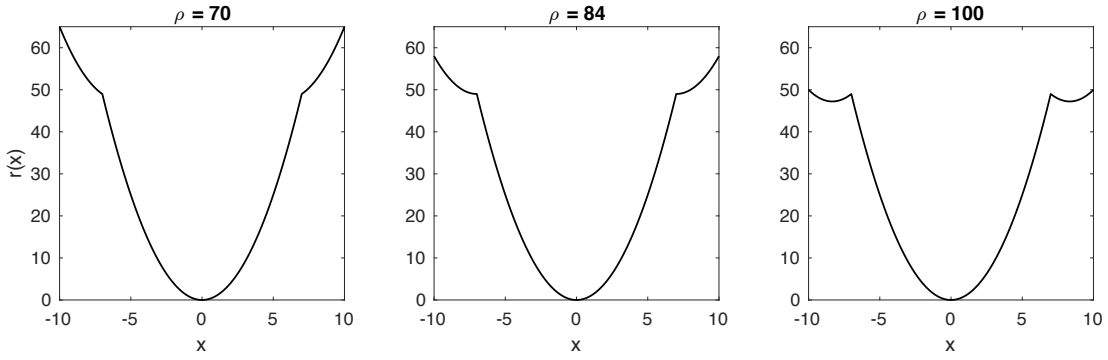


Figure 4.5: Plots of r in Example 4.6.1 for three different values of ρ: 70, 84, and 100. As expected when $\rho = 100$, the function does not satisfy the nondecreasing condition.

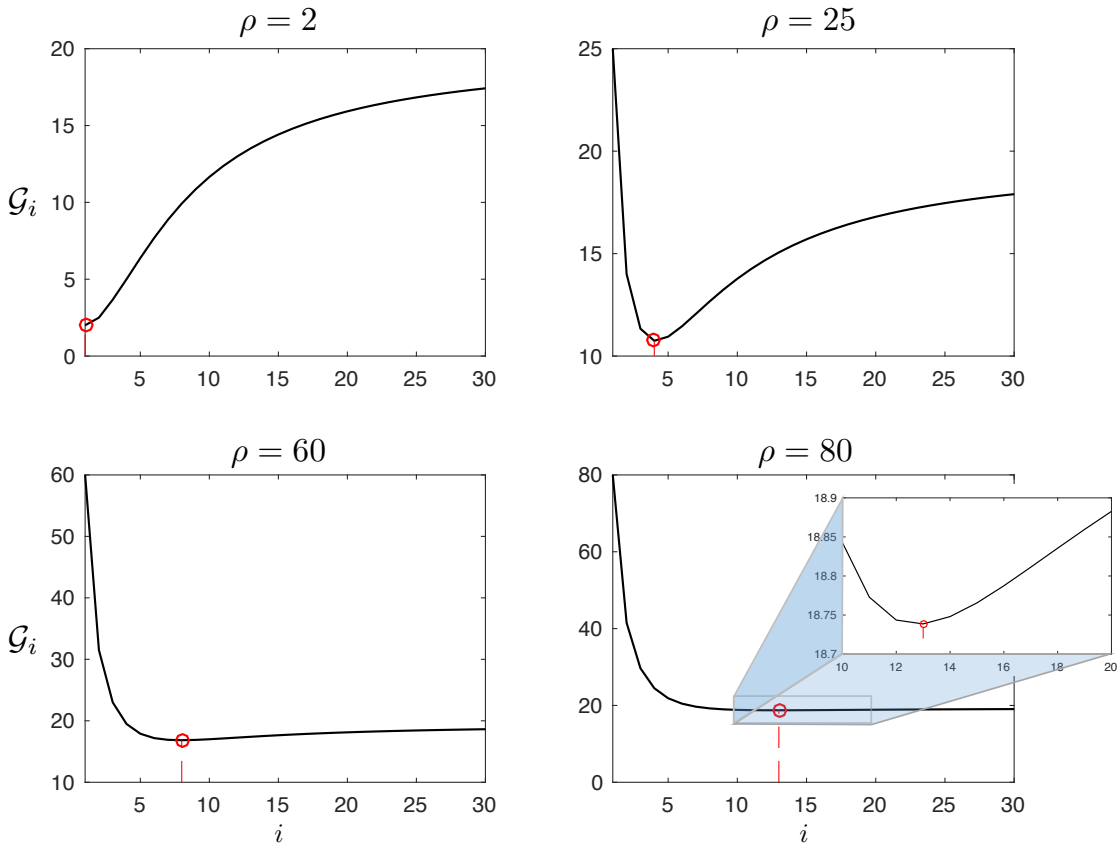


Figure 4.6: $\mathcal{G}_i$ in Example 4.6.1 for four different values of ρ: 2, 25, 60, and 80.

4.6.2 Gaussian Noise

Recall from the beginning of the this section that we have a denied set $\mathbb{D}$ on the real line given by $\mathbb{D} = \{x \in \mathbb{R} \ : \ -w \leq x \leq w\}$ for some w in $\mathbb{R}_+$, that the stage cost is given by $\bar{g}(x) = x^2$, and the process has the following update rule:

$$X_{k+1} = \begin{cases} X_k + N_k, & U_k = 0 \text{ and } X_k \in \mathbb{D} \\ R_k \ , & U_k = 1 \text{ or } X_k \notin \mathbb{D}. \end{cases}$$

In this subsection, R_k and N_k are two independent sequences of independent and identically distributed *Gaussian* random variables with zero mean and variance σ^2.

What distinguishes the Gaussian from the uniform case is that the conditional `pdf`s of interest will always satisfy the stochastic ordering condition for any combination of w and σ. As with the uniform case, this is difficult to show analytically, but has been verified numerically.

With regards to the function r, we can calculate the following probability:

$$\mathbb{P}\big(|N + x| > w\big) = 1 + \frac{1}{2}\left[\text{erf}\left(\frac{-w - x}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{w - x}{\sigma\sqrt{2}}\right)\right], \tag{4.47}$$

and write the r as follows:

$$r(x) = x^2 - \rho\left\{1 + \frac{1}{2}\left[\text{erf}\left(\frac{-w - x}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{w - x}{\sigma\sqrt{2}}\right)\right]\right\}$$

The first derivative of r is given by:

$$r'(x) = 2x - \frac{\rho}{\sqrt{2\pi}\sigma}\left[- e^{\frac{-(w+x)^2}{2\sigma^2}} + e^{\frac{-(w-x)^2}{2\sigma^2}}\right]$$

Therefore, the function r is nonnegative in $[0, w]$ for values of ρ that satisfy the following for all values of x in $[0, w]$:

$$\rho \leq 2\sqrt{2\pi}\sigma x \left[- e^{\frac{-(w+x)^2}{2\sigma^2}} + e^{\frac{-(w-x)^2}{2\sigma^2}} \right]^{-1}$$

Therefore:

$$\rho \leq \min_{0 \leq x \leq w} 2\sqrt{2\pi}\sigma x \left[- e^{\frac{-(w+x)^2}{2\sigma^2}} + e^{\frac{-(w-x)^2}{2\sigma^2}} \right]^{-1} \tag{4.48}$$

Example 4.6.2. *Consider the denied set* $\mathbb{D}$ *given in (4.35) where* $w = 10$*, that is:*

$$\mathbb{D} = \{x \in \mathbb{R} \ : \ -10 \leq x \leq 10\}$$

The process is updated according to (4.36) where the noise and reset variables are Gaussian with mean zero and variance 4 *(that is,* $\sigma = 2$*).*

We proceed by checking the conditional `pdf`s of interest, which can be seen in Fig. 4.7.

Next, solving the right hand side of (4.48) we find that ρ must be less than 98.1955 so that the associated function r is nondecreasing in $[0, w]$. The plots of r for different values of ρ can be seen in Fig. 4.8.

Finally, Fig. 4.9 shows the cost $\mathcal{G}_i$ for four values of ρ for which the function r satisfies the nondecreasing requirement. As expected, when $\rho = 3.9 \leq 3.9984 = \frac{q_1 \eta_1}{p_2 + q_1^2}$, the optimal reset parameter is $i^* = 1$ (by Corollary 4.4.3).

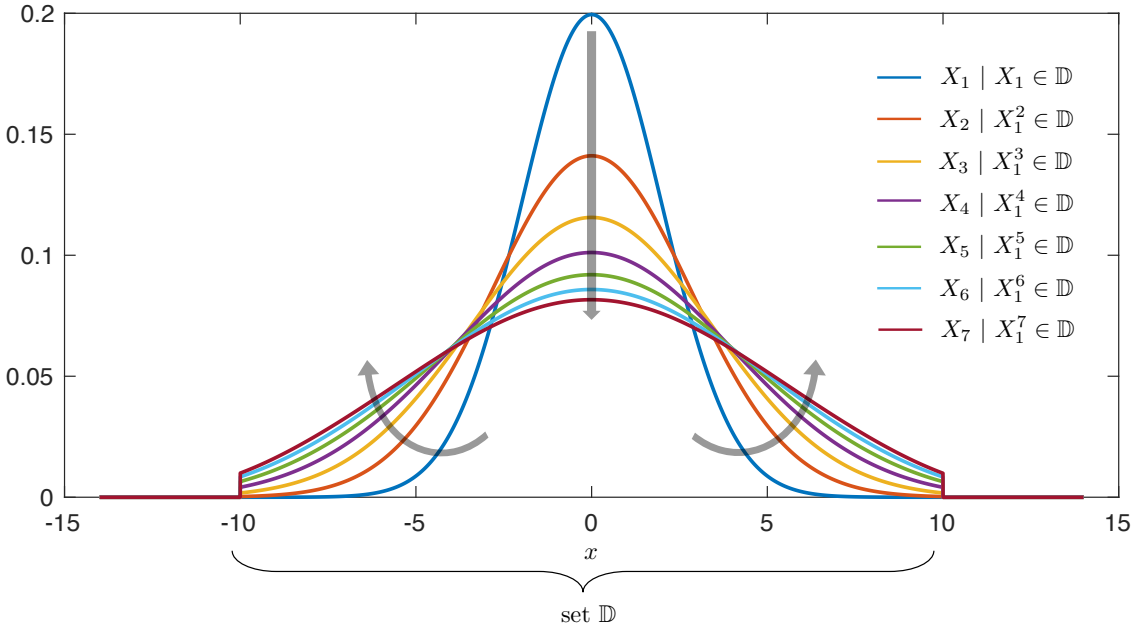


Figure 4.7: Evolution of conditioned `pdf`s for Example 4.6.2.

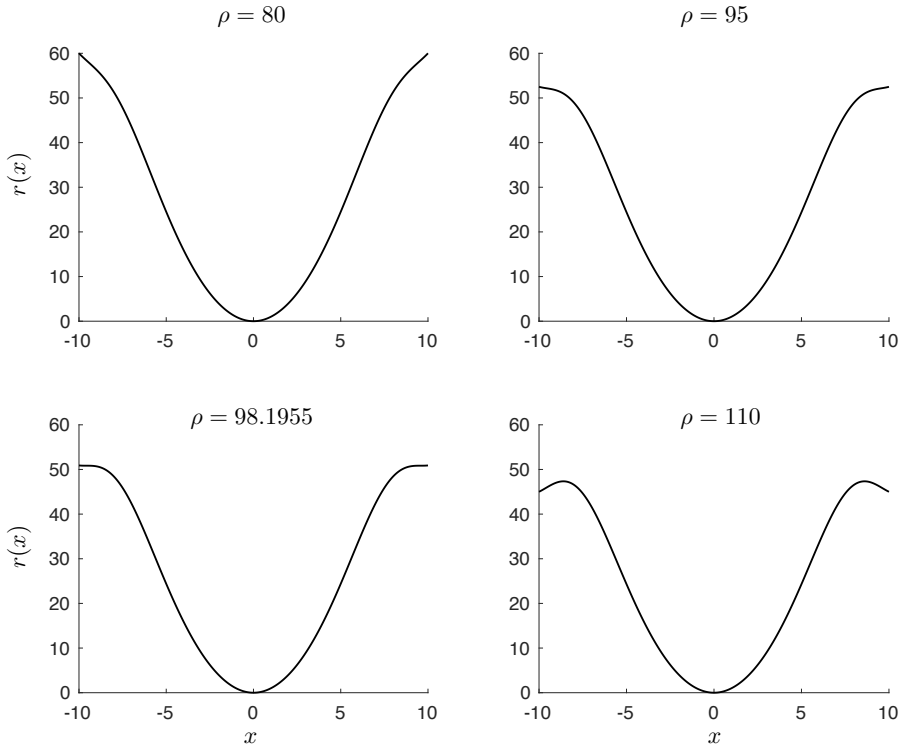


Figure 4.8: Plots of r in Example 4.6.2 for four different values of ρ: 80, 95, 98.1955, and 110. As expected when $\rho = 110$, the function does not satisfy the nondecreasing condition.

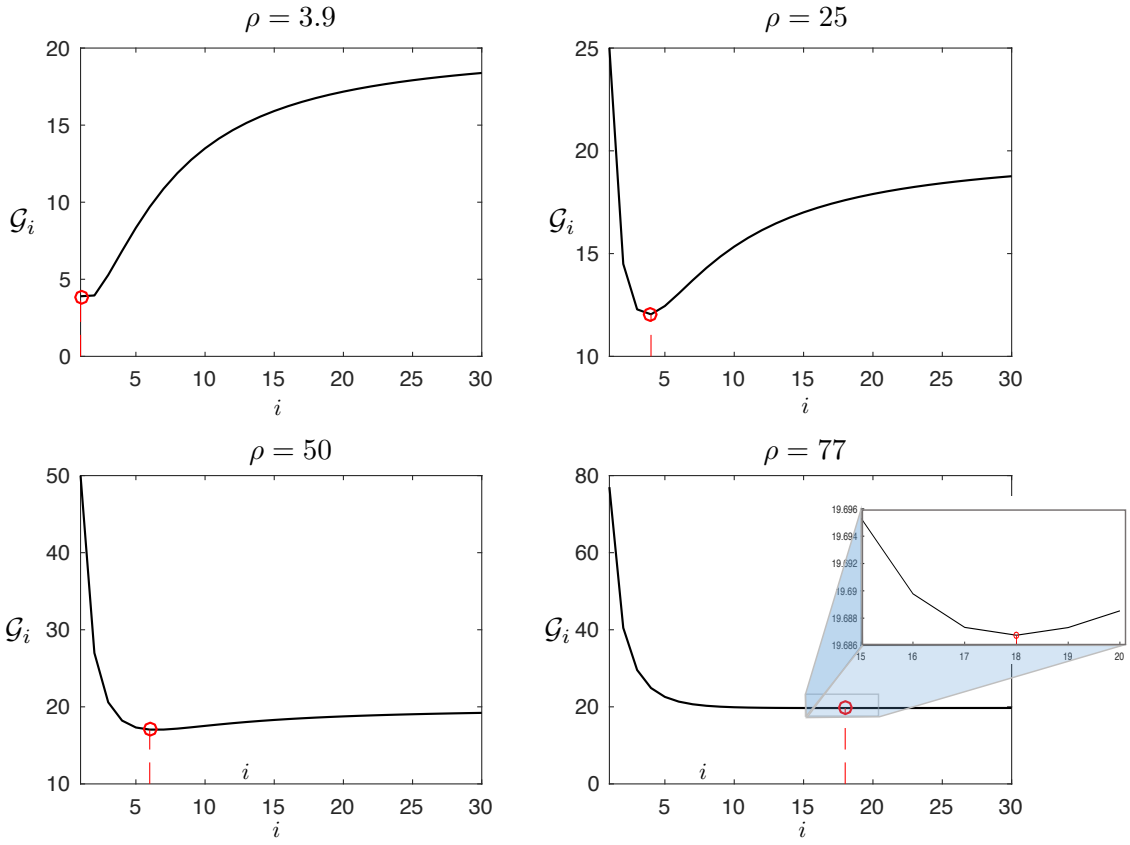


Figure 4.9: $\mathcal{G}_i$ in Example 4.6.2 for four different values of ρ: 3.9, 25, 50, and 77.

4.7 Concrete Cases: the Real Plane

To conclude this chapter, we show an example with a stochastic process with additive Gaussian noise in $\mathbb{R}^2$.

Example 4.7.1. *Consider the denied set given by:*

$$\mathbb{D} = \{(x, y) \in \mathbb{R}^2 \ : ||(x, y)|| \leq 10\} \tag{4.49}$$

The reset penalty is 50 *and state cost is given by* $\bar{g}(x, y) = x^2 + y^2$ *so that the overall stage cost is:*

$$g(x, y, u) = \begin{cases} 50 & u = 1 \\ 0 & x \notin \mathbb{D} \\ x^2 + y^2 & otherwise \end{cases} \tag{4.50}$$

The process is given by Z_k *in* $\mathbb{R}^2$ *and has the following update rule:*

$$Z_{k+1} = \begin{cases} Z_k + N_k, & U_k = 0 \text{ and } Z_k \in \mathbb{D} \\ R_k \ , & U_k = 1 \text{ or } Z_k \notin \mathbb{D}. \end{cases}$$

where R_k *and* N_k *are two independent sequences of independent and identically distributed Gaussian random vectors with mean zero and variance* Σ*, where:*

$$\Sigma = 4 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.51}$$

Fig. 4.10 shows the evolution of the `pdf` *of* $Z_k | Z_1^k \in \mathbb{D}$*. A cross-section of the* `pdf`*s can be seen in Fig. 4.11. As expected, the radial stochastic order is increasing.*

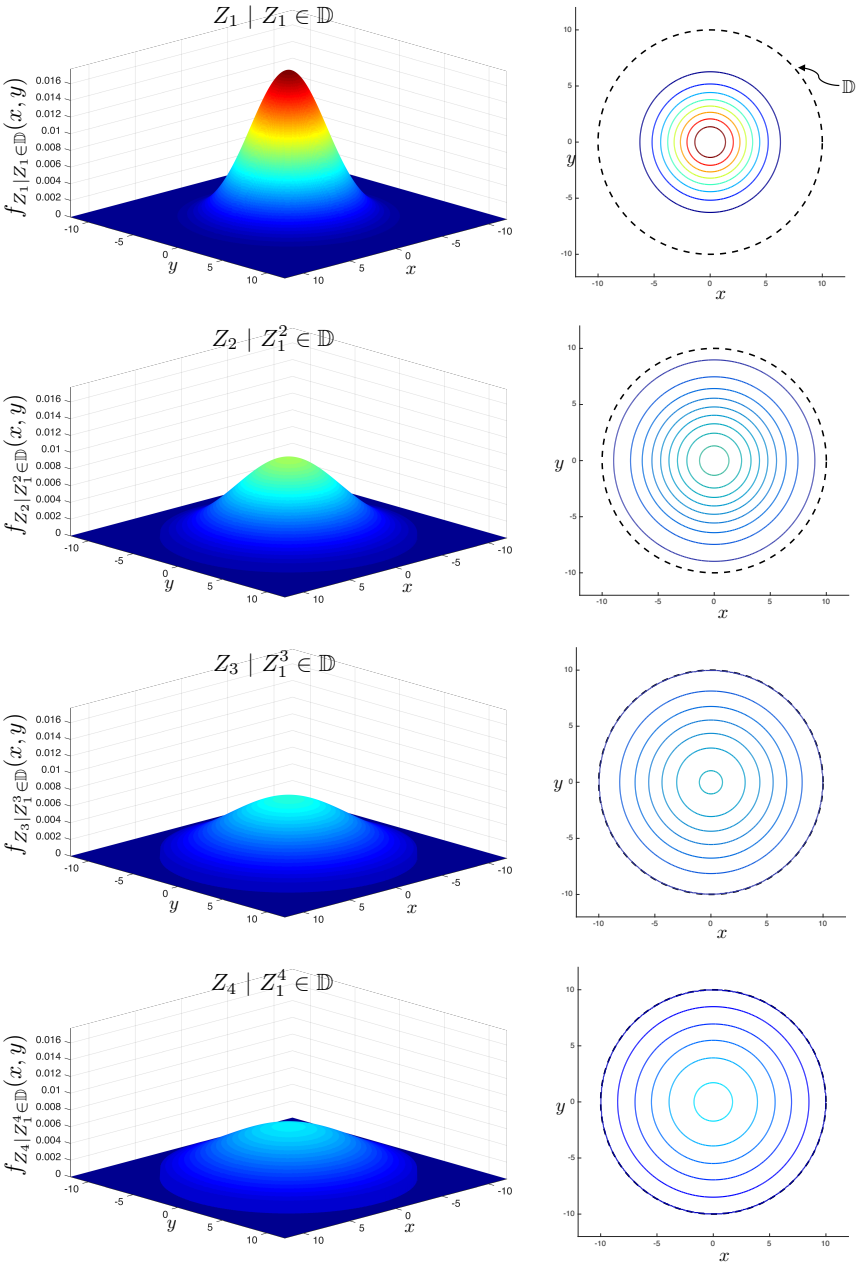


Figure 4.10: Evolution of the condtional `pdf`s of $Z_k|Z_1^k \in \mathbb{D}$. The right side shows the contour plots.

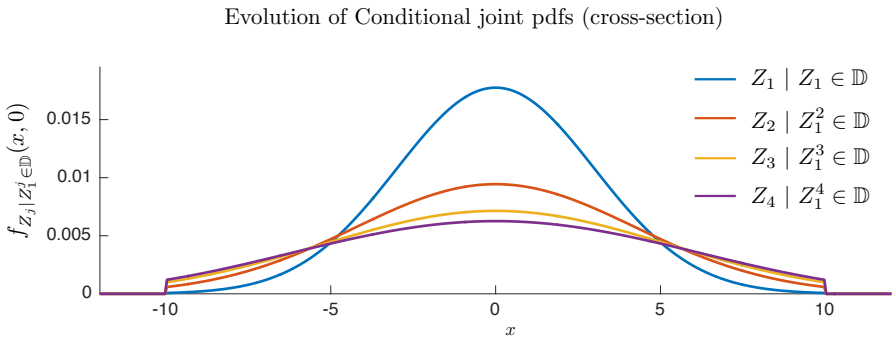


Figure 4.11: The cross-section of the evolution of the condtional `pdf`s of $Z_k|Z_1^k \in \mathbb{D}$

The cross-section of the function r for different values of ρ can be seen in Fig. 4.12. The figure shows that for $\rho = 50$, the function is radially increasing. The optimal reset parameter was found to be $i^* = 4$, as can be seen in the plot of $\mathcal{G}_i$ in Fig.4.13.

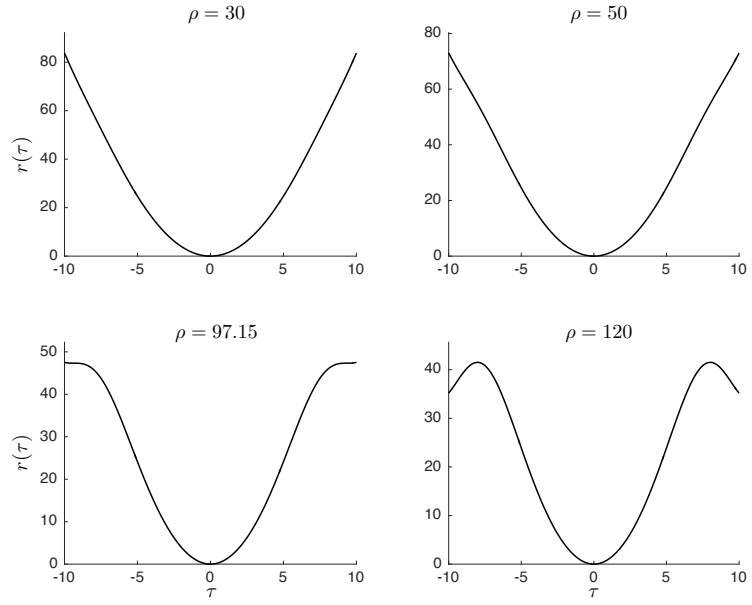


Figure 4.12: Cross-section of the function $r(x, y)$ for diferent values of ρ.

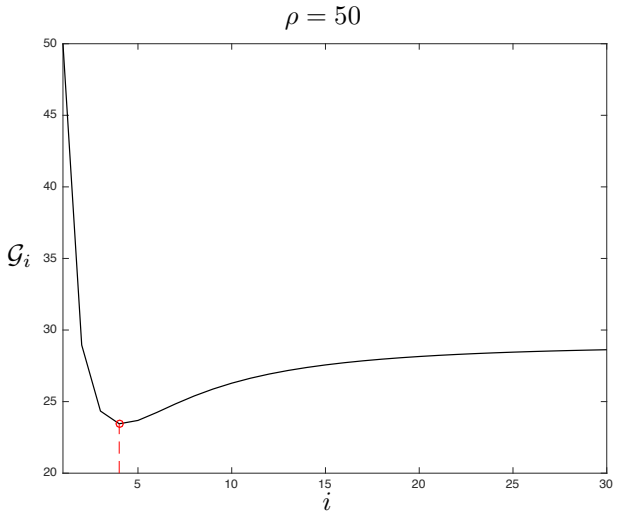
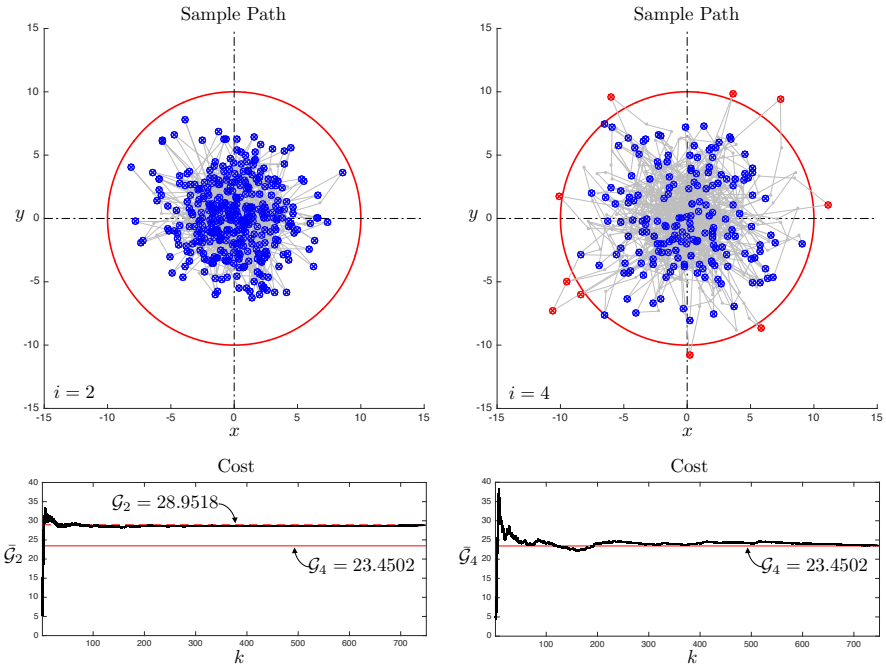


Figure 4.13: $\mathcal{G}_i$ for Example 4.7.1

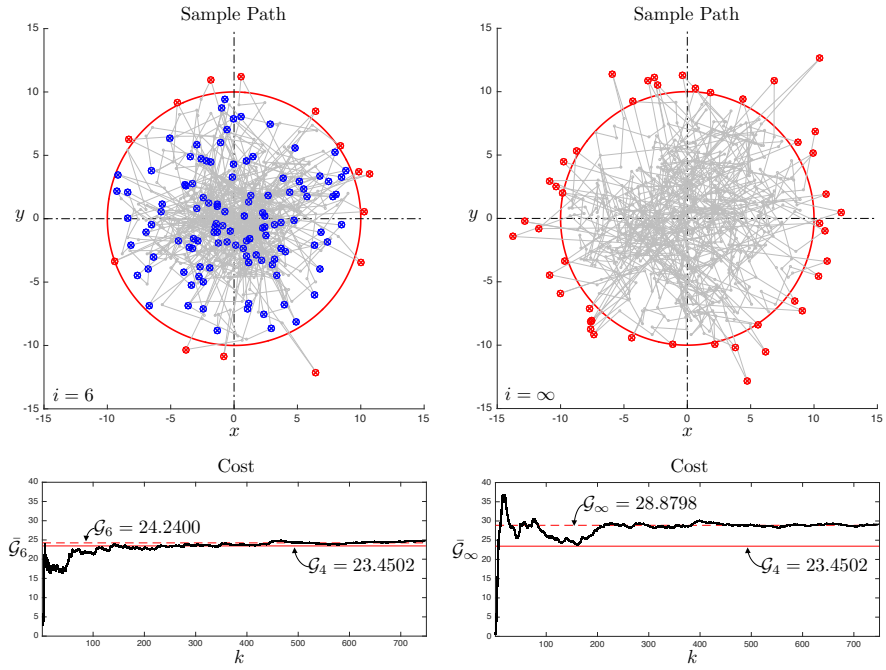
Four sample paths are shown in Fig. 4.14 for four different reset parameters: 2, 4, 6 *and "never" ($i = \infty$). Below each sample path, we plot the associated running cost given by:*

$$\bar{\mathcal{G}}_i(k) = \frac{1}{k} \sum_{j=1}^{k} g(x_j, y_j, u_j) \tag{4.52}$$



(a) $i = 2$

(b) $i = i^* = 4$



(c) $i = 6$

(d) $i = \infty$

Figure 4.14: Sample paths for Example 4.7.1. The red circles represent *passive* resets while the blue circles represent *controlled* resets.

4.8 Conclusion

In this chapter we proposed a problem of optimal sensor usage in denied environments. We used the framework of renewal reward processes to establish a recursive expression for the cost we wished to minimize, and established conditions in Theorem 4.4.1 under which any local minimum, if it exists, is also global. We showed that when an auxiliary process has increasing radial stochastic order, the conditions for which Theorem 4.4.1 holds simplifiy to checking if a certain function is radially increasing. We illustrated the results with concrete scalar and vector examples.

Chapter 5: Sensor Usage in Denied Enviroments: Markov Chains

In this chapter, we adapt the theory developed in Chapter 4 for the case when the state space is discrete and the state update is ruled by a Markov chain. Specifically, we focus on chains whose state space is indexed by integers, and at each time step the chain can stay at its current state, or move, with equal probability, to a neighboring state.

This approach overcomes a difficulty present in the previous chapter: checking whether the process $X_k|X_1^k \in \mathbb{D}$ ha increasing radial stochastic order. In Chapter 4, this task involved computing an integral for all radiuses inside the denied environment $\mathbb{D}$, whereas here, as we will show, the task simplifies to computing finitely many dot products. Moreover, by eigen-decomposing a submatrix of the probability transition matrix of the Markov chain, we aim at better understanding the evolution of the conditioned probability mass function (`pmf`) of $X_k|X_1^k \in \mathbb{D}$, which is essential to establishing the radial stochastic order of the process.

The main focus of this chapter is to establish conditions on the initial `pmf` of the Markov chain that guarantee that the radial stochastic order of $X_k|X_1^k \in \mathbb{D}$ is increasing for all time steps so that previous results, such as Theorem 4.4.1, can be applied. The results of Chapter 4, which can be easily adapted to Markov chains,

are reproduced in this chapter (with appropriate modifications) for completeness, but their proofs are omitted since they are nearly identical to those found in Chapter 4.

The chapter is organized as follows: Section 5.1 provides the problem formulation. In Section 5.1, we reformulate the problem as a renewal process and, in Section 5.2, we provide results from Chapter 4 adapted to Markov chains. The main results of this chapter can be found in Section 5.3, where we characterize initial `pmf`s that guarantee increasing radial stochastic order of the conditioned process. Finally, we conclude with final remarks in Section 5.4.

5.1 Problem Formulation

In this chapter, we adopt most of the notation used in Chapter 4. In addition, we reserve bold-face small-case letters, such as $\boldsymbol{v}$ for vectors, and bold-face upper-case letters, such as $\boldsymbol{P}$, for matrices.

Consider a Markov chain whose state at time k is given by X_k in $\mathbb{Z}$, and whose transition probabilities are given by:

$$\mathbb{P}\big(X_{k+1} = x \;\big|\; X_k = x\big) = \alpha \tag{5.1}$$

$$\mathbb{P}\big(X_{k+1} = x+1 \;\big|\; X_k = x\big) = \beta \tag{5.2}$$

$$\mathbb{P}\big(X_{k+1} = x-1 \;\big|\; X_k = x\big) = \beta \tag{5.3}$$

for some α and β in $[0, 1]$ such that $\alpha \geq 2\beta$ and $\alpha + 2\beta = 1$.

The denied set $\mathbb{D}$ is a subset of $\mathbb{Z}$ that is symmetric around 0 and is given by:

$$\mathbb{D} = \{-n, ..., 0, ...n\} \tag{5.4}$$

for some positive integer n. The transition diagram of the Markov chain and the denied set $\mathbb{D}$ are depicted in Fig. 5.1

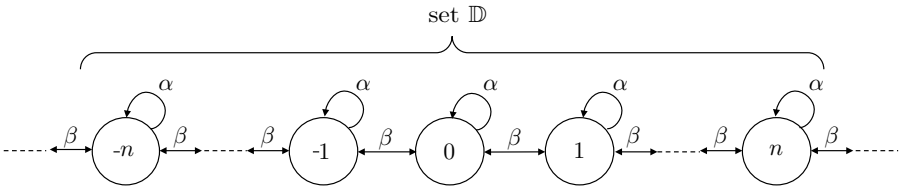


Figure 5.1: Diagram of Markov chain

The chain is initialized to some *intial* `pmf` $\tilde{f}_1$ that is symmetric around 0, such that

$$\mathbb{P}(X_1 = x) = \tilde{f}_1(x) \tag{5.5}$$

for all x in $\mathbb{Z}$.

Once the chain in initialized, it evolves according to the transition probabilities described in (5.1) - (5.3), until it undergoes a *reset*. When a reset happens, the chain is *reinitialized* to the same initial `pmf` $\tilde{f}_1$. There are two distinct situations when a reset takes place: *i)* when the reset variable U_k is set to one (*active* or *controlled* reset); and *ii)* when the chain leaves the denied set $\mathbb{D}$ (*passive* reset).

One can think of an active reset as an agent's decision to pay a price ρ to use an expensive sensor and observe its location, enabling the agent to return to its initial position (represented by the initial `pmf` $\tilde{f}_1$). Similarly, a passive reset happens when

the agent wanders outside the denied set $\mathbb{D}$, where state observations are readily available (and free). The agent can therefore return to its original location without paying for sensor usage.

The reset variable U_k cannot depend on state of the chain X_k since direct observations of the process are not available in $\mathbb{D}$. It can, however, use knowledge of the last time the chain exited the denied set. Therefore, the reset variable is determined according to a policy $\mathcal{U}$ of the following type:

$$U_k = \mathcal{U}\big(\{\mathcal{I}_j(X_j \notin \mathbb{D})\}_{j=0}^{k}\big) \tag{5.6}$$

where $\mathcal{U}$ maps the history up to time k of the indicator function $\mathcal{I}_j$ of the event $X_j \notin \mathbb{D}$ to a binary value 0 (wait) or 1 (reset).

For a given reset price ρ in $\mathbb{R}_+$, consider the following cost:

$$\mathcal{G}_{\mathcal{U}} \stackrel{def}{=} \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\sum_{k=1}^{N} g(X_k, U_k)\right], \tag{5.7}$$

where the reset variable U_k is determined according to the policy $\mathcal{U}$, and the state cost g is some non-negative real-valued function that has the following structure:

$$g(x,u) = \begin{cases} 0 & x \notin \mathbb{D} \\ \rho & u = 1 \\ \bar{g}(x) & \text{otherwise} \end{cases} \tag{5.8}$$

for some non-negative real-valued function $\bar{g} : \mathbb{Z} \to \mathbb{R}_+$.

A stage cost of this type means that when a passive reset occurs ($x \notin \mathbb{D}$), no cost is incurred, whereas when a controlled reset takes place ($u = 1$), a cost of ρ is

incurred. Finally when the stage is reset-free, a cost of $\bar{g}$ is incurred which depends only on the current state of the Markov chain.

We are ready to state the main problem:

Problem 5.1.1. *Given the Markov chain described in (5.1) and real-valued cost described in (5.7), a positive reset penalty ρ and set $\mathbb{D}$ as in (5.4), find an optimal reset policy $\mathcal{U}^*$ as in (5.6) such that:*

$$\mathcal{U}^* = \arg\min_{\mathcal{U}} \mathcal{G}_{\mathcal{U}} \tag{5.9}$$

5.1.1 The Renewal Process Formulation

As we did in Chapter 4, the first step in solving Problem 5.1.1 is to recognize that the underlining stochastic process is in fact a *renewal process*, since it renews when $U_k = 1$ or when X_k exits the set $\mathbb{D}$. With this approach, instead of using the cost in (5.7), we consider an equivalent cost structure that uses the concept of *cycles*. The idea is simple: we rewrite the cost using the expectation of *a* cycle whose length is a random variable T, which is known as a *stopping time*. This means we can focus on only one cycle (of random length T) rather than on whole infinite horizon. This is a known result in renewal process theory and is given below:

Lemma 5.1.2. *The cost in (5.7) is equivalent to the following cost:*

$$\mathcal{G}_{\mathcal{U}} = \frac{\mathbb{E}\left[\sum_{k=1}^{T} g(X_k, U_k)\right]}{\mathbb{E}[T]} \tag{5.10}$$

where the random variable T is the length of the cycle.

Proof. See [70], Theorem 3.6.1 □

From now on and with some abuse of notation, we will use the time index k to refer to time steps inside of a cycle rather than those in the complete infinite time horizon. Therefore whenever the cycle reaches time step $k = 4$, we can assume that a reset has *not* happened in $k = 0, 1, 2$ or 3 (otherwise the cycle would have ended and a new one begun).

Furthermore, we can now consider an equivalent Markov chain with a finite state space, whose transition probabilities are given in Fig. 5.2 and below:

$$\mathbb{P}\big(X_{k+1} = x \;\big|\; X_k = x\big) = \alpha, \qquad x \in \mathbb{D} \tag{5.11}$$

$$\mathbb{P}\big(X_{k+1} = x + 1 \;\big|\; X_k = x\big) = \beta, \qquad x \in \mathbb{D} \tag{5.12}$$

$$\mathbb{P}\big(X_{k+1} = x - 1 \;\big|\; X_k = x\big) = \beta, \qquad x \in \mathbb{D} \tag{5.13}$$

$$\mathbb{P}\big(X_{k+1} = x \;\big|\; X_k = x\big) = 1, \qquad x \in \{-n-1,\ n+1\} \tag{5.14}$$

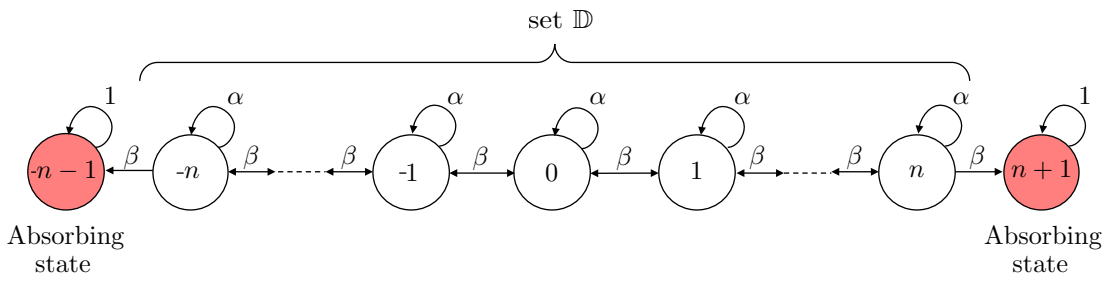


Figure 5.2: Diagram of Markov chain with absorbing states.

The transition probabilities of the modified chain is given by the $(2n + 3) \times$

$(2n+3)$ matrix $\bar{\boldsymbol{P}}$ given below:

$$\bar{\boldsymbol{P}} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \beta & \alpha & \beta & 0 & \cdots & \cdots & 0 \\ 0 & \beta & \alpha & \beta & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & \beta & \alpha & \beta & 0 \\ 0 & \cdots & \cdots & 0 & \beta & \alpha & \beta \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \tag{5.15}$$

Finally, the equivalent chain is initialized according to the `pmf` f_1 given by:

$$f_1(x) = \begin{cases} \tilde{f}_1(x) & x \in \mathbb{D} \\ \frac{1}{2}\sum_{\tau \notin \mathbb{D}} \tilde{f}_1(\tau) & x \in \{-n-1,\ n+1\} \end{cases} \tag{5.16}$$

The probability distribution of the cycle length T will depend on the `pmf` f_1, the probabilities α and β, *and* on the reset policy $\mathcal{U}$, which is our design parameter. It should now be clear that the reset policy $\mathcal{U}$ can only depend on how many time steps have elapsed without a reset occurring. Moreover, any reset policy of interest can be parameterized by a single positive integer i, namely the number of time steps to wait without a passive reset having occurred before triggering a controlled reset. In other words, the reset policy parameterized by i at (cycle) time k is given by:

$$\mathcal{U}_i\big(\{\mathcal{I}_j(X_j \notin \mathbb{D})\}_{j=0}^{k}\big) = \begin{cases} 1, & k = i \text{ and } \sum_{j=0}^{k} \mathcal{I}_j(X_j \notin \mathbb{D}) = 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.17}$$

In light of this discussion, we can reparameterize the cost in terms of i, as follows:

$$\mathcal{G}_i = \frac{\mathbb{E}\left[\sum_{k=1}^{T_i} g(X_k, U_k)\right]}{\mathbb{E}[T_i]} \tag{5.18}$$

where U_k is determined according to the policy $\mathcal{U}_i$ in (5.17), and T_i is the appropriate stopping time for X_k under policy $\mathcal{U}_i$.

We are now ready to state an equivalent problem to Problem 4.1.1:

Problem 5.1.3. *Given the Markov chain described in (5.1) and real-valued cost described in (5.18), a positive reset penalty ρ and set $\mathbb{D}$ as in (5.4), find the optimal reset parameter i^* such that:*

$$i^* = \arg\min_i \mathcal{G}_i \tag{5.19}$$

5.2 Finding the Optimal Reset Parameter

The results in this section rely solely on expectations and not on the nature of the underlining stochastic process. Therefore these results are direct adaptation of those in Chapter 4. They have been reproduced here (with appropriate modifications where needed) for completeness, but the proofs have been omitted.

Theorem 5.2.1. *Consider Problem 5.1.3 with stage cost $\bar{g}$ and reset penalty ρ. Define the function $r : \mathbb{Z} \to \mathbb{R}$ where:*

$$r(x) \stackrel{def}{=} \bar{g}(x) - \rho\beta\mathcal{I}(x \in \{-n, n\}) \tag{5.20}$$

where $\mathcal{I}$ is the indicator function. Further suppose that:

$$\mathbb{E}\big[r(X_i) \mid X_1^i \in \mathbb{D}\big] \quad \leq \quad \mathbb{E}\big[r(X_{i+1}) \mid X_1^{i+1} \in \mathbb{D}\big] \tag{5.21}$$

for all i.

The following holds: if there exists a parameter i such that $\mathcal{G}_i \geq \mathcal{G}_{i-1}$, then $\mathcal{G}_j \geq \mathcal{G}_{j-1}$ for all $j > i$

Proof. See proof of Theorem 4.4.1. □

Theorem 5.2.1 states that, under certain conditions, once the function $\mathcal{G}_i$ starts increasing it always increases from then on. A direct result of Theorem 4.4.1 related to the optimal reset parameter is the following:

Corollary 5.2.2. *Under the same conditions of Theorem 5.2.1, if a local minimum exists it is global.*

5.2.1 Radial Stochastic Order for Markov Chains

Checking if a finitely valued discrete random process has increasing radial stochastic order consists on computing finitely many dot products for each time step k and comparing their values.

Let $\boldsymbol{\eta}_0, ..., \boldsymbol{\eta}_{n-1}$ be $2n+1$-dimensional column vectors such that the l^{th} entry of $\boldsymbol{\eta}_r$ is given by:

$$\boldsymbol{\eta}_r[l] = \begin{cases} 1 & -r \leq l \leq r \\ 0 & \text{otherwise} \end{cases} \tag{5.22}$$

Definiton 5.2.3. *A Markov chain Y_k is said to be increasing in the radial stochastic order, written $Y_k \leq_r Y_{k+1}$, if*

$$\boldsymbol{\eta}_r^T \boldsymbol{p}_k \geq \boldsymbol{\eta}_r^T \boldsymbol{p}_{k+1} \tag{5.23}$$

for all $k \geq 1$ and all r in $\{0, ...n\}$; where $\boldsymbol{p}_k$ is the `pmf` *of Y_k.*

Definiton 5.2.4. *A function $l : \mathbb{Z} \to \mathbb{R}$ is radially nondecreasing function when, for all x and y with $|x| \leq |y|$, the following holds:*

$$l(x) \leq l(y) \tag{5.24}$$

Lemma 5.2.5. *Suppose Y_k is a Markov chain in $\mathbb{Z}$ with increasing stochastic order, and let $l : \mathbb{Z} \to \mathbb{R}$ be a radially nondecreasing function. The following holds:*

$$\mathbb{E}[l(Y_{k+1})] \geq \mathbb{E}[l(Y_k)] \tag{5.25}$$

for all i.

Corollary 5.2.6. *Condition (5.21) in Theorem 5.2.1 is satisfied for all i if the following holds:*

- *The conditioned process X_k given that $X_1^k \in \mathbb{D}$ has increasing stochastic order.*
- *The function $r : \mathbb{R}^n \to \mathbb{R}$ where:*

$$r(x) \stackrel{def}{=} \bar{g}(x) - \rho\beta\mathcal{I}(x \in \{-n, n\}) \tag{5.26}$$

  is radially nondecreasing.

Proof. Follows from Lemma 5.2.5. □

5.3 Characterizing initial pmfs

In this section, the goal is to characterize initial pmfs for which the resultant Markov chain has increasing stochastic order. This is motivated by Corollary 5.2.6, which guarantees application of Theorem 5.2.1 and its derivations. Since the application of these results has been extensively covered in Chapter 4, we omit this treatment here and focus solely on the task of identifying conditions on the initial pmf that guarantees that the radial stochastic order is increasing.

Before proceeding with the main results, we need to introduce a vector notation for the pmf of the state of the Markov chain at any given time k given that the chain has not exited the denied set $\mathbb{D}$. The aim is to establish conditions so that the radial stochastic order of these pmfs is increasing.

Let $\boldsymbol{v}_1$ a $(2n+1)$-dimensional column vector that represents the pmf of X_1 given that $X_1 \in \mathbb{D}$, where entries of $\boldsymbol{v}_1$ are indexed from $-n$ to n. The vector $\boldsymbol{v}_1$ is given by:

$$\boldsymbol{v}_1 \stackrel{def}{=} \frac{1}{\sum_{x\in\mathbb{D}} f_1(x)} \begin{bmatrix} f_1(-n) & \cdots & f_1(0) & \cdots & f_1(n) \end{bmatrix}^T. \tag{5.27}$$

For simplicity, without loss of generality, we consider from now on only initial pmfs that have support in $\mathbb{D}$, and therefore:

$$\boldsymbol{v}_1 = \begin{bmatrix} f_1(-n) & \cdots & f_1(0) & \cdots & f_1(n) \end{bmatrix}^T. \tag{5.28}$$

For $k \geq 2$, let $\boldsymbol{v}_k$ be a column vector that represents the pmf of X_k given that

$X_1^k \in \mathbb{D}$. The evolution of the conditional pmfs $\boldsymbol{v}_k$ are given by:

$$\boldsymbol{v}_{k+1} = \gamma_k^{-1} \boldsymbol{P} \boldsymbol{v}_k \tag{5.29}$$

$$= \gamma_k^{-1} \boldsymbol{P}^k \boldsymbol{v}_1 \tag{5.30}$$

where $\boldsymbol{P}$ is a $(2n+1) \times (2n+1)$ matrix given by:

$$\boldsymbol{P} = \begin{bmatrix} \alpha & \beta & 0 & \cdots & \cdots & \cdots & 0 \\ \beta & \alpha & \beta & 0 & \cdots & \cdots & 0 \\ 0 & \beta & \alpha & \beta & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & \beta & \alpha & \beta & 0 \\ 0 & \cdots & \cdots & 0 & \beta & \alpha & \beta \\ 0 & \cdots & \cdots & \cdots & 0 & \beta & \alpha \end{bmatrix}, \tag{5.31}$$

and γ_k is normalization factor given defined by:

$$\gamma_k \stackrel{def}{=} \mathbb{1}^T \boldsymbol{P}^k \boldsymbol{v}_1 \tag{5.32}$$

where $\mathbb{1}$ is the unit column vector of appropriate dimension.

Remark Note that the matrix $\boldsymbol{P}$ represents the non-absorbing states of the Markov chain $\bar{\boldsymbol{P}}$ and as such is *not* stochastic. Therefore the normalization terms γ are necessarily strictly less than 1.

5.3.1 Decomposition of $\boldsymbol{P}$

We proceed by establishing an eigen-decomposition of the matrix $\boldsymbol{P}$, which will serve as the foundation of the main results of this chapter. Because of the

special nature of the matrix $\boldsymbol{P}$, its eigenvectors are independent of the parameters α and β, which will enable us to establish a few general results for the whole class of matrices of the type considered here.

Lemma 5.3.1. *For any matrix $\boldsymbol{P}$ as in (5.31) with parameters α and β, its eigenvalues are given by:*

$$\lambda_j = \alpha + 2\beta \cos\left(\frac{j\pi}{2n+2}\right) \tag{5.33}$$

for $j = 1, ..., 2n+1$.

Moreover, for each eigenvalue λ_j we can write an associated eigenvector as follows:

$$\boldsymbol{u}_j = \frac{1}{\sqrt{n+1}} \begin{bmatrix} \sin\left(\frac{j\pi}{2n+2}\right) \\ \sin\left(\frac{2j\pi}{2n+2}\right) \\ \vdots \\ \sin\left(\frac{(2n+1)j\pi}{2n+2}\right) \end{bmatrix}. \tag{5.34}$$

Finally, the eigenvectors $\boldsymbol{u}_j$'s are orthonormal.

Proof. It suffices to show that $\boldsymbol{P}\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j$ for all j, which is straightforward using the fact that $2\sin\theta\cos\phi = \sin(\theta-\phi) + \sin(\theta+\phi)$. For a direct construction of the eigenvalues and eigenvectors from the matrix $\boldsymbol{P}$, see [72].

The orthogonality of the eigenvectors follows from the symmetry of $\boldsymbol{P}$. To

check that $\|\boldsymbol{u}_j\| = 1$, note that:

$$\begin{aligned}
\|\boldsymbol{u}_j\|^2 &= \frac{1}{n+1}\sum_{l=1}^{2n+1}\sin^2\left(\frac{lj\pi}{2n+2}\right) \\
&= \frac{1}{n+1}\sum_{l=1}^{2n+1}\frac{1}{2}-\frac{1}{2}\cos\left(\frac{2lj\pi}{2n+2}\right) \\
&= \frac{2n+1}{2n+2}-\frac{1}{2n+2}\sum_{l=1}^{2n+1}\cos\left(\frac{2lj\pi}{2n+2}\right) \\
&= \frac{2n+1}{2n+2}-\frac{1}{2n+2}\left(\frac{\sin\left(\frac{4n+3}{2n+2}j\pi\right)}{2\sin\left(\frac{j\pi}{2n+2}\right)}-\frac{1}{2}\right) \qquad (5.35) \\
&= \frac{2n+1}{2n+2}-\frac{1}{2n+2}\left(-\frac{1}{2}-\frac{1}{2}\right) \qquad (5.36) \\
&= \frac{2n+1}{2n+2}+\frac{1}{2n+2} \\
&= 1
\end{aligned}$$

where the line (5.35) is due to Lagrange's trigonometric identity, and (5.36) is reached by noting that

$$\begin{aligned}
\sin\left(\frac{4n+3}{2n+2}j\pi\right) &= \sin\left(2j\pi-\frac{1}{2n+2}j\pi\right) \\
&= -\sin\left(\frac{j\pi}{2n+2}\right)
\end{aligned}$$

□

Remark Note that the eigenvectors $\boldsymbol{u}_j$ do not depend on the parameters α and β. This means that all matrices of this type have the same eigenvectors.

Finally, the matrix $\boldsymbol{P}$ can be written according to its eigenvectors and eigenvalues, as follows:

$$\boldsymbol{P} = \sum_{j=1}^{2n+1}\lambda_j\boldsymbol{u}_j\boldsymbol{u}_j^T \qquad (5.37)$$

Due to the orthonormality of the eigenvectors $\boldsymbol{u}_j$'s, the $k-th$ power of $\boldsymbol{P}$ is given by:

$$\boldsymbol{P}^k = \sum_{j=1}^{2n+1} \lambda_j^k \boldsymbol{u}_j \boldsymbol{u}_j^T, \tag{5.38}$$

and the conditional `pmf` $\boldsymbol{v}_{k+1}$ by:

$$\boldsymbol{v}_{k+1} = \gamma_k^{-1} \sum_{j=1}^{2n+1} \lambda_j^k \boldsymbol{u}_j \boldsymbol{u}_j^T \boldsymbol{v}_1 \tag{5.39}$$

5.3.2 Initial `pmf`s for Increasing Radial Stochastic Order

Any initial `pmf` $\boldsymbol{v}_1$ can be written as a linear combination of the eigenvectors $\boldsymbol{u}_j$'s since they form a basis in $\mathbb{R}^{2n+1}$. Therefore, $\boldsymbol{v}_1$ can be represented by the weights b_j's of the linear combination:

$$\boldsymbol{v}_1 = \sum_{j=1}^{2n+1} b_j \boldsymbol{u}_j \tag{5.40}$$

And therefore, the conditional `pmf` 's $\boldsymbol{v}_{k+1}$ can be written as:

$$\boldsymbol{v}_{k+1} = \gamma_k^{-1} \sum_{j=1}^{2n+1} \lambda_j^k b_j \boldsymbol{u}_j, \tag{5.41}$$

Recall from the definition of radial stochastic order for Markov chains that $\boldsymbol{\eta}_0, ..., \boldsymbol{\eta}_{n-1}$ are column vectors in $\mathbb{R}^{2n+1}$ such the l^{th} entry is given by:

$$\boldsymbol{\eta}_r[l] = \begin{cases} 1 & -r \leq l \leq r \\ \\ 0 & \text{otherwise} \end{cases} \tag{5.42}$$

and consider the following matrices $\boldsymbol{\Gamma}_r(t)$ indexed by r in $\{0, ..., n-1\}$ and t in $\mathbb{R}+$:

$$\boldsymbol{\Gamma}_r(t)[i,j] \stackrel{def}{=} \frac{1}{2}(\lambda_i \lambda_j)^t \ln(\lambda_i/\lambda_j)\big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big) \tag{5.43}$$

where $d_j \stackrel{def}{=} \mathbb{1}^T \boldsymbol{u}_j$.

The following theorem provides a condition on the weights b_j's that guarantee that the Markov chain has increasing radial stochastic order. Although the condition is hard to check, it functions as the basis of the results that follow, which are more tractable.

Theorem 5.3.2. *Suppose $\boldsymbol{b}$, a column vector whose entries are the weights b_j's in (5.40), is such that*

$$\boldsymbol{b}^T \boldsymbol{\Gamma}_r(t) \boldsymbol{b} \leq 0 \tag{5.44}$$

for all r in $\{0, ..., n-1\}$ and t in $\mathbb{R}_+$. The following holds:

$$\{X_k \mid X_1^k \in \mathbb{D}\} \ \leq_r \ \{X_{k+1} \mid X_1^{k+1} \in \mathbb{D}\} \tag{5.45}$$

for all k.

Proof. Suppose by contradiction that there exists an r and a k such that:

$$\boldsymbol{\eta}_r^T \boldsymbol{v}_{k+1} > \boldsymbol{\eta}_r^T \boldsymbol{v}_k \tag{5.46}$$

(which would imply that (5.45) does not hold).

Recall that the conditional `pmf` 's $\boldsymbol{v}_k$ can be written as:

$$\boldsymbol{v}_{k+1} = \gamma_k^{-1} \sum_{j=1}^{2n+1} \lambda_j^k b_j \boldsymbol{u}_j, \tag{5.47}$$

and therefore we have that:

$$\gamma_k^{-1} \sum_{j=1}^{2n+1} \lambda_j^k b_j \boldsymbol{\eta}_r^T \boldsymbol{u}_j - \gamma_{k-1}^{-1} \sum_{j=1}^{2n+1} \lambda_j^{k-1} b_j \boldsymbol{\eta}_r^T \boldsymbol{u}_j > 0. \tag{5.48}$$

Let the function $h_r : \mathbb{R}_+ \to [0,1]$ be given by:

$$h_r(t) = \frac{1}{\hat{\gamma}(t)} \sum_{j=1}^{2n+1} \lambda_j^t b_j \boldsymbol{\eta}_r^T \boldsymbol{u}_j \tag{5.49}$$

where $\hat{\gamma} : \mathbb{R}_+ \to [0,1]$ is given by:

$$\hat{\gamma}(t) \stackrel{def}{=} \sum_{j=1}^{2n+1} \lambda_j^t b_j d_j \tag{5.50}$$

where $d_j \stackrel{def}{=} \mathbb{1}^T \boldsymbol{u}_j$.

Therefore, the assumption which we wish to contradict becomes:

$$h_r(k) - h_r(k-1) > 0. \tag{5.51}$$

and therefore:

$$\int_{k-1}^{k} h'(t)dt > 0 \tag{5.52}$$

where the derivative is given by:

$$h_r'(t) = \frac{1}{\hat{\gamma}^2(t)} \sum_{j=1}^{2n+1} \sum_{i<j} b_i b_j (\lambda_i \lambda_j)^t \ln(\lambda_i/\lambda_j) \big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j \big) \tag{5.53}$$

$$= \frac{1}{\hat{\gamma}^2(t)} \boldsymbol{b}^T \boldsymbol{\Gamma}_r(t) \boldsymbol{b} \tag{5.54}$$

$$\leq 0 \quad \text{for all } t, \text{ by the condition in (5.44)}, \tag{5.55}$$

which is a contradiction. □

The constraint in (5.44) is, in general, not easy to check since it has to hold for all r and *all* t. Below we provide a sufficient condition that does not depend on t and guarantees that (5.44) holds for all t.

Corollary 5.3.3. *Suppose* $\boldsymbol{b}$*, a column vector whose entries are the weights* b_j*'s in (5.40), is such that*

$$\mathbb{1}^T \Psi_r(\boldsymbol{b}) \mathbb{1} \leq 0 \tag{5.56}$$

for all r *in* $\{0, ..., n-1\}$*, and where:*

$$\Psi_r(\boldsymbol{b})[i,j] \stackrel{def}{=} \begin{cases} \frac{1}{2} b_i b_j \ln(\lambda_i/\lambda_j)\big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big) & (i,j) \in \{1,3\} \\ \left| \frac{\lambda_i \lambda_j}{2\lambda_1 \lambda_3} b_i b_j \ln(\lambda_i/\lambda_j)\big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big) \right| & \textit{otherwise} \end{cases}. \tag{5.57}$$

The following holds:

$$\{X_k \mid X_1^k \in \mathbb{D}\} \ \leq_r \ \{X_{k+1} \mid X_1^{k+1} \in \mathbb{D}\} \tag{5.58}$$

Proof. The proof relies on computing a bound on $\boldsymbol{b}^T \boldsymbol{\Gamma}_r(t) \boldsymbol{b}$ and showing that if the condition of the corollary is satisfied, then (5.44) of Theorem 5.3.2 is also satisfied.

$$\boldsymbol{b}^T \boldsymbol{\Gamma}_r(t) \boldsymbol{b} = \sum_{j=3}^{2n+1} \sum_{i<j} b_i b_j (\lambda_i \lambda_j)^t \ln(\lambda_i/\lambda_j)\big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big) \tag{5.59}$$

$$= b_1 b_3 (\lambda_1 \lambda_3)^t \ln(\lambda_1/\lambda_3)\big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) + \sum_{j=5}^{2n+1} \sum_{i<j} b_i b_j (\lambda_i \lambda_j)^t \ln(\lambda_i/\lambda_j)\big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big) \tag{5.60}$$

Note that the indices in the summation are odd because the even-indexed terms of

the matrix are zero.

$$
\begin{aligned}
\boldsymbol{b}^T\boldsymbol{\Gamma}_r(t)\boldsymbol{b} &= (\lambda_1\lambda_3)^t\Big\{b_1b_3\ln(\lambda_1/\lambda_3)\big(d_3\boldsymbol{\eta}_r^T\boldsymbol{u}_1 - d_1\boldsymbol{\eta}_r^T\boldsymbol{u}_3\big)+ \\
&\qquad \sum_{j=5}^{2n+1}\sum_{i<j} b_ib_j\Big(\frac{\lambda_i\lambda_j}{\lambda_1\lambda_3}\Big)^t\ln(\lambda_i/\lambda_j)\big(d_j\boldsymbol{\eta}_r^T\boldsymbol{u}_i - d_i\boldsymbol{\eta}_r^T\boldsymbol{u}_j\big)\Big\} \\
&\le (\lambda_1\lambda_3)^t\Big\{b_1b_3\ln(\lambda_1/\lambda_3)\big(d_3\boldsymbol{\eta}_r^T\boldsymbol{u}_1 - d_1\boldsymbol{\eta}_r^T\boldsymbol{u}_3\big)+ \\
&\qquad \sum_{j=5}^{2n+1}\sum_{i<j} \Big|b_ib_j\Big(\frac{\lambda_i\lambda_j}{\lambda_1\lambda_3}\Big)^t\ln(\lambda_i/\lambda_j)\big(d_j\boldsymbol{\eta}_r^T\boldsymbol{u}_i - d_i\boldsymbol{\eta}_r^T\boldsymbol{u}_j\big)\Big|\Big\} \\
&\le (\lambda_1\lambda_3)^t\Big\{b_1b_3\ln(\lambda_1/\lambda_3)\big(d_3\boldsymbol{\eta}_r^T\boldsymbol{u}_1 - d_1\boldsymbol{\eta}_r^T\boldsymbol{u}_3\big)+ \\
&\qquad \sum_{j=5}^{2n+1}\sum_{i<j} \Big|b_ib_j\frac{\lambda_i\lambda_j}{\lambda_1\lambda_3}\ln(\lambda_i/\lambda_j)\big(d_j\boldsymbol{\eta}_r^T\boldsymbol{u}_i - d_i\boldsymbol{\eta}_r^T\boldsymbol{u}_j\big)\Big|\Big\} \\
&= (\lambda_1\lambda_3)^t\mathbb{1}^T\Psi_r(\boldsymbol{b})\mathbb{1} \\
&\le 0 \qquad \text{for all } t.
\end{aligned}
$$

And so, by Theorem 5.3.2, the stochastic order is increasing. □

Although Corollary 5.3.3 is easy to check, it is somewhat conservative since it relies on the triangle inequality. This conservativeness worsens as the dimension of the problem grows. Nevertheless, we provide the next example to demonstrate a case when it applies.

Example 5.3.4. *Let $n = 5$, $\alpha = 0.6$, and the initial* `pmf` $\boldsymbol{v}_1$ *be such that the even indexed entries of* $\boldsymbol{b}$ *be zero, and $b_1 = 0.3580$, $b_3 = 0.1192$, $b_5 = 0.0137$, $b_7 = -0.0002$, $b_9 = 0.0002$ and $b_{11} = 0.0001$.*

For such parameters, $\mathbb{1}^T\Psi_r(\boldsymbol{b})\mathbb{1} = -0.0063, -0.0180, -0.0215, -0.0129, -0.0042$

for $r = 0, 1, 2, 3, 4$, respectively. Therefore, Corollary 5.3.3 holds and

$$\{X_k \mid X_1^k \in \mathbb{D}\} \ \leq_r \ \{X_{k+1} \mid X_1^{k+1} \in \mathbb{D}\} \tag{5.61}$$

for all k.

Figure 5.3 shows the `pmf`*s $\boldsymbol{v}_k$ for $k = 1, 5, 10, 20$ on the left and the evolution of $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ until $k = 30$ on the right. The figure shows that the radial stochastic order is indeed increasing.*

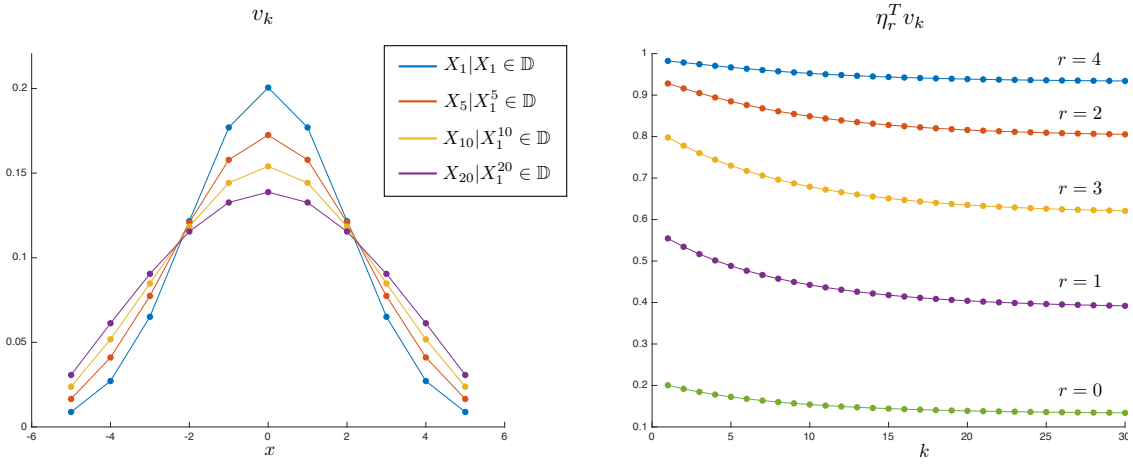


Figure 5.3: The evolution of the `pmf`s of $X_k \mid X_1^k \in \mathbb{D}$ (left) and $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ (right) for Example 5.3.4. The figure shows that the radial stochastic order is increasing, as guaranteed by Corollary 5.3.3.

5.3.3 A Special Case

This subsection explores a special case, when the initial `pmf` is restricted to be a linear combination of eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_3$, that is, the symmetric eigenvectors associated with the largest and third-largest eigenvalues. The study of `pmf`s derived

from these eigenvectors is important because, being associated with large eigenvalues, these eigenvectors dominate in the general case with all symmetric eigenvectors, as we will see in the next subsection. For now, we provide the following result:

Corollary 5.3.5. *Let the initial* `pmf` *be a linear combination of* $\boldsymbol{u}_1$ *and* $\boldsymbol{u}_3$ *with weights* b_1 *and* b_3*, such that:*

$$b_1 b_3 \big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) \leq 0 \tag{5.62}$$

for all r *in* $\{0, ..., n-1\}$*. The following holds:*

$$\{X_k \mid X_1^k \in \mathbb{D}\} \ \leq_r \ \{X_{k+1} \mid X_1^{k+1} \in \mathbb{D}\} \tag{5.63}$$

Proof. The proof is straightforward by computing:

$$\Psi_r(\boldsymbol{b}) = b_1 b_3 \ln(\lambda_1/\lambda_3)\big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) \tag{5.64}$$

$$\leq 0, \tag{5.65}$$

since $\ln(\lambda_1/\lambda_3) \geq 0$ for any choice of α and β, and $b_1 b_3 \big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) \leq 0$. The result follows from Lemma 5.3.4. □

Remark For all odd $n \leq 10^4$, we have found numerically that

$$\big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) > 0 \tag{5.66}$$

for all r in $\{0, ..., n-1\}$. Therefore initial `pmfs` with $b_1 > 0$ and $b_3 \leq 0$ will result in $X_k | X_1^k \in \mathbb{D}$ having increasing radial stochastic order for all k. When b_3 is positive, the radial stochastic order will be decreasing for all k.

Remark Note that the assumption in Corollary 5.3.5 does not rely on the eigenvalues of $\boldsymbol{P}$. It relies only only its eigenvectors, which are the same for all $\boldsymbol{P}$ of the type considered in this chapter.

In the following example the initial `pmf` is a linear combination of eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_3$ for $n = 5$ (see Fig. 5.4). Two cases are shown: $b_1 > 0$, $b_3 < 0$; and $b_1 > 0$, $b_3 > 0$.

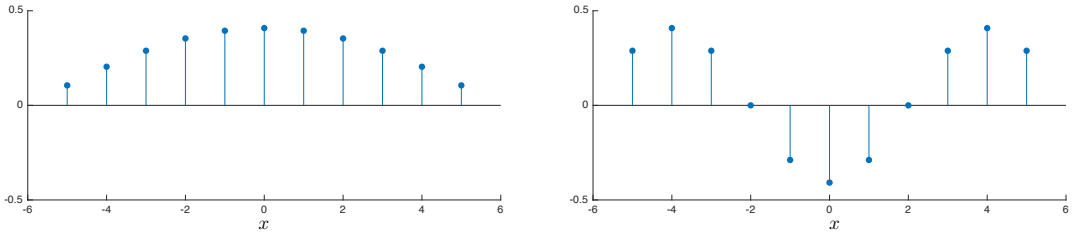


Figure 5.4: Eigenvectors u_1 and u_3 for $n = 5$.

Example 5.3.6. *Let the initial* `pmf` $\boldsymbol{v}_1$ *be a linear combination of* $\boldsymbol{u}_1$ *and* $\boldsymbol{u}_3$ *with weights* b_1 *and* b_3*:*

$$\boldsymbol{v}_1 = b_1\boldsymbol{u}_1 + b_3\boldsymbol{u}_3 \tag{5.67}$$

We consider two cases:

- $b_1 = 0.3649$ *and* $b_3 = -0.1336$
- $b_1 = b_3 = 0.2447$.

As shown in Fig. 5.5, when $b_1 > 0$ *and* $b_3 < 0$*, the radial stochastic order is increasing, as guaranteed by Corollary 5.3.5. When* $b_1 > 0$ *and* $b_3 > 0$*, the radial stochastic order is decreasing.*

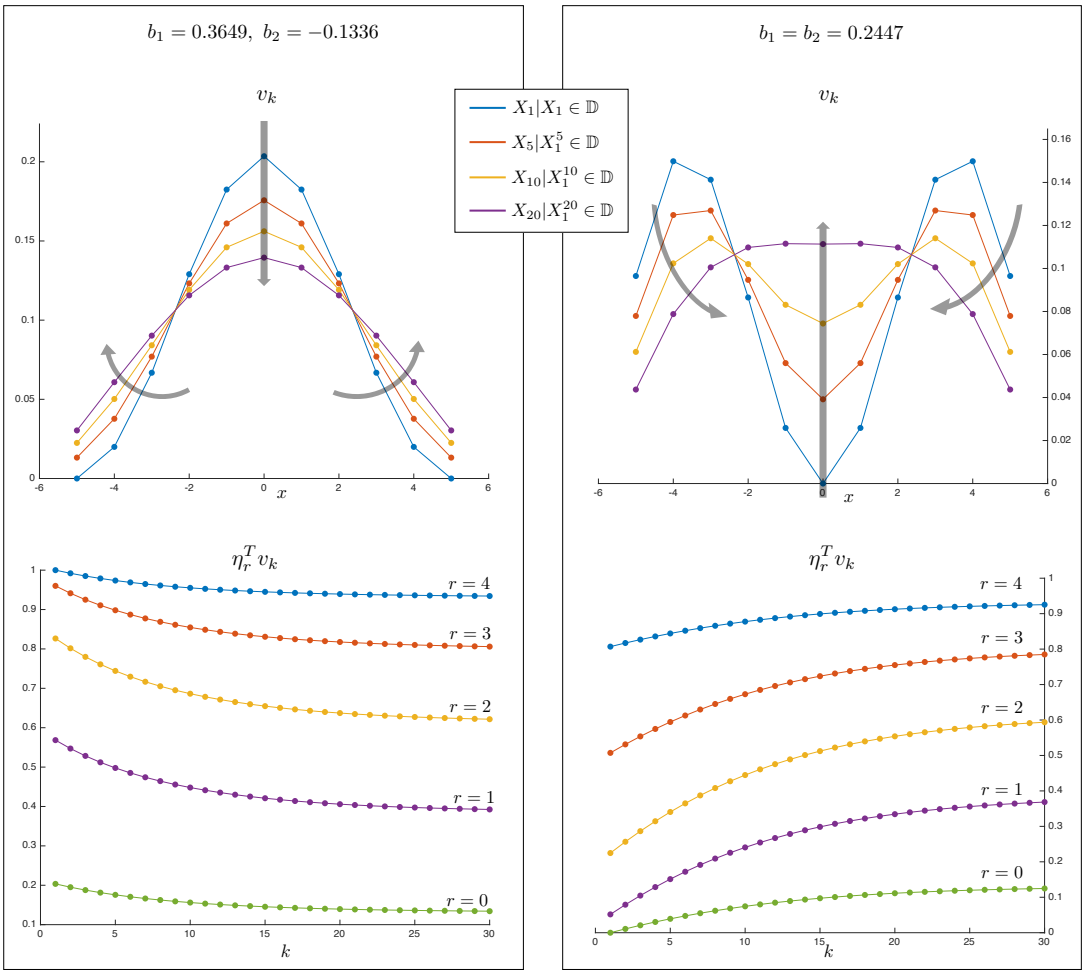


Figure 5.5: The evolution of $X_k \mid X_1^k \in \mathbb{D}$ and $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ for Example 5.3.6. On the left, with $b_1 > 0$, $b_3 < 0$, the figure shows that the radial stochastic order is increasing, as guarateed by Corollary 5.3.5. On the right, with $b_1 > 0$, $b_3 > 0$, the figure shows that the radial stochastic order is decreasing.

5.3.4 A Bound on k for Increasing Stochastic Order

So far we have considered radial stochastic orders that rely on the sign of $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ for each r and *for all time* $k \geq 1$. However, a natural extension is to consider situations when the radial stochastic order is increasing only after some $\tilde{k} > 1$, after

which the results developed so far apply.

Definiton 5.3.7. *A Markov chain Y_k is said to be increasing in the radial stochastic order after time $\tilde{k}$, written $Y_k \leq_{r,\tilde{k}} Y_{k+1}$, if*

$$\boldsymbol{\eta}_r^T \boldsymbol{p}_k \geq \boldsymbol{\eta}_r^T \boldsymbol{p}_{k+1} \tag{5.68}$$

for all $k \geq \tilde{k}$ and all r in $\{0, ...n\}$; where $\boldsymbol{p}_k$ is the `pmf` *of Y_k.*

The results presented in the section build upon the results of the previous section and provide a bound on k after which the radial stochastic order is guaranteed to be increasing.

Corollary 5.3.8. *Consider any valid weight vector $\boldsymbol{b}$ such that:*

$$b_1 b_3 \big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3\big) \leq 0 \tag{5.69}$$

for all r in $\{0, ..., n-1\}$. There exists a $\tilde{k}$ such that

$$\{X_k \mid X_1^k \in \mathbb{D}\} \ \leq_{rst} \ \{X_{k+1} \mid X_1^{k+1} \in \mathbb{D}\} \tag{5.70}$$

for all $k \geq \tilde{k}$.

Moreover, one such $\tilde{k}$ is given by:

$$\tilde{k} = ceil\Big(\arg\min_t \big\{ \ t \mid \Phi_{r,\boldsymbol{b}}(t) \leq 0, \ r \in \{0, ..., n-1\} \big\} \Big) \tag{5.71}$$

where:

$$\Phi_{r,\boldsymbol{b}}(t) \stackrel{def}{=} c_{13}^r + \sum_{j=5}^{2n+1} \sum_{i<j} \Big(\frac{\lambda_i \lambda_j}{\lambda_1 \lambda_3} \Big)^t \big| \ c_{ij}^r \ \big|, \tag{5.72}$$

and $c_{ij}^r \stackrel{def}{=} b_i b_j \ln(\lambda_i/\lambda_j) \big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j\big)$

Proof. Recall from the proof of Corollary 5.3.3 that:

$$\begin{aligned}
\boldsymbol{b}^T \boldsymbol{\Gamma}_r(t) \boldsymbol{b} &\leq (\lambda_1 \lambda_3)^t \Big\{ b_1 b_3 \ln(\lambda_1/\lambda_3) \big(d_3 \boldsymbol{\eta}_r^T \boldsymbol{u}_1 - d_1 \boldsymbol{\eta}_r^T \boldsymbol{u}_3 \big) + \\
&\qquad \sum_{j=5}^{2n+1} \sum_{i<j} \Big| b_i b_j \Big(\frac{\lambda_i \lambda_j}{\lambda_1 \lambda_3} \Big)^t \ln(\lambda_i/\lambda_j) \big(d_j \boldsymbol{\eta}_r^T \boldsymbol{u}_i - d_i \boldsymbol{\eta}_r^T \boldsymbol{u}_j \big) \Big| \Big\} \\
&= (\lambda_1 \lambda_3)^t \Big\{ c_{13}^r + \sum_{j=5}^{2n+1} \sum_{i<j} \Big(\frac{\lambda_i \lambda_j}{\lambda_1 \lambda_3} \Big)^t \Big| \; c_{ij}^r \; \Big| \Big\} \\
&= (\lambda_1 \lambda_3)^t \Phi_{r,\boldsymbol{b}}(t)
\end{aligned}$$

Clearly, there exists a $\tilde{t}$ for which:

$$\Phi_{r,\boldsymbol{b}}(t) \leq 0 \tag{5.73}$$

for all r and $t \geq \tilde{t}$. This can be seen by noting that c_{13} is negative by assumption, and that, since the ratio of eigenvalues is less than one for all (i, j) in question, all other summands, although positive, go to zero exponentially fast. The smallest $\tilde{t}$ for which $\Phi_{r,\boldsymbol{b}}(t)$ is less or equal to zero is:

$$\tilde{t} = \arg\min_t \Big\{ \; t \mid \Phi_{r,\boldsymbol{b}}(t) \leq 0, \; r \in \{0, ..., n-1\} \Big\} \tag{5.74}$$

The result follows with $\tilde{k} = \text{ceil}(\tilde{t})$. □

We finish this discussion with a numerical example:

Example 5.3.9. *Let $n = 5$, $\alpha = 0.6$ and consider an initial* `pmf` $\boldsymbol{v}_1$ *such that*

$$\mathbb{P}(X_1 = -2) = \mathbb{P}(X_1 = 2) = 0.4 \tag{5.75}$$

$$\mathbb{P}(X_1 = -4) = \mathbb{P}(X_1 = 4) = 0.1 \tag{5.76}$$

We begin by finding the weight vector $\boldsymbol{b}$ by projecting $\boldsymbol{v}_1$ on the eigenvectors $\boldsymbol{u}_j$'s. Since $\boldsymbol{v}_1$ is symmetric, the entries associated with the odd eigenvectors are

zero, and $b_1 = 0.3563$, $b_3 = -0.1493$, $b_5 = 0.1254$, $b_7 = 0.0437$, $b_9 = -0.3126$ *and* $b_{11} = 0.2746$

Based on a previous discussion and the fact that $b_1 > 0$ *and* $b_3 < 0$, *the inequality in (5.69) is satisfied. Computing the bound in (5.74), we get that* $\tilde{t} = 6.4550$ *and therefore the stochastic order is guaranteed to be increasing for all* k *greater than* 7.

Fig. 5.6 shows the plots of $\Phi_{r,\boldsymbol{b}}(t)$*for each* r *as well as the evolution of* $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ *for each* r. *The figure shows that the radial stochastic order is indeed increasing after* $k = 7$. *The evolution of the conditioned* `pmf`*s* $\boldsymbol{v}_k$ *can be seen in Fig. 5.7. The figure on the left represents the evolution up until* $k = 7$, *and the one on the right depicts the evolution from* $k = 7$ *until* $k = 30$ *(with increasing radial stochastic order).*

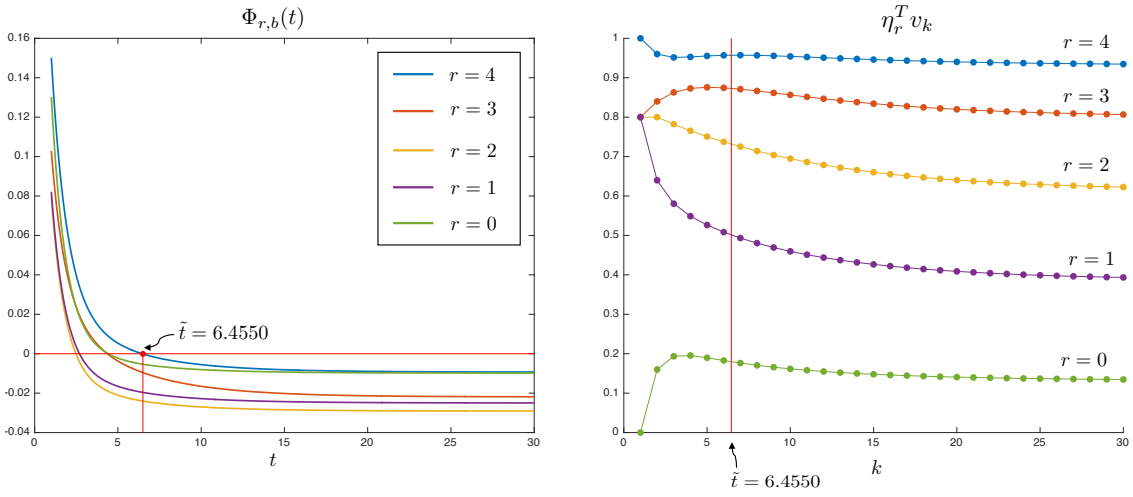


Figure 5.6: The plots of $\Phi_{r,\boldsymbol{b}}(t)$ (left) and the evolution of $\boldsymbol{\eta}_r^T \boldsymbol{v}_k$ (right) for Example 5.3.9. This figure shows that the radial stochastic order is increasing after $k = 7$.

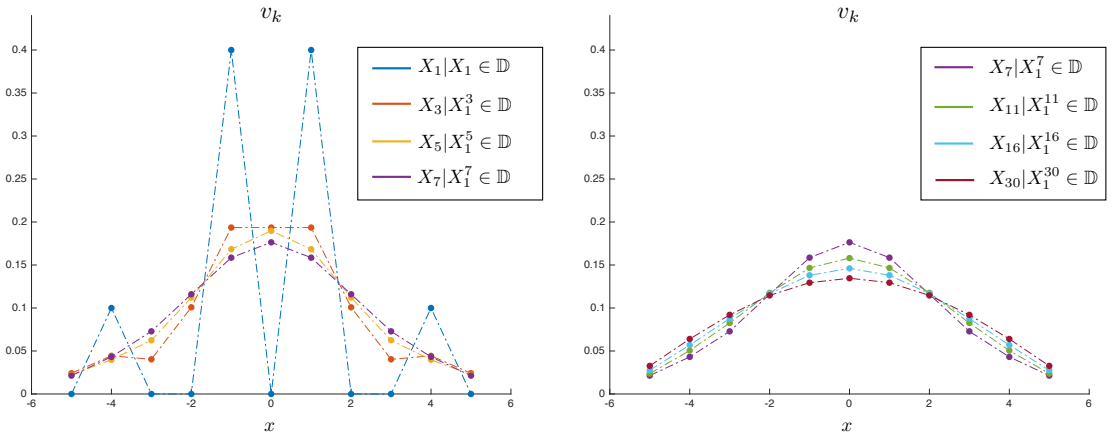


Figure 5.7: The evolution of $X_k \mid X_1^k \in \mathbb{D}$ for Example 5.3.9. The image on the left shows $\boldsymbol{v}_k$ for $k = 1, 3, 5$ and 7. On the right, $\boldsymbol{v}_k$ for $k = 7, 11, 16$ and 30. As expected, the radial stochastic order in increasing after $k = 7$.

5.4 Conclusion

In this chapter we have extended the results of Chapter 4 for Markov chains, and focused specifically on establishing conditions on the initial `pmf` of the chain for which the radial stochastic order of $X_k|X_1^k \in \mathbb{D}$ is increasing and, therefore, the results established in Chapter 4 can be applied.

The results have been achieved by writing the `pmf`s of the conditioned process in terms of the eigenvectors of a sub-matrix of the Markov chain. We established a main result in Theorem 5.3.2, from which other results follow. We showed that there exists an easy-to-check, albeit restrictive, sufficient condition for the theorem; we established conditions for initial `pmf`s that are linear combinations of the two symmetric eigenvectors associated with the largest eigenvalues; and we provided

conditions on the initial `pmf` and a bound on k after which the radial stochastic order of the Markov chain is guaranteed to be increasing.

Chapter 6: Conclusions

This thesis had two main focuses: *i*) the study of control design for persistent surveillance, where the goal is to design policies that achieve surveillance of the largest possible area, while respecting certain constraints; and *ii*) the problem of sensor usage for monitoring of denied environments, inside which state observations are costly. Our guiding philosophy was to theoretically analyze seemingly simple problems and, by doing so, to unearth *fundamental* concepts and design principles.

We addressed the design of full-state feedback memoryless policies for MDPs with finite state and action spaces. The main problem is to design policies that lead to the largest set of recurrent states, subject to convex constraints on the set of invariant `pmf`s. We described a finitely parametrized convex program that solves the problem via entropy maximization principles. Our approach has the advantage of yielding a closed-loop Markov chain with least number of recurrent classes.

Next, we have applied the results obtained in Chapter 2 to the problem of maximal persistent surveillance for robots whose dynamics are governed by MDPs. We dealt with safety constraints by casting a convex constraint on the invariant `pmf`, which allowed the application of our method. The simple structure of the resulting controllers enables their implementation in small robots.

In the second part of the thesis, we tackled the problem of optimal sensor usage in denied environments. We used the framework of renewal reward processes to establish a recursive expression for the cost we wished to minimize, and established conditions under which any local minimum, if it exists, is also global. We showed that when an auxiliary process has increasing radial stochastic order, the conditions for which the result holds simplify.

Next, we have extended the results on sensor usage in denied environments for Markov chains, and focused specifically on establishing conditions on the initial `pmf` of the chain for which the radial stochastic order of the auxiliary process is increasing. The results have been achieved by writing the `pmf`s of the conditioned process in terms of the eigenvectors of a sub-matrix of the Markov chain. We established a main result in Theorem 5.3.2, from which other results follow. We showed that there exists an easy-to-check, albeit restrictive, sufficient condition for the theorem; we established conditions for initial `pmf`s that are linear combinations of the two symmetric eigenvectors associated with the largest eigenvalues; and we provided conditions on the initial `pmf` and a bound on k after which the radial stochastic order of the Markov chain is guaranteed to be increasing.

6.1 Future Directions

A natural extension of our work, both in regards to the problem of maximal persistent surveillance and the problem of optimal sensor usage, is the validation of the results in a test-bed environment. The first problem would be best suited

for small robots with limited on-board processing and actuation capabilities. The second can be tested on small aerial vehicles with the aid of motion capture systems. We believe the results developed in this thesis would serve well as guiding principles in real world implementations.

On the theoretical front, the problem of optimal sensor usage can be further developed in many different directions, amongst which we highlight two. *i*) Although the work was developed for discrete-time systems, we believe it would extend nicely to the continuous case. The underlying process would be characterized by the diffusion equation and the auxiliary process, whose radial stochastic order is of importance, would be characterized by a diffusion equation with absorbing boundaries. *ii*) Since we considered binary control policies (reset or not), it would be interesting to see what happens in a traditional framework, in which the choice of control can affect the state in a more general fashion, such as in the context of linear time-invariant systems.

Bibliography

[1] B. Grocholsky, J. Keller, V. Kumar, and G. Pappas, "Cooperative air and ground surveillance," *IEEE Robotics Automation Magazine*, vol. 13, no. 3, pp. 16–25, Sep. 2006.

[2] L. Murray, J. Timmis, and A. Tyrrell, "Modular self-assembling and self-reconfiguring e-pucks," *Swarm Intelligence*, vol. 7, no. 2-3, pp. 83–113, May 2013. [Online]. Available: http://link.springer.com/article/10.1007/s11721-013-0082-y

[3] K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, and Y. Hada, "Redesign of rescue mobile robot Quince," in *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, Nov. 2011, pp. 13–18.

[4] D. F. Hougen, S. Benjaafar, J. C. Bonney, J. R. Budenske, M. Dvorak, M. Gini, H. French, D. G. Krantz, P. Y. Li, F. Malver, B. Nelson, N. Papanikolopoulos, P. E. Rybski, S. A. Stoeter, R. Voyles, and K. B. Yesin, "A miniature robotic system for reconnaissance and surveillance," in *IEEE International Conference*

on Robotics and Automation, 2000. Proceedings. ICRA '00, vol. 1, 2000, pp. 501–507 vol.1.

[5] S. d'Oleire Oltmanns, I. Marzolff, K. D. Peter, and J. B. Ries, "Unmanned Aerial Vehicle (UAV) for Monitoring Soil Erosion in Morocco," *Remote Sensing*, vol. 4, no. 11, pp. 3390–3416, Nov. 2012. [Online]. Available: http://www.mdpi.com/2072-4292/4/11/3390

[6] G. Zhou, C. Li, and P. Cheng, "Unmanned aerial vehicle (UAV) real-time video registration for forest fire monitoring," in *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, vol. 3, Jul. 2005, pp. 1803–1806.

[7] M. Dunbabin and L. Marques, "Robots for Environmental Monitoring: Significant Advancements and Applications," *IEEE Robotics Automation Magazine*, vol. 19, no. 1, pp. 24–39, Mar. 2012.

[8] I. Maza, F. Caballero, J. Capitn, J. R. Martnez-de Dios, and A. Ollero, "Experimental Results in Multi-UAV Coordination for Disaster Management and Civil Security Applications," *Journal of Intelligent & Robotic Systems*, vol. 61, no. 1-4, pp. 563–585, Dec. 2010. [Online]. Available: http://link.springer.com/article/10.1007/s10846-010-9497-5

[9] N. Michael, S. Shen, K. Mohta, Y. Mulgaonkar, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, K. Ohno, E. Takeuchi, and S. Tadokoro, "Collaborative mapping of an earthquake-damaged

building via ground and aerial robots," *Journal of Field Robotics*, vol. 29, no. 5, pp. 832–841, Sep. 2012. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/rob.21436/abstract

[10] M. K. Habib and Y. Baudoin, "Robot-assisted risky intervention, search, rescue and environmental surveillance," *International Journal of Advanced Robotic Systems*, vol. 7, no. 1, pp. 1–8, 2010. [Online]. Available: http://cdn.intechopen.com/pdfs-wm/6961.pdf

[11] Z. Li, Y. Liu, R. Walker, R. Hayward, and J. Zhang, "Towards automatic power line detection for a UAV surveillance system using pulse coupled neural filter and an improved Hough transform," *Machine Vision and Applications*, vol. 21, no. 5, pp. 677–686, Sep. 2009. [Online]. Available: http://link.springer.com/article/10.1007/s00138-009-0206-y

[12] A. P. Sabelhaus, D. Mirsky, L. M. Hill, N. C. Martins, and S. Bergbreiter, "TinyTeRP: A Tiny Terrestrial Robotic Platform with modular sensing," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 2600–2605.

[13] H. Aoyama, A. Himoto, O. Fuchiwaki, D. Misaki, and T. Sumrall, "Micro hopping robot with IR sensor for disaster survivor detection," in *IEEE International Safety, Security and Rescue Rototics, Workshop, 2005.*, Jun. 2005, pp. 189–194.

[14] N. Boillot, D. Dhoutaut, and J. Bourgeois, "Using Nano-wireless Communications in Micro-Robots Applications," in *Proceedings of ACM The First Annual International Conference on Nanoscale Computing and Communication*, ser. NANOCOM' 14. New York, NY, USA: ACM, 2007, pp. 10:1–10:9. [Online]. Available: http://doi.acm.org/10.1145/2619955.2619967

[15] P. Dario, R. Valleggi, M. C. Carrozza, M. C. Montesi, and M. Cocco, "Microactuators for microrobots: a critical survey," *Journal of Micromechanics and Microengineering*, vol. 2, no. 3, p. 141, 1992. [Online]. Available: http://stacks.iop.org/0960-1317/2/i=3/a=005

[16] M. J. Kuhlman, E. Arvelo, S. Lin, P. A. Abshire, and N. C. Martins, "Mixed-signal architecture of randomized receding horizon control for miniature robotics," in *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2012, pp. 570–573.

[17] C. Zheng, Q. Huang, W. Sang, Y. Liu, L. Wang, and Y. Huang, "Mechanical Design and Control System of a Miniature Autonomous Surveillance Robot," in *2007 International Conference on Mechatronics and Automation*, Aug. 2007, pp. 1752–1757.

[18] H. Choset, "Coverage for robotics – A survey of recent results," *Annals of Mathematics and Artificial Intelligence*, vol. 31, pp. 113–126, 2001.

[19] N. Nigam, S. Bieniawski, I. Kroo, and J. Vian, "Control of multiple UAVs for persistent surveillance: algorithm and flight test results," *IEEE Transactions*

on Control Systems Technology, vol. 20, no. 5, pp. 1236–1251, 2012.

[20] N. Nigam and I. Kroo, "Persistent surveillance using multiple unmanned air vehicles," in *IEEE Aerospace Conference*, 2008, pp. 1–14.

[21] P. F. Hokayem, D. Stipanovic, and M. W. Spong, "On persistent coverage control," in *IEEE Conference on Decision and Control*, 2007, pp. 6130–6135.

[22] B. Bethke, J. Redding, J. P. How, M. A. Vavrina, and J. Vian, "Agent capability in persistent mission planning using approximate dynamic programming," *American Control Conference*, pp. 1623–1628, 2010.

[23] X. C. Ding, S. L. Smith, C. Belta, and D. Rus, "MDP optimal control under temporal logic constraints," in *IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 532–538.

[24] S. L. Smith, M. Schwager, and D. Rus, "Persistent robotic tasks: monitoring and sweeping in changing environments," *IEEE Transactions on Robotics*, vol. 28, no. 2, pp. 410–426, 2012.

[25] N. Michael, E. Stump, and K. Mohta, "Persistent surveillance with a team of MAVs," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 2708–2714.

[26] A. Arapostathis, R. Kumar, and S. Tangirala, "Controlled Markov chains with safety upper bound," *IEEE Transactions on Automatic Control*, vol. 48, no. 7, pp. 1230–1234, Jul. 2003.

[27] A. Arapostathis, R. Kumar, and S. Hsu, "Control of Markov Chains With Safety Bounds," *IEEE Transactions on Automation Science and Engineering*, vol. 2, pp. 333–343, Oct. 2005.

[28] S. Hsu, A. Arapostathis, and R. Kumar, "On controlled Markov chains with optimality requirement and safety constraint," *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 6, Jun. 2010.

[29] W. Wu, A. Arapostathis, and R. Kumar, "On non-stationary policies and maximal invariant safe sets of controlled Markov chains," in *43rd IEEE Conference on Decision and Control*, 2004, pp. 3696–3701.

[30] E. Arvelo and N. C. Martins, "Maximizing the set of recurrent states of an MDP subject to convex constraints," *Automatica*, vol. 50, no. 3, pp. 994–998, Mar. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109814000211

[31] E. Arvelo, E. Kim, and N. C. Martins, "Maximal persistent surveillance under safety constraints," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 4048–4053.

[32] R. He, S. Prentice, and N. Roy, "Planning in information space for a quadrotor helicopter in a GPS-denied environment," in *IEEE International Conference on Robotics and Automation, 2008. ICRA 2008*, May 2008, pp. 1814–1820.

[33] S. Ahrens, D. Levine, G. Andrews, and J. P. How, "Vision-based guidance and control of a hovering vehicle in unknown, GPS-denied environments," in *IEEE*

International Conference on Robotics and Automation, 2009. ICRA '09, May 2009, pp. 2643–2648.

[34] Y. Watanabe, P. Fabiani, and G. L. Besnerais, "Simultaneous visual target tracking and navigation in a GPS-denied environment," in *International Conference on Advanced Robotics, 2009. ICAR 2009*, Jun. 2009, pp. 1–6.

[35] L. Doitsidis, S. Weiss, A. Renzaglia, M. W. Achtelik, E. Kosmatopoulos, R. Siegwart, and D. Scaramuzza, "Optimal surveillance coverage for teams of micro aerial vehicles in GPS-denied environments using onboard vision," *Autonomous Robots*, vol. 33, no. 1-2, pp. 173–188, Mar. 2012. [Online]. Available: http://link.springer.com/article/10.1007/s10514-012-9292-1

[36] T.-L. Lee and C.-J. Wu, "Fuzzy Motion Planning of Mobile Robots in Unknown Environments," *Journal of Intelligent and Robotic Systems*, vol. 37, no. 2, pp. 177–191. [Online]. Available: http://link.springer.com/article/10.1023/A:1024145608826

[37] A. Bry, A. Bachrach, and N. Roy, "State estimation for aggressive flight in GPS-denied environments using onboard sensing," in *2012 IEEE International Conference on Robotics and Automation (ICRA)*, May 2012, pp. 1–8.

[38] J. K. Lee, D. A. Grejner-Brzezinska, and C. Toth, "Network-based Collaborative Navigation in GPS-Denied Environment," *The Journal of Navigation*, vol. 65, no. 3, pp. 445–457, Jul. 2012. [Online]. Available: https://www.cambridge.org/core/journals/journal-of-navigation/

article/network-based-collaborative-navigation-in-gps-denied-environment/E374295572912C9F0377FF293CE43866

[39] K. J. Astrm, "Event Based Control," in *Analysis and Design of Nonlinear Control Systems*, A. Astolfi and L. Marconi, Eds. Springer Berlin Heidelberg, 2008, pp. 127–147, dOI: 10.1007/978-3-540-74358-3_9. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-74358-3_9

[40] J. Sijs and M. Lazar, "On Event Based State Estimation," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds. Springer Berlin Heidelberg, Apr. 2009, no. 5469, pp. 336–350, dOI: 10.1007/978-3-642-00602-9_24. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-00602-9_24

[41] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec. 2012, pp. 3270–3285.

[42] M. Lemmon, "Event-Triggered Feedback in Control, Estimation, and Optimization," in *Networked Control Systems*, ser. Lecture Notes in Control and Information Sciences, A. Bemporad, M. Heemels, and M. Johansson, Eds. Springer London, 2010, no. 406, pp. 293–358, dOI: 10.1007/978-0-85729-033-5_9. [Online]. Available: http://link.springer.com/chapter/10.1007/978-0-85729-033-5_9

[43] T. Henningsson, E. Johannesson, and A. Cervin, "Sporadic event-based control of first-order linear stochastic systems," *Automatica*, vol. 44, no. 11, pp. 2890–2895, Nov. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109808002550

[44] O. C. Imer and T. Basar, "Optimal Estimation with Limited Measurements," in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec. 2005, pp. 1029–1034.

[45] G. M. Lipsa and N. C. Martins, "Remote State Estimation With Communication Costs for First-Order LTI Systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, Sep. 2011.

[46] R. Cogill, S. Lall, and J. P. Hespanha, "A Constant Factor Approximation Algorithm for Event-Based Sampling," in *Perspectives in Mathematical System Theory, Control, and Signal Processing*, ser. Lecture Notes in Control and Information Sciences, J. C. Willems, S. Hara, Y. Ohta, and H. Fujioka, Eds. Springer Berlin Heidelberg, 2010, no. 398, pp. 51–60, dOI: 10.1007/978-3-540-93918-4_5. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-93918-4_5

[47] Y. Xu and J. P. Hespanha, "Optimal communication logics in networked control systems," in *43rd IEEE Conference on Decision and Control, 2004. CDC*, vol. 4, Dec. 2004, pp. 3527–3532 Vol.4.

[48] A. Arapostathis, V. Borkar, E. Fernandez-Guacherand, M. Ghosh, and S. Marcus, "Discrete-time controlled markov processes with average cost criterion: a survey," *SIAM Journal on Control and Optimization*, vol. 31, no. 2, pp. 282–344, Mar. 1993.

[49] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY: John Wiley and Sons, 1994.

[50] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx/, Apr. 2011.

[51] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 1990.

[52] P. Tabuada, *Verification and Control of Hybrid Systems: A Symbolic Approach*. Springer, 2010.

[53] I. Csiszar and P. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 1–116, 2004.

[54] K. Toh, R. Tutuncu, and M. Todd, "On the implementation and usage of SDPT3 a Matlab software package for semidefinite-quadratic-linear programming, version 4.0," Jul. 2006. [Online]. Available: http://www.math.nus.edu.sg/~mattohkc/sdpt3/guide4-0-draft.pdf

[55] E. D. Andersen, "Complexity of solving conic quadratic problems," Nov. 2013. [Online]. Available: http://erlingdandersen.blogspot.com/2013/11/complexity-of-solving-conic-quadratic.html

[56] V. Garg, R. Kumar, and S. Marcus, "A probabilistic language formalism for stochastic discrete-event systems," *IEEE Transactions on Automatic Control*, vol. 44, pp. 280–293, 1999.

[57] O. Hernandez-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, ser. Applications of Mathematics Series. Springer Verlag, 1996. [Online]. Available: http://books.google.com/books?id=NvPFQgAACAAJ

[58] P. Wolfe and G. Dantzig, "Linear Programming in a Markov Chain," *Operations Research*, vol. 10, no. 5, pp. 702–710, Oct. 1962.

[59] V. Borkar, "Controlled Markov chains with constraints," *Sadhana*, vol. 15, no. 4-5, pp. 405–413, Dec. 1990.

[60] B. Fox, "Markov Renewal Programming by Linear Fractional Programming," *SIAM Journal on Applied Mathematics*, vol. 14, no. 6, pp. 1418–1432, Nov. 1966.

[61] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

[62] V. Borkar, "Convex Analytic Methods in Markov Decision Processes," in *Handbook of Markov Decision Processes*, E. A. Feinberg and A. Shwartz, Eds. Boston, MA: Springer US, 2002, pp. 347–375.

[63] D. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*, 3rd ed. Athena Scientific, 2005.

[64] A. Hordijk and L. Kallenberg, "Linear programming and Markov decision chains," *Management Science*, vol. 25, no. 4, pp. 352–362, 1979.

[65] P. Kumar and P. Varaiya, *Stochastic systems*, ser. Estimation, identification, and adaptive control. Prentice Hall, 1986.

[66] E. Arvelo, E. Kim, and N. Martins, "Maximal Persistent Surveillance under Safety Constraints," in *IEEE International Conference on Robotics and Automation*, May 2013.

[67] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," *International Joint Conference on Artificial Intelligence*, 1995.

[68] A. Undurti and J. P. How, "A decentralized approach to multi-agent planning in the presence of constraints and uncertainty," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 2534–2539.

[69] J. Burlet, O. Aycard, and T. Fraichard, "Robust motion planning using Markov decision processes and quadtree decomposition," in *IEEE International Conference on Robotics and Automation*, 2004, pp. 2820–2825.

[70] S. M. Ross, *Stochastic Processes*, 2nd ed. Wiley, 1996.

[71] A. Mller and D. Stoyan, *Comparison methods for stochastic models and risks*. John Wiley & Sons Inc, 2002, vol. 389. [Online]. Available: http://scholar.google.com/scholar?cluster=12431483416228060085&hl=en&oi=scholarr

[72] W. F. Trench, "Banded symmetric Toeplitz matrices: where linear algebra borrows from difference equations," Tech. Rep., 2009.