Universidad
Carlos III de Madrid

PH.D. THESIS

Restless Bandit Index Policies for Dynamic Sensor Scheduling Optimization

Author:
Sofía S. Villar

Advisor:
José Niño-Mora

STATISTICS DEPARTMENT

Getafe, April 2012

To pipi,
(from sisi.)

Acknowledgements

I would like to thank my advisor Prof. José Niño-Mora for introducing me to a wonderful research topic which offers mathematical challenges as well as several relevant motivating applications. Had it not been for my fascination with this topic, I would never have had the eagerness and enthusiasm I have shown to complete this dissertation.

During the four years I worked on this thesis, I had the privilege of exchanging ideas with people who I consider to be excellent researchers and who not only helped me keeping my motivation but also provided me with constructive feedback which improved my work significantly. A special thank to Peter Jacko, who has provided, with endless kindness and patience, his insight and suggestions. I want to express my sincere gratitude to Bernardo D'Auria, who was always available to talk about research. It was a privilege and a pleasure to meet and have an enriching conversation with John Gittins.

Thanks to all the members of the Statistics Department, from faculty, staff and fellow Ph.D. students who made me feel a little bit at home during these years. I am sincerely grateful for having met and befriended superb colleagues from which I always learnt something new: Dae-Jin Lee, who despite the huge distance always managed to cheer me on; André, who was a great course mate, and Miguel Angel, who offered me endless support and great company in the final sprint. To my office mate Javi, I do not think I will ever have an office mate who writes fight club quotes with me on the whiteboard. To the not so dead Ph.D. students, especially to Leo and Alb, I am sorry we became dead by the end of my road, but I wish you come back to life soon! To Mei and Liu, my recent office mates, good luck.

Thanks to the people in Leganés who shared with me countless meals and coffees. Special thanks to María Durbán, it was a pleasure for me working with her as a teacher assistant and as a coordinator. I learnt a lot from her, professionally and as a person. I am also really happy for having shared many moments with people like Gema García, Elisa Molanes, Sergio García, Emilio Letón (thanks for all the *boletines* you forwarded me!), Belén Martín, Javier Nogales, Ismael Sánchez and Ignacio Cascos. A sincere thank

you to Esther Ruiz and Javier Prieto for their constant availability and kindness. To faculty from the Business Department, specially to Pablo Ruiz-Verdú and Luis Gómez Mejía who set for me an example of excellence in teaching and in research.

These last three years have been especially difficult for me. Those that know me better know how hard it was for me to let all my personal circumstances aside, and concentrate on making my greatest effort to fulfill my life-time dream. I would certainly not have made it without all the meals, the squash games, the mountain walks, etc., I shared with friends like Sara and Vincezo, Goki and Asli, Kenedy, Ana-María and Marius, Patricia, Agata and Lucas, Ana Laura, Pau, Cristina, Gabi. To my friends in Argentina, those who I always visited and those who I do not see as often as I would like. To Paz, Lola, and Gabi, every moment I spend with you, no matter how short, is amazing. You are great friends and I have missed you everyday since I arrived to Spain. To Victor who came to visit me also, and to Diego, Dalmira, Maru, Vero and all "Edudis". To Marcos, I would not have been here if it weren't for you. I will always be grateful for your support. To gabi, wherever you are.

To my family in Argentina, who always asked why my visits were so short. These pages have the answer. To my grandma Elba and to my godfather Negro, thanks for teaching me all you know about life. To my grandma Sara, who as Elba and Negro, came to visit me to see by themselves I was doing fine. To my parents, Néstor y Adriana, and to my sisters, Julie y Mati and to my brother Santiago. To my niece Sarita, who always complained when my visits were over. It would have made me immensely happy if you could have been here to share this moment with me. To my uncles Cristian and Leo who have been like my older brothers. To the rest of my family in Spain, to whom I never visited enough, Marta, Graciela, María Angeles, Gracielita, Carla, Varón y Gabi, and my Spanish cousins: Shyri and Alvaro. A special thank to my uncle Omar, you have always been an example for me. To the rest of my uncles, aunts, and cousins thanks!

This period of personal development was possible thanks to two scholarships and the funding of the Statistics Department of this University. My work has been supported in part by the Spanish Ministry of Education and Science project MTM2007-63140 and by the Ministry of Science and Innovation project MTM2010-20808.

Abstract

This dissertation addresses two complex stochastic and dynamic resource allocation problems, with application in modern sensor systems: (i) hunting multiple elusive hiding targets and (ii) tracking multiple moving targets. These problems are naturally formulated as Multi-armed Restless Bandit Problems (MARBPs) with real-state variables, which introduces technical difficulties that cause its optimal solution to be intractable. Hence, in this thesis we focus on designing tractable and well-performing heuristic policies of priority-index type.

We consider the above MARBPs as Markov Decision Processess (MDPs) with special structure, and we deploy recent extensions to the unifying principle to design a dynamic priority index policy based on a Lagrangian relaxation and decomposition approach. This approach allows to design an index rule based on a structural property of the optimal solution to the decomposed parametric-optimization subproblems. The resulting index is a measure of the Marginal Productivity (MP) of resources invested in the subproblems, and it is then used to define a heuristic priority rule for the original intractable problems.

For each of the problems under consideration we perform such a decomposition, to analyze the conditions under which the index recovering the optimal policies for the subproblems exists. We further obtain formulae for the indices which do not admit a closed form expression, but which are approximately computed by a tractable evaluation method.

Apart from the practical contribution of deriving the tractable sensor scheduling polices which improve on existing heuristics, the main contributions of this thesis are the following: (i) deploying the recent extensions of Sufficient Indexability Conditions (SIC) to the real state case, for two problems in which direct verification of the SIC and obtaining a closed-form index formula are not possible, (ii) addressing the technical difficulties to analyze PCL-indexability introduced by the uncountable state space of the MARBPs of concern, and the state evolution over it given by non-linear dynamics by exploiting the special structure of the trajectories of the state and the action processes under a threshold policy using properties of Möbius Transformations, and (iii) providing with

a tractable approximate evaluation method for the resulting index policies.

Resumen

Esta tesis estudia dos problemas dinámicos y estocásticos de asignación de recursos, con aplicación a sistemas modernos de sensores: (i) localización de mútiples objetivos evasivos que se ocultan y (ii) el rastreo de mútiples objetivos que se mueven. Estos problemas son modelizados naturalmente como problemas de "Multi-armed Restless Bandit" con variable de estado real, lo que introduce dificultades técnicas que causan que su solción óptima no sea computacionalmente tratable. Debido a esto, en esta tesis nos concentramos en cambio en diseñar políticas heurísticas de prioridad que sean computacionalmente tratables y cuyo rendimento sea casi óptimo.

Modelizamos los problemas arriba mencionados como problemas de decisión Markovianos con estructura especial y les aplicamos resultados existentes en la literatura, los que constituyen un principio unificador para el diseño de políticas de índices de prioridad basadas en la relajación Lagrangiana y la descomposición de esos problemas. Este enfoque nos permite considerar una propiedad de los subproblemas: la indexabilidad, por la cual podemos resolverlos de manera óptima mediante una política índice. El índice resultante es una medida de productividad de los recursos invertidos en los subproblemas, y es usado luego como medida de la prioridad dinámica para los problemas originales intratables.

Para cada uno de los problemas bajo estudio realizamos tal descomposición, y analizamos las condiciones bajo las que una política índice que recupere la solución óptima de los subproblemas existe. Además obtenemos fórmulas para los índices, las que a pesar de no admitir una expresión cerrada, son calculadas aproximadamente de manera eficiente meadiante un método tratable.

Aparte de la contribución práctica de obtener reglas heurísticas de índices de prioridad para el funcionamiento de sistemas de múltiples sensores en el contexto de los dos problemas analizados, las principales contribuciones teóricas son las siguientes: (i) la aplicación de las extensiones recientes de las condiciones suficientes de indexabilidad para el caso de variable de estado real, para dos problemas en los que tanto la verificación directa de ellas como la obtención de fórmulas cerradas no son posibles, (ii) el tratamiento de las dificultades técnicas para establecer la indexabilidad introducidas por el espacio de estado infinito de los problemas bajo consideración, y por la evolución sobre este estado dada por dinámicas no lineales, explotando propiedades estructurales de los procesos de la variable de estado y trabajo bajo políticas de umbral como recursiones de Transformaciones de Möbius, and (iii) un método aproximado de evaluación de las políticas de índices resultantes.

Contents

List of Figures

Acronyms

MDP	Markov Decision Processes	iii
MARBP	Multi-armed Restless Bandit Problem	iii
MABP	Multi-armed Bandit Problem	4
RB	Restless Bandit	3
PRI	Pulse Repetition Interval	7
TEV	Tracking Error Variance	9
POMDP	Partially Observed Markov Decision Process	10
JMPD	Joint Multi-target Probability Density Function	11
HMM	Hidden Markov Model	11
MSE	Mean Square Error	13
MCP	Markov Control Process	16
ET	Expected Total	20
ETD	Expected Total Discounted	20
LRA	Long Run Average	20
DPE	Dynamic Programming Equations	20
DP	Dynamic Programming	21
PCL	Partial Conservation Law	27
MP	Marginal Productivity	iii
TEV	Tracking Error Variance	9
STEV	Scaled Tracking Error Variance	79
LFT	Linear Fractional Transformation	125
SIC	Sufficient Indexability Conditions	iii
w.p	with probability	

Part I

Background

A journey of a thousand miles
must begin with a single step.
Lao Tzu.

Chapter 1

Introduction

1.1 Research Contributions and Thesis Organization

The fundamental economic problem arises because resources are scarce in relation to their alternative uses. *Scarcity* forces economic agents to make decisions among those possible uses, and therefore to sacrifice the benefits of the unselected ones. Moreover, these decisions must be made under *uncertain* conditions, which further evolve over *time*. In economics, the value of the second best alternative forgone is known as *opportunity cost* and rational decision agents are assumed to make choices which yield the minimum expected opportunity cost over time. Such an economic problem, namely how to optimally allocate scarce resources under uncertainty over time, is ubiquitous and leads us to ponder how to set priorities among the activities competing for our limited resources.

This dissertation addresses two concrete problems of this sort, which arise in modern sensing systems, thereby making a twofold contribution. First, at the broadest level, a major contribution of this dissertation is the design of novel and well-performing tractable sensing policies for two of the most challenging applications in *sensor management*: *smart target detection* and *multitarget tracking*. Second, at the deepest level, this dissertation contributes to the indexation literature for Restless Bandits (RBs) by favorably solving the challenges posed by the specific technical difficulties of these applications, as, e.g., its real state-space.

Modern sensing technologies offer the possibility of efficiently performing tasks by adaptively deploying its sensing resources based on the information extracted from past measurements. Yet, realizing such system's overall performance gains requires appropriate *on-line* sensing rules. Thus, the general problem in *sensor management* is to design sensing algorithms that allow for the fruitful adoption of cutting edge technologies. A natural procedure to derive those rules is to represent the underlying resource allocation problem by some stochastic dynamic optimization model, whose optimal solution is traditionally characterized by a dynamic programming framework. However, those formulations, at least for realistic scenarios, typically have a prohibitively large size (possibly infinite), which dramatically hinders its practical application. Thus, fully exploiting the performance advantages offered by the new technologies by means of *active* dynamic sensing policies remains very challenging, mainly due to the well known *curse of dimensionality*. For this reason, the design of both computationally feasible and nearly-optimal sensing strategies, as the ones proposed in this thesis, continues to be a highly active applied research area. For a comprehensive review of the most general issues originating the sensor management literature, see e.g. Xiong and Svensson (2002); Ng and Ng (2000).

The approach followed in this work to achieve such a practical contribution is to formulate both applications as stochastic dynamic optimization models within a special class of Markov Decision Processes (MDP): the Multi-armed Restless Bandit Problem (MARBP) with real state projects, seeking to exploit the special structure of its optimal policy (when possible) to design a tractable heuristic of *priority index-type*. Such a class of policies defines, for each alternative, an index which represents the priority that allocating resources to that use should have, given its state at a given period of time. Naturally, the resulting priority index policy distributes the available resources to the alternatives yielding the currently largest index values, as long as they are profitable. Probably the most fruitful example of this procedure is the optimality of a priority-index rule for the *classical* Multi-armed Bandit Problem (MABP) in Gittins and Jones (1974). For alternative proofs of this fundamental result in the discrete state case, each offering complementary insights see, e.g., Whittle (1980), Varaiya et al. (1985), Weber (1992),Bertsimas and Niño-Mora (1996).

However, as shown in this dissertation, deriving a solution strategy based on such indexation approach for these two specific *sensor management* applications, as well as many others, raises substantial research challenges. Specifically, a fundamental issue is that both models call for the use of the *restless* variant of the MABP, for which the existence of an index rule that yields its optimal solution is not guaranteed. Furthermore, even if such an index exists, for the sake of practical implementation, providing with an efficient index evaluation method is of key relevance. This dissertation solves these challenges, among others, by application of the methodology based on a Lagrangian relaxation and decomposition approach introduced by Whittle (1988) and further developed by Niño-Mora into a systematic framework in work reviewed in Niño-Mora (2007a).

The present work illustrates the advantages of Niño-Mora’ s approach to alleviate the often baffling effort required to exploit special structure of resource allocation problems of this sort by deploying it to the formulations of concern. Additionally, the index policies proposed in this dissertation constitute, together with work in Niño-Mora (2009), preeminent early examples of the application and effectiveness of the general indexability conditions for real-state RBs introduced in Niño-Mora (2008).

Consequently, the contributions of this thesis are relevant to: (a) practitioners, who may wish to augment the productivity of their sensing systems; (b) researchers, who are challenged to design efficient algorithms to optimize a dynamic and stochastic systems, and (c) researchers struggling to design a mathematically based priority-index rule for real state RB models.

The thesis is structured as follows: the first three chapters provide the reader with an introductory description of the applied problems of concern as well as with the necessary overview of the basic methodological aspects of their theoretical formulations. Thus, the remainder of Chapter 1 discusses some of the main challenges and open problems in *sensor management*, and describes the ones related to the specific applications of concern. The chapter concludes with an overview of the two algorithms for recursive estimation most commonly used in signal processing applications, which provide the state dynamics of both proposed dynamic optimization models. Next, Chapter 2 reviews theoretical aspects of RB problems that naturally precede the background on RB indexation presented in Chapter 3.

The rest of the thesis presents the methodological and applied results, and it is structured as follows. Chapter 4 introduces the target hunt formulation and its indexability analysis while Chapter 5 discusses the corresponding computational experiments. Chapter 6 introduces the multitarget formulation and its indexability analysis while Chapter 7 discusses the corresponding computational experiments.

To conclude, Chapter 8 provides a summary of our work and its main research contributions, and discusses directions for future research.

There are 3 appendices at the end of this thesis. The first Appendix presents a review of basic concepts and properties of Möbius Transformations while the other two Appendices presents the detailed proofs of the indexability sections of Chapter 4 and Chapter 6.

1.2 Sensor Management Problems

Over the past few years, advances in sensing technology have provided modern multisensor systems with an increased operating flexibility to achieve different performance objectives, for instance when performing tasks as target *detection*, *tracking* or *identi-*

fication. The common element in the novel features introduced by such technological advances reduces to an increased agility of selection among possible sensing actions. In traditional sensing systems, parameters such as beam direction or waveform mode among others, are typically hard-wired, i.e. they are selected by fixed *off-line* approaches. In contrast to that, agile systems are capable of electronically controlling their sensing parameters during system operation so as to best extract information from the scene. Such an unprecedented feature provides new systems with the possibility of rapidly adapting its functioning to suit a variety of highly dynamic environments, which in turn raises a general s*ensor management* question: how to dynamically allocate sensing resources and modalities to optimize the system's overall performance?

The widespread adoption of these cutting-edge technologies has therefore led to research activity (both in academia and in industry) that seeks to improve modern system's performance through an adequate design of *on-line* active sensing schemes. The decision problem can be summarized as follows: the system's manager must select the parameters of the system's sensing resources sequentially over time, where each decision will provide him/her with a reward (in terms of the information gained) which is uncertain. Thus, the system's manager goal is to select parameters over time so as to maximize the total expected reward generated as a result of the system's operation.

Within Operations Research the interest has been mostly concentrated in the development of appropriate *scheduling algorithms* or *heuristics* to fully exploit the benefits of flexible systems. The design of such active sensing rules has been thus naturally posed as the optimal solution to some stochastic sequential resource allocation problem. See Williams (2007); Washburn et al. (2002); Krishnamurthy and Evans (2001); Castanon (1997) for examples of the application of these ideas. Yet, these specific applied problems have also motivated significant research efforts in related areas such as signal processing, statistics, or machine learning, thus becoming a multidisciplinary research literature which is now currently known as *sensor management*. A notorious example in such a vein is given by the book Hero et al. (2006).

Besides the specific challenges posed by a particular *sensor management* problem, like multiple target tracking, there are some general issues which play a prominent role in the design of active system's operating policies for the inherent benefits of these flexible systems to be fully realized. Such general issues are the following: (i) the real-time operational management of modern sensing systems requires *implementable* scheduling algorithms, which ideally run in polynomial time, since they will be *on-line*; (ii) the need to account for the long term effects of current actions to achieve greater performance gains calls for non-myopic policies; (iii) when the system is to be used in fairly distinct environments, *robustness* of scheduling methods is of vital importance (i.e., rules leading to near-optimal performance in one environment should not yield in a poor performance in another environment) and; (iv) policy design should take into account that a low system utilization may become highly advantageous, either for the sake of maximizing system's lifetime in battery-constrained networks or simply because idle radar time can be allocated to other tasks in multi-function radars.

The development of *sensor management* scheduling policies going from a theoretical approach based on stochastic control theory to its practical application exposes a stark *trade-off* between issue (i) and the issues (ii)-(iv): optimal dynamic stochastic decision rules which are robust, non-myopic and cost efficient have computationally intensive requirements, whereas computationally efficient suboptimal heuristics are implementable at the expense of losing robustness, cost efficiency or long-run performance gains. The scheduling rules proposed in this thesis, as illustrated by the computational experiments reviewed in Chapter 5 and Chapter 7, successfully address such a give-and-take between implementability and optimality loss.

In the next subsections, a basic description of the two specific motivating applications is provided together with a brief overview of previous related work.

1.2.1 Hunting Elusive Hiding Targets

In Chapter 4 we formulate and investigate the following problem.

Problem 1. There are N independent locations (or sites), each containing (at most) one target (or object) hidden in it. There are M $(1 \leq M \leq N)$ sensors, each of which at every discrete period can search at most one of those locations. All sensors in the system are synchronized to operate over time slots $t = 0, 1, \ldots$, where a time slot corresponds to a Pulse Repetition Interval (PRI).

Each target can choose between two possible *visibility* states: a *hidden* state, in which it is invisible to sensors but cannot perform its tasks, and an *exposed* state, in which it can perform its tasks but is detectable by sensors. Targets are such that: 1) they perceive they are being sensed; 2) they do not wish to be found, but they wish to perform their tasks; and 3) they react to sensing by becoming *elusive*. Thus, if a target n is in the *hidden* state in period t it becomes *exposed* in period $t + 1$ with probability p_n^0 if its location is not sensed in period t and with probability p_n^1 if sensed. Further, if a target is in the *exposed* state in period t it becomes *hidden* with probability q_n^0 if its location is not sensed in period t and with probability q_n^1 if sensed. Finally, to model the elusive reaction of the targets we assume that $p_n^0 > p_n^1$ and $q_n^0 < q_n^1$. See Figure 1.1.

The probability that a sensor searching target n finds it when it is *visible* is $0 < \alpha_n \leq 1$, and hence the probability that an unfound target is visible at slot t changes by Bayes' theorem as the sensor's detection output is observed. The cost of a single search of a location that possibly contains target n is $c_n \geq 0$ and yields a reward $r_n \beta^t$ when it succeeds at finding target n in slot t, where $0 \leq \beta \leq 1$ is a discount factor.

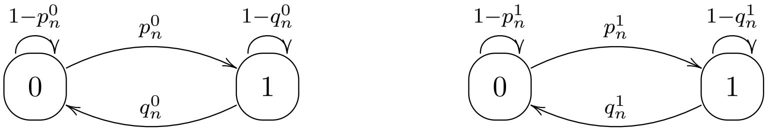


Figure 1.1: A model of a 2-state Markov chain. The arrows represent one-period transitions among the states 0 (hidden) and 1 (exposed) with given probabilities under actions 0 (on the left) and 1 (on the right).

The goal is to design a tractable policy which addresses the following question:

How should the N locations be scheduled for being sensed so as to be close to maximizing the total expected discounted reward of finding all targets, using at most M sensors at each time slot?

The main concern in Problem 1 is to determine how to conduct the search of the targets with the available sensors so as to achieve the stated performance objective. In this sense, the problem is a *search problem*. Search problems have been the subject of scientific research for more than sixty years now, constituting one of the oldest areas of *Operations Research*. Actually, initial research on search problems can be traced to work done by Bernard Koopman during World War II, when the term *Operations Research* was coined, to refer to the attempts of finding most efficient and effective ways of conducting military missions. In fact, at that time one of the most important military operations during war was *searching*. Koopman (1946), through his work for the US Navy trying to provide efficient methods for detecting submarines, laid the basis for later developments in search theory.

Developments of the theory have taken place since then along many directions, and have appeared scattered through the literature of operations research, applied mathematics, optimization theory and statistics. A comprehensive presentation of major results in search theory published during the 30 years that followed Koopman's work is the book by Stone (1975). The most complete results refer to the optimal search problem for a unique stationary object hidden within a discrete set of locations and when no false targets (i.e., objects capable of causing a detection but which are not objective targets) are present. Two examples of this type of problems of particular relevance to this dissertation, are the search problems in Gittins (1989, chap. 8) and in Song and Teneketzis

(2004), which are formulated into a *classic* MABP and are thus optimally solved by an index policy.

There have also been efforts to address search problems dealing with more general situations, e.g. with multiple objects, or with mobile objects, and even to include false targets. Yet, despite this abundant literature, the case in which targets may evade the searcher, as it occurs in Problem 1, remains understudied today. The main obstacle to deploying the strategy followed by Gittins (1989, chap. 8) and in Song and Teneketzis (2004) is the fact that when modeling moving or elusive targets the natural MARBP formulation is *restless*. There are some papers implementing alternative approaches. For instance, in Savage and La Scala (2009) game theory is applied to formulate and solve search problems with reactive targets. Reinforcement learning is used in Kreucher et al. (2006) to derive a non-myopic scheduling rule for both detection and tracking of "*smart*" targets, a particle filter approach is used in Liu et al. (2009), while in Rucker (2006) agent-based modeling is used to address an application model similar to Problem 1.

1.2.2 Multitarget Tracking

In Chapter 6 we formulate and investigate the following problem.

Problem 2. There are N independent targets whose position on the real line $x_{n,t}$ changes stochastically over time periods $t = 1, 2, \ldots$, following linear Gauss-Markov dynamics, i.e., with the increments corresponding to a zero-mean white noise with variance q_n. There are M $(1 < M \leq N)$ phased array radars, each of which at every discrete period of time can track at most one of those targets. All radars in the system are synchronized to operate over time slots $t = 0, 1, \ldots$, where a time slot corresponds to a PRI. Any radar, if allocated to measure target n's position, provides a noisy *measurement* $y_{n,t}$ of its true position $x_{n,t}$. Measurements are also generated by linear Gauss-Markov dynamics having zero-mean white noise with variance r_n.
The optimal minimum-variance predicted position of target n at slot t for this model is given by the Kalman Filter prediction and updating equations, depending on whether a radar's measurement for that target is available at time t or not. The Tracking Error Variance (TEV) $p_{n,t}$, measuring the uncertainty in target n's track, will be larger when no measurement of that target's position is available at the beginning of the slot.
The system incurs a sensing cost $c_n \geq 0$ when measuring target n's position for a single slot and, when predicting target n's position at slot t, it incurs a precision cost/error equal to $p_{n,t}\beta^t$, where $0 \leq \beta \leq 1$ is a discount factor.
The goal is to design a tractable policy which addresses the following question:

How should the targets be scheduled for being measured so as to be (at least) close to minimizing the total expected discounted sensing and precision cost of tracking all N targets using at most M sensors at each time slot?

In this problem, the main concern is to derive the most precise and cost efficient target track updates scheduling policy. Multiple moving target tracking models such as this one have been one of the earliest and most challenging applications of sensor management. Early work on the subject dealt with the minimization of radar energy required for track maintenance, see, e.g., van Keuk and Blackman (1993), Stromberg (1996), Hong and Jung (1998). Since the 1960's there have been many solutions proposed for addressing this problem, mostly based on the use of the Kalman Filter. However, recent progress on particle filter approaches has been extended to Bayesian multi-target tracking problems as well.

In Krishnamurthy and Evans (2001) a beam scheduling algorithm is derived from a discrete-time and discrete-state Partially Observed Markov Decision Process (POMDP) model, assuming that targets' motion from one PRI to the next is negligible (i.e., targets are *stationary*). Exploiting the special structure of the resulting POMDP as a classic MABP, the optimal policy is characterized in terms of an index policy. Of particular relevance to this dissertation is the work of La Scala and Moran (2006), in which the inadequacy of assuming the negligibility of targets' motion is pointed out and the authors extend the results in Howard et al. (2004) on optimality of the myopic-index scheduling policy for tracking two symmetric targets to more general linear dynamical systems under the same finite-horizon total TEV performance objective. Despite remarking that such a problem falls within the framework of the restless MABP, they suggest a heuristic policy which is not based on the indexation approach deployed in this work and which, as shown by simulation experiments in Niño-Mora and Villar (2009), does not perform well in the case of multiple asymmetric targets.

1.3 Statistical Signal Processing

Recursive estimation plays a central role in many applications of signal processing which are commonly encountered in sensor management problems, e.g., in target tracking and in navigation applications. Whenever we must infer the knowledge about some parameters which are indirectly observable from the outcome of a related experiment, and this knowledge can be updated as new measurements are collected, the underlying estimation is naturally done recursively.

In the following we review two particular estimation algorithms for optimally and recursively estimating an underlying parameter of interest: the general Bayes filter and a special case of that filter for the multivariate normal distributions: the Kalman filter. The former is used when formulating the MABP model for Problem 1, while the latter is used for the formulation of Problem 2.

The Bayesian approach to filtering has recently become widely adopted in sensor management applications, especially in multi-target tracking problems. For the latter applications, a Joint Multi-target Probability Density Function (JMPD) describing the posterior density given past measurements (also known as the *belief* state) is defined together with the construction of a filter to update it as measurements become available according to the usual rules of Bayesian filtering. The main issue with this approach is naturally the computational cost of the JMPD as the number of targets increases. The incorporation of such filters within this literature has been justified by its capability of handling situations which the Kalman Filter fails to address, as for example non-linear states or non Gaussian measurements.

However, the importance of the Bayesian Filter for sensor management applications is actually enlarged by the fact that in most sensor management applications, the full state of the system is not directly observable, instead a noisy measurement is available. Hence, when formulating such sensor management problems as Markov Decision Processes (MDP), the resulting models fall within the framework of POMDPs. The POMDP is equivalent to a standard MDP whose state variable corresponds to a belief state which evolves according to a Bayes rule. In subsection 2.1.5 we deal with this issue in more detail.

The General Recursive Bayesian Filter is an algorithm used for estimating the current unobservable state variable in a Hidden Markov Model (HMM) given past observations. A HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. HMMs are especially known for their application to robotics and bioinformatics.

Let the true state variable, denoted as $X_t \in \mathbb{R}^k$, follow an unobserved Markov process over time, that is

$$\mathcal{P}(X_t|X_{t-1},\ldots,X_0) = \mathcal{P}(X_t|X_{t-1}), \tag{1.1}$$

and assume a stochastic measurement process $Y_t \in \mathbb{R}^k$ such that:

$$\mathcal{P}(Y_t|X_t,\ldots,X_0) = \mathcal{P}(Y_t|X_t). \tag{1.2}$$

Then, under these assumptions, the joint probability distribution for a vector of measurements and unobserved states is computed as follows:

$$\mathcal{P}(Y_t, \ldots, Y_0, X_t, \ldots, X_0) = \mathcal{P}(X_0) \prod_{j=1}^{t} \mathcal{P}(Y_j|X_j)\mathcal{P}(X_j|X_{j-1}). \tag{1.3}$$

Thus, a natural way to *predict* the unobservable state in t given information up to $t-1$ is to use the probability distribution of the state variable at period t given the measurements available at period $t-1$. That distribution is given by:

$$\mathcal{P}(X_t|Y_{t-1}) = \int \mathcal{P}(X_t|X_{t-1})\mathcal{P}(Y_{t-1}|X_{t-1})dX_{t-1}. \tag{1.4}$$

Next, once the measurements at period t becomes available, the prediction of the state at t may be *updated*, by means of the probability distribution associated with the state for t. That distribution is given by:

$$\mathcal{P}(X_t|Y_t) = \frac{\mathcal{P}(Y_t|X_t)\mathcal{P}(X_t|Y_{t-1})}{\int \mathcal{P}(Y_t|X_t)\mathcal{P}(X_t|Y_{t-1})dX_t}. \tag{1.5}$$

Thus, the *predictive* step of the filter is the conditional expectation $\hat{X}_{t|t-1} \triangleq \mathsf{E}\left[X_t|Y_{t-1}\right]$, while the *updating* step of the filter is $\hat{X}_t \triangleq \mathsf{E}\left[X_t|Y_t\right]$.

In the special case of a HMM in which both the unobservable state variable and the measurement processes follow linear dynamics perturbed by a Gaussian noise, then the Bayes Filter becomes the Kalman Filter. In its most general version, the associated linear unobserved component model can be written as the following state space model:

$$X_t = F_t X_{t-1} + C_t + \Omega_t, \quad t \geq 1, \tag{1.6}$$

$$Y_t = H_t X_t + D_t + N_t, \quad t \geq 0, \tag{1.7}$$

where Ω_t and N_t are two independent and identically distributed (i.i.d) zero-mean Gaussian white noise with variance Q_t and R_t, respectively called as the *position-noise process* and *measurement-noise process*, and F_t and H_t are in $\mathbb{R}^{k\times k}$ and C_t and D_t are in $\mathbb{R}^k$. Q_t, R_t, F_t, H_t, C_t and D_t are predetermined parameters, in the sense that they are known at time $t-1$. In the case they are fixed for all t, the model is said to be *time-invariant*. In this Gaussian state space model, equation (1.6) is commonly known as *transition equation* while equation (1.7) as the *measurement equation*. The initial state X_0 is assumed to be normally distributed with mean $\boldsymbol{\mu_0}$ and variance $\boldsymbol{\Sigma_0}$ and both noise processes are assumed to be uncorrelated with the initial state and also with each other (which due to the normality assumption is ensured if $\mathsf{E}(X_0, \Omega_t) = \mathbf{0}$ and $\mathsf{E}(X_0, \nu_s) = \mathbf{0}$ for all $t = 1, 2, \ldots, T$ and $\mathsf{E}(\Omega_t, \nu_s) = \mathbf{0}$ for all $s, t = 1, 2, \ldots, T$.)

The verification of properties (1.1) and (1.2) is straightforward, in fact in this case those probabilities distributions are computed using properties of the multivariate normal distribution to be:

$$\mathcal{P}(X_t|X_{t-1}) = \mathcal{N}\left((F_{t-1}\hat{X}_{t-1} + C_t), Q\right) \tag{1.8}$$

$$\mathcal{P}(Y_t|X_t) = \mathcal{N}\left((H_{t-1}\hat{X}_t + D_t), R\right) \tag{1.9}$$

Therefore, the predictive $\hat{X}_{t|t-1}$ and updating $\hat{X}_t$ steps defined as in (1.4) and (1.5) result in the following estimation equations which constitute the Kalman filter algorithm

$$\mathsf{E}\left[X_t|Y_{t-1}\right] = F_t\hat{X}_{t-1} + C_t \tag{1.10}$$

$$\mathsf{E}\left[X_t|Y_t\right] = \hat{X}_{t|t-1} + P_{t|t-1}Y_tS_t^{-1}i_t, \tag{1.11}$$

where $P_{t|t-1} \triangleq \mathsf{E}\left[\left(X_t - \hat{X}_{t|t-1}\right)^2 |Y_{t-1}\right]$ is the Mean Square Error (MSE) associated to the predictive step estimator, $i_t \triangleq \left(Y_t - \hat{Y}_t\right) = H_t\left(X_t - \hat{X}_{t|t-1}\right) + N_t$, is known as the *innovation* or *measurement* residual. Notice that $i_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{S_t})$ with $S_t \triangleq \mathsf{E}(i^2) = H_tP_{t|t-1}H_t' + R_t$. Define also the MSE associated to the updating step estimator as $P_t \triangleq \mathsf{E}\left[\left(X_t - \hat{X}_t\right)^2 |Y_t\right]$. Both MSE are respectively computed to be:

$$P_{t|t-1} = F_tP_{t-1}F_t' + Q_t \tag{1.12}$$

$$P_t = P_{t|t-1} - P_{t|t-1}H_t'S_t^{-1}H_tY_tP_{t|t-1} \tag{1.13}$$

These results follow from the properties of the multivariate normal distribution. In particular, results (1.11) and (1.13) require properties of the conditional distributions and moments of the random vector $(X_t, Y_t)|Y_{t-1}$.

Under the assumptions that both the initial state and the disturbances are normally distributed, the estimators for the state vectors (as each measurement becomes available), (1.10) and (1.10), are optimal, in the sense that they yield the minimum MSE. This optimality still holds when dropping the normality assumption yet it is restricted to the class of linear estimators.

Motivation is what gets you started.
Habit is what keeps you going.
Jim Rohn.

Chapter 2

The Real-State MARBP

In this chapter we survey key theoretical aspects of the MARBP, with special emphasis on its real-state variant, highlighting the mathematical interest of the research challenges it raises and on its diverse variety of possible applications. The main goal of such a summary of ideas and methods is to provide the reader with a clear perspective from which to assess the contribution of the present work as well as its motivation. The summary is completed with a brief account of the historical background and an overview of the most influential previous research effort on MARBP. The chapter concludes with a concise description of the specific challenges posed by the real-state MARBP applications investigated in this dissertation.

2.1 The MARBP

The MARBP represents in a simplified way the overarching concern of how to best allocate scarce resources over time under uncertainty. It can be simply formulated as follows. Imagine a manager who must decide over some infinite horizon of discrete time slots $t = 0, 1, \ldots$ how to best allocate a fixed (and limited) endowment of resources $\bar{W}$ to a finite number of binary-action (*active/passive*) stochastic projects labeled by $n = 1, \ldots, N$. Specifically, assume that the resource scarcity forces the manager to choose at the start of each period a subset of those projects worth (at most) $\bar{W}$ to form the realized portfolio. Each project yield rewards in time, depending both on the manager's action and on the project's state. The project's state lives on a state space $\mathbb{X}$ and evolves randomly over it according to an *active/passive transition law*, based on the manager's action. In some cases, it may be more convenient to consider that projects incur costs instead of rewards when active. In the remainder of this chapter we focus on the case of reward yielding projects for ease of presentation and interpretation of the reviewed concepts. Further, throughout the remainder of this work we shall stick to the following notational conventions: uppercase letters denote random variables (X, Y) while a realization of that random variable will be denoted with the corresponding lowercase (x, y).

In the MARBP, every project is modeled as a discrete-time Markov Control Process (MCP) whose defining elements are described in detail in the following subsection.

2.1.1 The One-Arm Bandit Control Model

The discrete-time MCP representing project n's decision problem is given by the five-tuple:

$$(\mathbb{X}_n, A_n, P_n(.|x_n, a_n), R_n(x_n, a_n), W_n(x_n, a_n)), \tag{2.1}$$

consisting of

- The *state space* $\mathbb{X}_n$. In the most general setting, $\mathbb{X}_n$ may be a Borel space, although most frequent MARBP applications investigated on the recent literature have focused on the case in which $\mathbb{X}_n$ is a finite set (or an infinite denumerable set) of possible states which project n may occupy. Instead, the problems addressed in this dissertation have the distinguishing feature of dealing with a state variable X_n that lives in a closed interval (possibly unbounded) of the real line $\mathbb{R}$, i.e. $\mathbb{X}_n \subseteq \mathbb{R}$. Thus, X_n for such real-state MARBP applications admits infinite possible values.

- The *action set* A_n. For general Markov control models, the action set consists of a Borel space, yet for the standard MABP the action set is a binary set representing the work/rest of projects. i.e. $A_n \triangleq \{0, 1\}$, with $a_n = 1$: *active*; $a_n = 0$: *passive*. We denote by $A_n(x_n)$ the set of feasible actions at some $x_n \in \mathbb{X}_n$. Notice that in the case that there exists some state $x_n^u \in \mathbb{X}_n$ for which it either holds that $A_n(x_n^u) = \{0\}$ or that $A_n(x_n^u) = \{1\}$, then x_n^u is an *uncontrollable* state. Hence, if there exists at least one uncontrollable state, we define the set of controllable states, i.e., those states for which both actions are feasible, as follows

$$\mathbb{X}_n^{0,1} \triangleq \{x_n \in \mathbb{X} : A_n(x_n) = \{0, 1\}\}$$

- A Markovian *transition law* $P_n(.|x_n, a_n)$, describing the evolution of the state variable $X_{n,t}$ given $x_{n,t-1}$ and $a_{n,t-1}$. Thus, the state variable of the MCP is a $\mathbb{X}_n$-valued random variable taking values in $\mathbb{X}_n$ according to some probability kernel

$$P_n(B|x_n, a_n) \triangleq \mathcal{P}_n\{X_{n,t} \in B | x_{n,t-1} = x_n, a_{n,t-1} = a_n\}, \text{ for } B \subset \mathscr{B}(\mathbb{X}_n),$$

where $\mathscr{B}(\mathbb{X}_n)$ denotes the Borel subsets of $\mathbb{X}_n$.

Notice that the above dynamics definition allows for changes in the state of passive projects. When the MABP incorporates such a feature, the model is called *restless*. In the case where passive projects remain frozen in their current states, then the MABP is said to be *classic*, and $P_n(.|x_n, a_n)$ is such that:

$$P_n(B|x_n, 0) \triangleq \mathcal{P}_n\{X_{n,t} \in B|x_{n,t-1} = x_n, a_{n,t-1} = a_n\} = 1_{\{x_{n,t-1} \in B\}}, \text{ for } B \subset \mathscr{B}(\mathbb{X}_n)\,.$$

A fundamental feature of both types of MABP is the fact that the state transitions across projects are assumed to be independent. Such a feature is key to the Lagrangian relaxation and decomposition indexation approach that will be reviewed in the following chapter.

- *One-period expected reward* $R_n(x_n, a_n)$ and *work* (i.e. expected resource consumption) $W_n(x_n, a_n)$ functions respectively giving the one-period expected rewards and expected work when project n occupies some state $x_n \in \mathbb{X}_n$ and it is operated under action $a_n \in \{0, 1\}$. In the cases in which it is most natural to consider cost functions, we shall denote them as $C_n(x_n, a_n)$.
  Notice that given the Markovian transition law assumed, the resulting MCP has the property that, at any given time slot, reward and work transitions depend only on the current state of the project and on the selected action.

  Regarding these one-period work functions it is assumed that:

  (i) Resource consumption when active is non-negative, i.e. $W_n(x_n, 1) \geq 0$;

  (ii) Resource consumption when active is at least as large as when passive, which is also non-negative, i.e. $W_n(x_n, 1) \geq W_n(x_n, 0) \geq 0$.

  (iii) Idling of all projects is a feasible action, i.e. $\sum_{n=1}^{N} W_n(x_n, 0) \leq \bar{W}$;

  In both applications investigated in this thesis we let $W_n(x_n, a_n) \triangleq a_n$ as in Whittle (1988) and thus $\bar{W}$ is an integer between 1 and N. Therefore, verification of the above stated assumptions is straightforward. Henceforth we focus our discussion for the case $W_n(x_n, a_n) \triangleq a_n$.

2.1.2 The Multi-Armed Bandit Control Policies

Decisions on which projects to work on at each time slot t, if any, are based on a *control (scheduling) policy* π which defines a feasible sequence of actions $\{a_{n,t}\}$ for each project

n and at every state and period. In words, a policy is a rule that specifies how to act at each time for every possible state of the projects given the available limited resources. Feasible policies are drawn from the set of *history-dependent randomized policies* Π, which for the applications considered in this dissertation reduces to

$$\sum_{n=1}^{N} a_{n,t} \leq M, \quad t \geq 0 \tag{2.2}$$

We shall denote such a class of policies as $\Pi(M)$. Notice that within the class $\Pi(M)$ we can further define restricted sets of policies, such as *deterministic* policies, or *Markov* or *stationary* policies, both within the sets of *randomized* and *deterministic* policies. To clarify this idea, we review below some of these definitions.

Consider the space $H_{n,t}$ of all admissible t-histories (i.e. histories up to time t) for the above described project n MCP, where a t-history is a vector of the form:

$$h_t = (x_{n,0}, a_{n,0}, \ldots, x_{n,t-1}, a_{n,t-1}, x_{n,t})$$

Definition 2.1. A *history dependent randomized policy* is a set of functions $\pi_{n,t}$ which map any possible history $h_{n,t}$ to a probability distribution, denoted as $\gamma_{n,t}$, on A_n, from which the manager will draw a random action $a_{n,t}$, i.e. $\{\pi^R_{n,t} : H_{n,t} \rightarrow \gamma_{n,t}(A_n), t \geq 0\}$.

Definition 2.2. A *history dependent deterministic policy* is a set of functions $\pi_{n,t}$ which map any possible history $h_{n,t}$ to an action in A_n, i.e. $\{\pi^D_{n,t} : H_{n,t} \rightarrow A_n, t \geq 0\}$. Notice that deterministic policies correspond to the subset of randomized policies in which $\gamma_{n,t}$ has probability mass 1 concentrated on the corresponding action for any t and $h_{n,t}$

Definition 2.3. A *Markovian policy* is a set of functions $\pi_{n,t}$ which for all t depend only on the current state $x_{n,t}$ instead of the whole t-history h_t, i.e. $\{\pi^{RM}_{n,t} : x_{n,t} \rightarrow \gamma_{n,t}(A_n), t \geq 0\}$, if *randomized*, or $\{\pi^{DM}_{n,t} : x_{n,t} \rightarrow A_n, t \geq 0\}$, if *deterministic*.

Definition 2.4. A *stationary policy* is a set of functions $\pi_{n,t}$ which map any possible state either to an action or to a probability distribution on the action space, regardless of t, i.e.$\{\pi^{RS}_n : x_n \rightarrow \gamma_n(A_n), t \geq 0\}$, if *randomized* or $\{\pi^{DS}_n : x_n \rightarrow A_n, t \geq 0\}$, if *deterministic*.

For a rigorous presentation of these definitions see Hernández-Lerma and Lasserre (1996, Chapter 1). Respectively, denote by $\Pi^R(M)$, $\Pi^D(M)$, $\Pi^{RM}(M)$, $\Pi^{RS}(M)$, $\Pi^{DM}(M)$, $\Pi^{DS}(M)$ to the sets of all randomized, deterministic, randomized Markovian, randomized stationary, deterministic Markovian and deterministic stationary policies that satisfy the hard sample-path resource constraint (2.2). It is important to note that the following relationship holds:

$$\Pi^{DS}(M) \subset \Pi^{DM}(M) \subset \Pi^{D}(M) \subset \Pi(M) \tag{2.3}$$

With this fact in mind, it is worth pointing out here that the relevance of studying restricted families of polices is explained both by theoretical and computational reasons as when optimal policies can be guaranteed to exist within these reduced classes, implementation and interpretation becomes significantly simpler. In fact, this is a central topic in Markov control theory among which the restless bandit indexation approach researched and deployed in this thesis may be included.

2.1.3 The Performance Measures

To complete the specification of the real-state MARBP, in addition to the projects' dynamical systems (given by the previously described MCP), we must define a *performance measure* on the set of feasible control policies $\Pi(M)$ upon which the portfolio's response to the selected policies will be evaluated and allocation decisions will be made. Once we have included this final element, the MARBP optimal control problem is to find a feasible policy in $\Pi(M)$ that optimizes the selected performance measure.

Naturally, in the setting of a optimization problem we would like to choose a performance measure that in some way *maximizes* the total investment rewards over the manager's operating horizon. Thus, we may want to consider the optimal control problem of finding an expected total-optimal policy, which maximizes

$$\max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}(\boldsymbol{M})} \mathsf{E}_{\mathbf{x_0}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \sum_{n \in \mathsf{N}} R_n \left(X_{n,t}, a_{n,t} \right) \right] , \tag{2.4}$$

where $\mathbf{x_0} = (x_{n,0})_{n=1}^{N}$ is the initial joint belief state, and $\mathsf{E}_{\mathbf{x_0}}^{\boldsymbol{\pi}}[\cdot]$ denotes expectation under policy $\boldsymbol{\pi}$ conditional on the initial joint state being equal to $\mathbf{x_0}$ (for any possible joint initial state). We denote by $V_T^*(\mathbf{x_0})$ to the optimal total value function.

Yet, when considering infinite horizon problems, which is usually an adequate framework for many problems in which there is no natural stopping time specified a priori or there is just number of decision stages is really infinite, or at least a large number of decision periods, summing up the overall flow of rewards yielded may not converge, at least under some policies.[1] For this technical reason (or even just for the sake of the interpretation of the performance objective) it may be more convenient to use other

[1] Consider for instance the case in which resting all projects generates a nonnegative constant reward at any possible state x_t, then the policy of $a_t = 0$ for all t, then ET objective function does not converge.

performance measures for some infinite horizon problems rather than the Expected Total (ET) measures.

The applications researched in this thesis inherently call for infinite horizon formulations. Thus, we have considered two widely used performance measures for such problems: the Expected Total Discounted (ETD) rewards and the Long Run Average (LRA) rewards, for which the optimal control problems can be respectively defined as follows:

(1) find a discount-optimal policy,

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(\boldsymbol{M})} \mathsf{E}^{\pi}_{\mathbf{x_0}} \left[\sum_{t=0}^{\infty} \sum_{n\in\mathsf{N}} \beta^t R_n\big(X_{n,t}, a_{n,t}\big) \right], \tag{2.5}$$

where $0 < \beta < 1$ is the discount factor.

(2) find an average-optimal policy,

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(\boldsymbol{M})} \liminf_{T\to\infty} \frac{1}{T} \mathsf{E}^{\pi}_{\mathbf{x_0}} \left[\sum_{t=0}^{T-1} \sum_{n\in\mathsf{N}} R_n\big(X_{n,t}, a_{n,t}\big) \right], \tag{2.6}$$

focusing on the ETD problem (2.5). We respectively denote by $V^*_D(\mathbf{x_0})$ and $V^*_A(\mathbf{x_0})$ to the optimal discounted value function and the optimal average value function. For a detailed discussion of these performance optimization criteria see e.g., Hernández-Lerma and Lasserre (1999).

2.1.4 The Optimality Equation

For a given performance measure, the resulting MARBP can be analyzed using the ideas introduced in Bellman (1957). Consider for instance the β-discounted control problem (2.5). Its solution satisfies the following Dynamic Programming Equations (DPE)

$$V^*_D(\mathbf{x_0}) = \max_{a\in\{0,1\}} \left[R(\mathbf{x_0}, a) + \beta \int_{\mathbb{X}} V^*_D(y) P(dy|\mathbf{x_0}, a) \right], \forall \mathbf{x_0} \in \mathbb{X} \tag{2.7}$$

Under appropriate conditions on the one-stage reward (cost) and work functions and on the transition law P (see, e.g., Hernández-Lerma and Lasserre, 1996, Assumption 4.2.1 & 4.2.2) it can be shown that there exists an optimal β-discounted policy for (2.7) which is further *deterministic* and *stationary*. A similar, though somehow more technical analysis, can be performed for the long-run average control problem (2.6) to study the conditions under which the existence of *deterministic stationary* LRA policies is guaranteed.

Except for rare special cases in which the solution can be analytically derived in closed form, the most frequently used method for finding the optimal policy of such decision problems is the application of iterative procedures or algorithms. These traditional solution techniques include: *value iteration* (Bellman, 1957) and *policy iteration* (Howard, 1960), as well as its variation called *modified policy iteration* (Puterman, 1994). The Markovian property lying at the heart of these approaches allows for the reduction of the original problem's complexity by breaking it down into simpler subproblems at various moments of time.

Yet, this reduction may not be enough to ensure tractability of the resulting optimal policies. The burden of these algorithms lies within the cardinality of the state space where $\mathbf{x_0}$ lives, since its size determines the computation and storage requirements for solving (2.7). Further, even for problems in which the cardinality of $\mathbb{X}_n$ is finite and relatively small, the number of Dynamic Programming (DP) equations in (2.7) grows exponentially in the number of projects, severely hindering the conventional numeric DP approach. Indeed, the special case of MARBP in which a finite state space is considered, the transition laws are deterministic, $W_n(x_n, a_n) \triangleq a_n$ and $\bar{W} \triangleq M = 1$, has been shown by Papadimitriou and Tsitsiklis (1994) to be PSPACE-hard (i.e. computationally intractable) despite the deterministic state dynamics assumption.

This well known *curse of dimensionality* affects also alternative solution approaches, such as the mathematical programming technique based on solving a *linear programming* reformulation of the Bellman equations. Although this solution strategy exhibits an advantage over traditional dynamic programming algorithms to find the solution of constrained MDP, as it successfully exploits the reduction on the set of feasible policies imposed by the extra constraints, it still suffers from computational intractability for general MDP models. For a review of some of the most important results concerning this technique see e.g., Heilmann (1978).

In conclusion, despite optimal policies for these control problems are known to exist, their applicability to a great deal of relevant problems is severely hindered for realistic scenarios due to computational or technical reasons. The exact numerical solution to their corresponding MARBP formulations is usually unavailable not only because the DPE formulation is quickly rendered intractable but, most importantly to this dissertation, because considering a real-state space introduces special difficulties. This fact explains that when forced to dealing with practical applications, as the ones addressed in this thesis, the necessity for implementable approaches becomes excruciating. As well, such a fact highlights the relevance of relationship (2.3), since as long as conditions which ensure that a *deterministic stationary* solution exists hold, it is is sufficient to search for the optimal policy within such a reduced set of policies. This simple idea, motivates the research on alternative solution approaches based on establishing conditions for the existence of optimal policies within reduced classes of policies, so that by restricting attention e.g. to $\Pi^{DS}(M)$ (or even subsets of $\Pi^{DS}(M)$) both computational and analytical advantages are exploited. As it will be discussed in the following chapter, the indexation methodology deployed in this dissertation constitute an alternative approach to traditional dynamic programming techniques based on this key idea.

To illustrate the ideas on real-state MARBP reviewed up to this point, we propose the example below.

Example 2.1 (Treasure Hunting Problem: MARBP formulation)**.** (Bertsekas, 2007, p. 70) Consider N boxes, each of which may contain a *hidden and unmoving treasure* in it, and a searcher that may search one box at every time period. Let α_n be the probability that a search in box n finds a treasure provided that it is hidden at that box. Searching box n costs c_n and it produces two possible outcomes: either the treasure is found, yielding reward r_n to the searcher, or information is gained on the likelihood that the box contains a treasure. Hence, the probability that a treasure lies within each box changes by Bayes' theorem as boxes are successively searched. A natural question in this context is how to schedule the boxes for being searched so that the expected net reward of finding the objet is maximized?

Notice that after finding a treasure at box n, it does make sense to continue searching it, thus the project of searching it concludes after a random number of searches. Following Bertsekas, we model this by letting the probability state drop to zero after finding a treasure at site n. [2] To formulate this simple problem into a MARBP framework, we start by defining a generic project as searching box n with the following MCP representing its corresponding decision problem.

- The *state space* $\mathbb{X}_n \triangleq [0,1]$: the state is the probability that the object is hidden at box n at time t.
- The *action set* $A_n \triangleq \{0,1\}$, with $a_n = 1$: *Search box n*; $a_n = 0$: *Do not Search box n*.
- The Markovian *transition law* $P_n(.|x_n, a_n)$ given by

$$\text{if } a_{n,t} = 0, \qquad x_{n,(t+1)} = x_{n,t} \text{ w.p. } 1,$$

$$\text{if } a_{n,t} = 1, \qquad x_{n,(t+1)} = \begin{cases} \frac{\alpha_n x_{n,t}}{\alpha_n x_{n,t} + (1 - x_{n,t})} & \text{w.p. } \alpha_n x_{n,t} + (1 - x_{n,t}) \\ \\ x_{n,(t+1)} = 0 & \text{w.p. } (1 - \alpha_n) x_{n,t} \end{cases}$$

[2] The goal of this convention is to model that the project reaches an uncontrollable state once the treasure has been found in which the only possible action is not to search for that treasure.

- *One-period expected cost* and *work* (i.e. expected resource consumption) functions respectively given by

$$R_n(x_n, a_n) \triangleq (r_n(1-\alpha_n)x_n - c_n)\, a_n \quad \text{and} \quad W_n(x_n, a_n) \triangleq a_n$$

Notice that the above formulation is *classic*, in the sense that passive projects remain frozen.

• • •

2.1.5 POMDP and the MARBP

In general, the MDP model described in subsection 2.1.1 constitutes a widely-used framework for modeling decision making in complex stochastic dynamical systems. Yet, in such a theoretical context the unique source of uncertainty in the model proceeds from state transitions from one state value to another. Unfortunately, in many practical applications we cannot rely on having an exact observation of the state of the process to base our decisions and we are thus forced to estimate it as precisely as possible given some observational data.

This is the case of most *sensor management* problems, in which the state of the controlled process (e.g. a target's position or target's type) is only partially observed due to measurement errors or clutter degradations. More examples of this nature arise in *artificial intelligence* or *automated planning* applications (see e.g. Kaelbling et al., 1998). All these applied problems illustrate a central characteristic of the general optimal control problem for POMDPs: optimal resource allocation must be done while simultaneously estimating optimally the unobservable state of the system given the error-prone observations.

Formally, the discrete time MCP associated with project n's POMDP is given by the tuple:

$$(\mathbb{X}_n, A_n, P_n(.|x_n, a_n), \mathbb{Y}_n, \Omega_n(.|x_n, a_n), R_n(x_n, a_n), W_n(x_n, a_n)), \tag{2.8}$$

where $\mathbb{X}_n, A_n, P_n(.|x_n, a_n), R_n(x_n, a_n), W_n(x_n, a_n)$ are defined as in subsection 2.1.1 except for the fact that we focus on the case in which the cardinality $\mathbb{X}_n$ is finite and thus $P_n(.|x_n, a_n)$ stands for transition matrices. The novel elements on the tuple, namely $\mathbb{Y}_n$ and $\Omega_n(.|x_n, a_n)$, respectively stand for the set of observations and a probability law describing the probability that we observe $Y_{n,t} = y_{n,t}$ given $x_{n,t}$ and $a_{n,t}$.

Although the state of the process is not directly observed, a probability distribution over the states $b_n(x_n)$ can be maintained giving the probability or *belief*, that the unobservable process is in state x_n. Since x_n is Markovian, keeping such beliefs over the states is done in the following way: if action $a_{n,t}$ is taken and it is followed by observation $y_{n,t}$, the next belief state, denoted as $b_n(x_{n,(t+1)})$, is determined by updating each state probability using Bayes' theorem, as follows,

$$b_n(x_{n,(t+1)}) = \frac{\Omega(y_{n,(t)}|x_{n,(t+1)}, a) \sum_{s\in\mathbb{X}_n} P(x_{n,(t+1)}|s, a_{n,t})b(s)}{\sum_{s,s'\in\mathbb{X}_n} \Omega(y_{n,t}|s', a_{n,t})P(s', s, a_{n,t})b(s)} \tag{2.9}$$

Since a feasible policy for (2.8) maps any possible belief state to the action space, a convenient way to analyze such a problem is to reformulate the POMDP as a fully observable equivalent MDP with a real-state variable. The resulting MDP will be defined by the tuple

$$(\mathbb{B}_n, A_n, T_n(.|x_n, a_n), R_n(b_n, a_n), W_n(b_n, a_n)), \tag{2.10}$$

where $\mathbb{B}_n, \subseteq [0, 1]$ is the set of belief states over the original POMDP states, $T_n(.|x_n, a_n)$ is the belief state update function resulting from the Bayes' update rule, one-period reward and cost functions are now defined over the belief state set, and A_n is the same as in (2.8).

In conclusion, POMDPs often suit better than MDPs for many relevant applied problems yet, their optimal strategies are in general intractable, as its state space $\mathbb{B}_n$ is infinite. In practice, as exact optimal solutions are not derivable analytically, approximate solution methods (based discretizations) are the most frequently deployed solution techniques. Nontheless, a realistic state space discretization is unlikely to result implementable since POMDPs are PSPACE-complete problems.

As already announced in section 1.3 of the previous chapter, this chapter explains why the relevance of successfully solving the real-state MARBPs, as the ones addressed in this dissertation, goes beyond the scope of the concrete applications of concern. Specifically, as the POMDP described in (2.8) has a special structure, usually fitting into the framework of the continuous-state MABP (either in its classic version or, more often, in its restless variant), the effective design of solution strategies for real-state MARBP offers the potential impact of achieving simultaneously tractability and performance improvements for a large class of POMDPs.

2.2 Index Policies

All the appealing theoretical features of the MDP framework are clearly obscured by the lack of applicability of the resulting optimal polices to optimize the performance of modern technological systems. Such a state of affairs is the main motivation for investigating the design of *heuristic* policies which achieve tractability, possibly at the expense of optimality, but which nonetheless manage to achieve a preestablished performance objective. Within the astoundingly large family of possible heuristics, a class stands out as sound and natural for these allocation problems: the *priority index* polices.

A *priority index* policy assigns a value to each project as a function of its state, where such value *prioritizes* its access to the scarce resources in the following sense: projects at each time slot are incorporated into the realized portfolio ordered with respect to their index values as long as the resulting portfolio remains affordable, given the resource endowment. Hence, a *priority index* policy activates a number of projects, feasible in terms of the hard sample-path constraint, whose index is currently the largest.

Formally, an index rule of priority-index type requires a function $\lambda_n(.)$ which, for every project n, maps its state space to the set of real numbers $\mathbb{R}$, i.e.

$$\lambda_n : \mathbb{X}_n^{0,1} \mapsto \mathbb{R} \tag{2.11}$$

One of the simplest example, and also one of the most widely used heuristics of priority-index type, is the myopic index policy (sometimes also called the *greedy* heuristic in some literature) which defines $\lambda_n(.)$ as follows

$$\lambda_n^{Myopic}(x) \triangleq R(x,1), \quad \forall x \in \mathbb{X}_n^{0,1} \tag{2.12}$$

Notice that the myopic policy is equivalent to setting $\beta = 0$ in (2.7) for a given project n and selecting as its current index value the best possible next-step reward. Obviously, the computational feasibility of such an index is attained at the expense of ignoring the future consequences of today's selected actions.

Another simple example of a priority-index policy, which is also relevant for this thesis, is to consider as a priority index value the current state of the project, i.e.

$$\lambda_n(x) \triangleq x, \quad \forall x \in \mathbb{X}_n^{0,1} \tag{2.13}$$

In the POMDP context, this heuristic is simply to use the belief state of the projets as an index. In the multitarget tracking setting this heuristic corresponds to considering the tracking error variance for each target as the priority index, as it was proposed by Howard et al. (2004); La Scala and Moran (2006).

As we will illustrate in the parts II and III of this dissertation, there are situations in which we can expect these two simple heuristics to perform equally well and even to be as good as the index policies derived from the indexation methodology deployed to the applications of concern. However, in realistic settings these two heuristics achieve different performance results and moreover they are normally significantly outperformed by the index rules proposed by this thesis.

In the following chapter we will review the background material for designing mathematically-based priority-index polices based on Lagrangian optimization methods. To introduce and motivate that background, we finish this chapter with a section providing a historical account of the development of central ideas of the indexation framework that shall be deployed throughout chapters Chapter 4 and Chapter 6.

2.2.1 Overview of Historical Development

The persistence in time of this idea of addressing the computational challenges by designing heuristics of priority-index type can be mostly attributed to early results on the optimality of this sort of rules. Such results prompted researchers to take advantage of the special structure of specific problems to provide with a solution method which offered not only tractability but also economic insights and insightful interpretations of the underlying system. In such a vein, a classical result is given by the optimality of the $c\mu$-rule for the problem of optimally sequencing a batch of jobs with mean processing time $(1/\mu)$ and linear holding costs (c) in the single-machine case (Smith, 1956). The index rule in this case prescribes to schedule the job yielding the highest expected cost reduction per unit of effort, which can also be interpreted as the job achieving the maximum average productivity rate of the machine.

Regarding the origins of bandit indexation, Bradt et al. (1956) first showed the optimality of an index policy for the classic finite-horizon undiscounted *one-armed* Bernoulli bandit problem. The index was defined as in (2.11) but taking as an argument an augmented state including both the current project state and the number of remaining periods. Bellman (1956) in turn extended such a result to the infinite-horizon problem, establishing the existence of an optimal policy of index type, in which the index is a function of the state only. Yet, efforts to apply these indexation ideas to the *multi-armed* bandit case, still in its *classic* version, were long deemed sterile as it was not until Gittins and Jones (1974) that its optimal solution was first shown to be attained by a priority-index rule, which has become known as Gittins' index policy.

The *classical* MABP, was formulated during the Second World War, but its roots can be traced back to the early thirties in the seminal work of Thompson (1933). Thompson presented a characteristic dilemma arising in the area of sequential design of experiments in terms of the problem of deciding how to assign patients to two distinct clinical treatments. The difficulty of the decision, as he pointed out, lies within the fact that giving a patient a treatment that currently appears to be inferior instead of the treatment with the highest estimated probability of success, can eventually lead to obtain a more accurate belief estimate which might show that the apparently inferior treatment is actually the best one. In short, the dilemma is to assign the *present* patient a treatment that is most likely to succeed based on information gathered so far (*exploitation*) or to assign a treatment with less chances of being efficient but that might eventually turn out to outperform the other, thus allowing for a higher rate of *future* successes (*exploration*). This idea was followed later by Robbins (1952) who posed a family of optimal sequential estimation problems, in which optimization and information estimation ar simultaneously.

The optimality of the Gittins' index rule led the way to substantial generalizations of this optimality result, however, in Whittle's words, "one class seemed to remain unnameable": those in which projects continue to evolve even when rested. This particular class, corresponding to the MARBP, first announced by Whittle (1988) as an extension to the *classical* MABP, offered increased modeling power as well as many interesting and novel mathematical challenges to researchers. In models in which the project's state is an information state, as is the case of the applications studied in this thesis, it is adequate to assume that information is gained (or lost) when we work (or rest) on the project. Actually, Whittle illustrated this point with a multitarget tracking example, in which, as he put it, "the bandits are *restless* in the most literal sense".

Whittle proposed a Lagrangian relaxation and decomposition approach to develop de index heuristic. Yet, he also realized that existence of such index for MARBP was only ensured for those which satisfied a structural property, which he termed *indexability*. Those circumstances called for general sufficient indexability conditions, to allow for exploitation of the enhanced modeling power. Such conditions were successfully provided through work done for discrete-state bandits in Niño-Mora (2001, 2002, 2006b) by deploying an achievable region approach, based on the introduction of Partial Conservation Laws (PCLs) (See the review Niño-Mora, 2011a). Moreover, those results include an adaptive-greedy algorithm for indexability verification and index computation. As reviewed in Niño-Mora (2007a), the resulting approach has been generalized into a widely applicable unifying framework to design index policies which admit economic interpretation and which have been found to exhibit near-optimal results. Of special relevance for this dissertation are the extensions of those sufficient indexability conditions and other results to the continuous-state case introduced in Niño-Mora (2008).

For brief and basic introduction to MARBPs and also for a more detailed historical account of development of this emerging research area see e.g., Niño-Mora (2011b).

2.3 Research Challenges & Applications

In this section, in light of the previous reviewed concepts, we reexamine and summarize the main challenges for solving the applications proposed in problem 1 and problem 2 (as stated in Chapter 1) by formulating them as real-state MARBPs and solving these by means of a priority index policy.

In modern sensing systems, depending on the system's mission, we can associate to its operating performance some statical model capturing the unobservable process evolution and relating sensor output to the state of that process. Next, using sensor observations an associated optimal sequential estimation problem is to predict the state of the underlying physical process according to some sensible criteria (e.g. mean square error). In turn, sensor observations should result from solving an optimal resource allocation problem which distributes sensing resources among projects so as to best extract the required information.

Hence, both for the detection and tracking problems, the underlying optimal sequential estimation problem can be easily put into a MABP framework, in which the sensing decisions are made incrementally as additional information is received. This step will be respectively done in Chapter 4 and Chapter 6 of this thesis. However, after doing so we will be faced with the following general issues:

(i) The most adequate MABP formulation for both problems will be *restless*, since predicting a project's state will be generally different depending on the information provided by sensing actions.

(ii) The state variable (capturing project's information state) will naturally belong to a continuous state-space.

The first issue will subsequently lead us to the dealing with the challenging questions of establishing under which conditions the index exists for the problems at hand, how to compute it in a reasonable time, and since optimality of the index policy is not ensured for the restless case, how far is the resulting heuristic from the optimal. The second issue will be introducing extra technical difficulties that must be successfully addressed to answer all the previous questions.

Action expresses priorities.
Gandhi

Chapter 3

Real-State RB Index Policies: Lagrangian Relaxation and Decomposition approach

3.1 Whittle's Relaxed Problem: Lagrangian relaxation and Performance Bound

As reviewed in the previous chapter, Whittle (1988) was the first to extend the scope of the indexation approach beyond the framework of *classic* bandits by proposing a relaxation of the original problem's hard sample path constraint for the equality-constrained case. The relaxed problem would thus have an enlarged family of feasible policies in which the resource constraint, instead of being fulfilled at each period of time, is fulfilled in expectation in terms of the selected performance measure over the whole operating horizon. For the ETD problem (2.5), the Whittle relaxation, which was originally defined for the case $W_n(x_n, a_n) = a_n$ and $\bar{W} = M$, can be extended as (2.5)-(3.1) with

$$\max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}(\boldsymbol{M})} \mathsf{E}_{\mathbf{x_0}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \sum_{n \in \mathsf{N}} \beta^t R_n \big(X_{n,t}, a_{n,t} \big) \right], \qquad (2.5)$$

$$\mathsf{E}_{\mathbf{x_0}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \sum_{n=1}^{N} \beta^t a_{n,t} \right] \leq \mathsf{E}_{\mathbf{x_0}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \beta^t M \right] = \frac{M}{1-\beta}, \qquad (3.1)$$

where (3.1) only requires that the expected total β-discounted resource consumption does not exceed the total β-discounted resource availability endowment. Respectively, for the LRA problem, the Whittle relaxation is (2.6)-(3.2) with

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(M)} \liminf_{T\to\infty} \frac{1}{T}\mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{T-1}\sum_{n\in\mathsf{N}} R_n\big(X_{n,t},a_{n,t}\big)\right], \qquad (2.6)$$

$$\limsup_{T\to\infty} \frac{1}{T}\mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{T-1}\sum_{n=1}^{N} a_{n,t}\right] \le M. \tag{3.2}$$

Whittle's *relaxed primal ETD problem* is

$$V_D^{\mathrm{R}}(\mathbf{x_0}) \triangleq \max_{(3.1),\boldsymbol{\pi}\in\Pi} \mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{\infty}\sum_{n=1}^{N}\beta^t R_n\big(X_{n,t},a_{n,t}\big)\right]. \tag{3.3}$$

where Π is the class of history dependent randomized scheduling policies (which may engage in any number of projects at any time), and under the long-run average criterion we obtain Whittle's *relaxed primal LRA problem*

$$V_A^{\mathrm{R}}(\mathbf{x_0}) \triangleq \max_{(3.2),\boldsymbol{\pi}\in\Pi} \liminf_{T\to\infty} \frac{1}{T}\mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{T-1}\sum_{n=1}^{N} R_n\big(X_{n,t},a_{n,t}\big)\right]. \tag{3.4}$$

Note that the optimal values of (3.3) gives an *upper bound* on the optimal value of the original problem (2.5), i.e. $V_D^{\mathrm{R}}(\mathbf{x_0}) \ge V_D^*(\mathbf{x_0})$. Respectively, (3.4) gives an *upper bound* on the optimal value of (2.6), i.e., $V_A^{\mathrm{R}}(\mathbf{x_0}) \ge V_A^*(\mathbf{x_0})$.

To address the constrained MDPs defined by the relaxed problems we next deploy a Lagrangian approach, including as a coupling constraint the relaxed resource constraint by attaching a Lagrange multiplier $\lambda \ge 0$ to it. The resulting unconstrained MDPs are

$$V_D^{\mathrm{L}}(\mathbf{x_0};\lambda) \triangleq \max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{\infty}\sum_{n=1}^{N}\beta^t\left\{R_n\big(X_{n,t},a_{n,t}\big)-\lambda a_{n,t}\right\}\right]+\lambda\frac{M}{1-\beta}, \tag{3.5}$$

and

$$V_A^{\mathrm{L}}(\mathbf{x_0};\lambda) \triangleq \max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \liminf_{T\to\infty} \frac{1}{T}\mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x_0}}\left[\sum_{t=0}^{T-1}\sum_{n=1}^{N}\left\{R_n\big(X_{n,t},a_{n,t}\big)-\lambda a_{n,t}\right\}\right]+\lambda M \tag{3.6}$$

For any arbitrary nonnegative value of the multiplier λ, the optimal values of (3.5) and (3.6) respectively give an *upper bound* on the optimal values of (3.3) and (3.4), i.e. $V_D^{\mathrm{L}}(\mathbf{x_0};\lambda) \ge V_D^{\mathrm{R}}(\mathbf{x_0})$ and $V_A^{\mathrm{L}}(\mathbf{x_0};\lambda) \ge V_A^{\mathrm{R}}(\mathbf{x_0})$. Notice that a policy solving either (3.5) or (3.6) for a given $\lambda \ge 0$, performs at least as well as the optimal policy for the original MABP, yet it is important to bear in mind that they will not typically be feasible for that problem, since they may not satisfy the resource sample path constraint. [1]

The *Lagrangian dual problem* is to find an optimal value of the multiplier $\lambda^*(\mathbf{x_0})$ giving the best upper bound on $V_D^{\mathrm{R}}(\mathbf{x_0})$ or $V_A^{\mathrm{R}}(\mathbf{x_0})$, which we denote by $V_D^{\mathrm{D}}(\mathbf{x_0})$ or $V_A^{\mathrm{D}}(\mathbf{x_0})$. If such a $\lambda^*(\mathbf{x_0})$ exists, it solves the following scalar optimization problem for the ETD problem

$$V_D^{\mathrm{d}}(\mathbf{x_0}) = \min_{\lambda \geq 0} V_D^{\mathrm{L}}(\mathbf{x_0}; \lambda), \tag{3.7}$$

and for the LRA problem it solves

$$V_A^{\mathrm{d}}(\mathbf{x_0}) = \min_{\lambda \geq 0} V_A^{\mathrm{L}}(\mathbf{x_0}; \lambda). \tag{3.8}$$

Note that (3.7) and (3.8) are convex optimization problems, since $\lambda \mapsto V_D^{\mathrm{L}}(\mathbf{x_0}; \lambda)$ and $\lambda \mapsto V_A^{\mathrm{L}}(\mathbf{x_0}; \lambda)$ are convex. Under suitable regularity conditions, $\lambda^*(\mathbf{x_0})$ exists and strong duality holds. Notice that although *weak duality* ($V^{\mathrm{R}}(\mathbf{x_0}) \geq V^{\mathrm{d}}(\mathbf{x_0})$) is ensured, satisfaction of *strong duality*, i.e. $V^{\mathrm{R}}(\mathbf{x_0}) = V^{\mathrm{d}}(\mathbf{x_0})$, calls for further investigation.

3.2 Indexability and the Whittle Index Policy

Next, to introduce the notion of indexability, we will make use of the key assumption that projects' state variables evolve independently from one another. This allows us to *decompose* the problems (3.5) and (3.6) into N independent parts, denoted as $V_{D,n}^{\mathrm{L}}(x_{n,0}; \lambda)$ and $V_{A,n}^{\mathrm{L}}(x_{n,0}; \lambda)$, each representing a single-project subproblem, consisting of the following ETD problem considered for some project n in isolation,

$$V_{D,n}^{\mathrm{L}}(x_{n,0}; \lambda) \triangleq \max_{\boldsymbol{\pi_n} \in \boldsymbol{\Pi_n}} \mathsf{E}_{x_{n,0}}^{\boldsymbol{\pi_n}} \left[\sum_{t=0}^{\infty} \beta^t \{ R_n\big(X_{n,t}, a_{n,t}\big) - \lambda a_{n,t} \} \right], \tag{3.9}$$

or the following LRA problem for some project n in isolation,

$$V_{A,n}^{\mathrm{L}}(x_{n,0}; \lambda) \triangleq \max_{\boldsymbol{\pi_n} \in \boldsymbol{\Pi_n}} \mathsf{E}_{x_{n,0}}^{\boldsymbol{\pi_n}} \liminf_{T \to \infty} \frac{1}{T} \left[\sum_{t=0}^{T-1} \{ R_n\big(X_{n,t}, a_{n,t}\big) - \lambda a_{n,t} \} \right] \tag{3.10}$$

where Π_n denotes the class of admissible policies for operating a single project, i.e., deciding when it should be active ($a_{n,t} = 1$) and passive ($a_{n,t} = 0$), and where λ is the Lagrangian multiplier. incorporated into the project's original flow of rewards.

As project state transitions are independent, the optimal value of the Lagrange re-

[1] We stress at this point that the heuristic policy based on this indexation approach is always feasible in terms of the hard sample path constraint as it is not prescribing to operate the system based on a policy solving the Lagrange relaxation.

laxation is decomposed as follows

$$V_D^{\mathrm{L}}(\mathbf{x_0};\lambda) = \sum_{n=1}^{N} V_{D,n}^{\mathrm{L}}(x_{n,0};\lambda) + \frac{M\lambda}{(1-\beta)}, \tag{3.11}$$

or

$$V_A^{\mathrm{L}}(\mathbf{x_0};\lambda) = \sum_{n=1}^{N} V_{A,n}^{\mathrm{L}}(x_{n,0};\lambda) + M\lambda, \tag{3.12}$$

Now, we can present the concept of *indexability*, introduced by Whittle (1988) as a key structural property of subproblems (3.9) and (3.10). In the remainder of the chapter, for ease of presentation, we will focus the ensuing discussion on the β-discounted subproblem (3.9). The analysis for the LRA case can be completed bearing in mind that its corresponding Whittle's index policy can be derived by letting $\beta \nearrow 1$ in the β-discounted index policy, provided that the limit defining (2.6) exists.

Definition 3.1. (Indexability) We say that project n is indexable if there exists a index function $\lambda_n^* : \mathbb{X}_n^{0,1} \mapsto \mathbb{R}$ such that for any value of the multiplier $\lambda \in \mathbb{R}$ and any controllable state $x_n \in \mathbb{X}_n^{0,1}$, it is optimal in subproblem (3.9), regardless of its initial state, to work in the project when it occupies state x_n iff $\lambda_n^*(x_n) \geq \lambda$

Whittle (1988) introduced the notion of *indexable RB projects* for the special case in which $W_n(x_{n,t}, a_{n,t}) = a_{n,t}$, while Niño-Mora (2002) extended such a concept to the case of general one-period resource consumption functions $W_n(x_{n,t}, a_{n,t})$. When Definition 3.1 holds, there exist a family of optimal policies for subproblem n which have a special structure that can be exploited to reduce the complexity of relaxed problem (3.7) as it implies that a reduced class of admissible policies in $\boldsymbol{\pi_n}$ may considered in order to solve its individual parts. Clearly, π_n^* belongs to the family of *deterministic stationary* policies Π_n^{DS}, since indexability implies the optimality of these family of policies for each λ-subproblem.

Thus, we can narrow our focus down to those policies and conveniently represent them by means of active sets $S \subseteq \mathbb{X}_n^{0,1}$, i.e., the set of controllable states in which the optimal action is to be active. Further, within *deterministic stationary* policies we can even focus on those policies with a certain monotonicity property with respect to this λ parameter, i.e. the (maximal) optimal active set $S^*(\lambda)$ expands monotonically as λ decreases.

3.3 Sufficient Indexability Conditions and Index Evaluation

If Whittle's indexability holds for a project, it ensures that the optimal policy for the RB subproblem can be characterized by a scalar priority index. Yet, this structural property needs to be analytically established for the model at hand, which is more often than not a challenging task. Furthermore, the characterization of the index given by Definition 3.1 is only implicit, and hence index computation may also demand a significant effort.

Such a state of affairs, motivated research to develop tractable sufficient indexability conditions. The first of such conditions for discrete-state restless bandits, along with an index algorithm, were introduced, developed and deployed in Niño-Mora (2001, 2002, 2006b). Such sufficient conditions draw on polyhedral arguments of having a problem which satisfies the Partial Conservation Laws (PCLs) for a postulate family of active sets. Such an approach has proven to be fruitful both in theoretical and algorithmic aspects, as well as in terms of the wide scope of successfully addressed applications. (For a detailed review of this indexation framework, see Niño-Mora, 2007a).

The scope of such discrete-state restless bandits sufficient indexability conditions has been extended to the real-state case in results announced in Niño-Mora (2008), as reviewed next. The following discussion focuses on a single-bandit problem modeling the optimal resource allocation problem of an individual project, whose label n is henceforth dropped from the notation. The MDP formulation is the following:

- The *state space* (we will focus on the controllable states) is a closed interval (possibly unbounded) of the real line $\mathbb{X}^{0,1} \subseteq \mathbb{X} \subseteq \mathbb{R}$;

- The *action set* $A(x) \subseteq \{0, 1\}$, with $a = 1$: *active/work*; $a = 0$: *passive/rest*.

- *Active dynamics*: If the project is at state x and the active action is selected at a given period, then during that period the system generates $R(x, 1)$ consuming $W(x, 1)$ unit of resources and paying λ per each of them. Next, the project moves to another state $Y^1 = y^1$ according to a stochastic kernel $P^1(.|x, 1)$.

- *Passive dynamics*: If the project is at state x and the passive action is selected at a given period, then during that period the system generates $R(x, 0)$ consuming $W(x, 0)$ unit of resources and paying λ per each of them. Next, the project moves to another state $Y^0 = y^0$ according to a stochastic kernel $P^0(.|x, 0)$.

- *One-period net expected reward* $R(x, a) - \lambda a$ if action a is deployed in state x.

The key to analytically establishing indexability conditions is to guess a family of stationary deterministic policies among which an optimal policy for (3.9) exists for every λ. For such a purpose, we shall evaluate the performance of an admissible policy $\pi \in \Pi$ along two dimensions: the *work measure*

$$g(x, \pi) \triangleq \mathsf{E}_x^\pi \left[\sum_{t=0}^{\infty} \beta^t a_t \right],$$

giving the ETD resource consumption of the project under policy π starting at $x_0 = x$; and the *reward measure*

$$f(x, \pi) \triangleq \mathsf{E}_x^\pi \left[\sum_{t=0}^{\infty} \beta^t R(x_t, a_t) \right],$$

giving the corresponding ETD reward achieved.

Note that the project's optimal control problem (3.9) is then reexpressed in terms of these measures as

$$V^*(x; \lambda) = \max_{\pi \in \Pi} f(x, \pi) - \lambda g(x, \pi). \tag{3.13}$$

In order to show indexability of (3.9), we must consider the existence of a structural property of optimal policies for the real-state MDP (4.6) as a function of the parameter λ. Henceforth, we refer to (4.6) as the project's *λ-charge subproblem*.

We shall further focus attention on the family of *threshold policies*. More precisely, for a given *threshold level* $z \in \overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$, the *$z$-threshold policy* activates the project in state x iff $x > z$, so its active set is $B(z) \triangleq \{x \in \mathbb{X}^{0,1} \colon x > z\}$. We let $B(z) = \mathbb{X}^{0,1}$ for $z = -\infty$,, and $B(z) = \emptyset$ for $z = \infty$. We denote by $g(x, z)$ and $f(x, z)$ the corresponding work and reward measures.

For fixed z, work measure $g(x, z)$ is characterized as

$$g(x, z) = \begin{cases} 1 + \beta \int_{\mathbb{X}} g(y, z) P^1(dy|x, 1), & x > z \\ 0 + \beta \int_{\mathbb{X}} g(y, z) P^0(dy|x, 0), & x \leq z, \end{cases} \tag{3.14}$$

whereas reward measure $f(x, z)$ is characterized by

$$f(x, z) = \begin{cases} R(x, 1) + \beta \int_{\mathbb{X}} f(y, z) P^1(dy|x, 1), & x > z \\ R(x, 0) + \beta \int_{\mathbb{X}} f(y, z) P^0(dy|x, 0), & x \leq z. \end{cases} \tag{3.15}$$

We shall use the marginal counterparts of such measures. For threshold z and action a, denote by $\langle a, z \rangle$ the policy that takes action a in the initial slot and adopts the z-

threshold policy thereafter. Define the *marginal work measure*

$$\begin{aligned} w(x,z) &\triangleq g(x,\langle 1,z\rangle) - g(x,\langle 0,z\rangle), \\ &= 1 + \left[\int_{\mathbb{X}} g(y,z)P^1(dy|x,1) - \int_{\mathbb{X}} g(y,z)P^0(dy|x,0)\right] \end{aligned} \tag{3.16}$$

and the *marginal reward measure*

$$\begin{aligned} r(x,z) &\triangleq f(x,\langle 1,z\rangle) - f(x,\langle 0,z\rangle). \\ &= R(x,1) - R(x,0) + \left[\int_{\mathbb{X}} f(y,z)P^1(dy|x,1) - \int_{\mathbb{X}} f(y,z)P^0(dy|x,0)\right] \end{aligned} \tag{3.17}$$

These measures respectively represent the marginal increase in resource expended and the marginal increase in rewards earned resulting from working instead of resting/iddling in the initial period and following the z-threshold policy afterwards. If $w(x,z) \neq 0$, define further the Marginal Productivity (MP) measure

$$\lambda^{MP}(x,z) \triangleq \frac{r(x,z)}{w(x,z)}. \tag{3.18}$$

Niño-Mora (2006b) coined the term marginal productivity based on the economic interpretations of this indexation methodology that will be reviewed in the following section. The following definition extends to the real-state setting a corresponding definition introduced by Niño-Mora (2001) for discrete-state restless bandits.

Definition 3.2. We say that subproblem (4.6) is *PCL-indexable* (with respect to threshold policies) if:

(i) *positive marginal work*: $w(x,z) > 0, x \in \mathbb{X}^{0,1}, z \in \overline{\mathbb{R}}$;

(ii) *nondecreasing index*: the index defined by

$$\lambda^{MP}(x) \triangleq \lambda^{MP}(x,x), \quad x \in \mathbb{X}^{0,1}. \tag{3.19}$$

is monotone nondecreasing and continuous in x

The next result, which was first stated in Niño-Mora (2008), extends the scope of a corresponding result in Niño-Mora (2001) for discrete-state restless bandits to the real-state setting, states the validity of the PCL-based sufficient indexability conditions reviewed in this section. It further shows how to evaluate the Whittle's MP index. The full proof of this result will be included in a paper, which is currently under preparation.

Theorem 3.1. *If subproblem (4.6) is PCL-indexable, then it is indexable and its MP index $\lambda^{MP}(x)$ in (3.19) is its Whittle's index $\lambda^*(x)$.*

If Theorem 3.1 holds, then we define the extended Whittle index policy for the original MARBP as follows: a t time t, the Whittle MP index policy selects at most M projects to work on, using $\lambda_n^*(x_{n,t})$ as a priority index for working on project n (where a larger index value means a higher priority), among those projects, if any, for which the index exceeds the λ charge, i.e., $\lambda_n^*(x_{n,t}) > \lambda$, breaking ties arbitrarily.

3.4 Indexability: Economic Interpretation

The indexability approach revised in this chapter admits an interesting geometric interpretation which highlights the economic meaning of the index. For some initial state x, consider project's (4.6) *achievable work-reward region*, as defined by Niño-Mora (2002, 2006b),

$$\mathbb{H}_x \triangleq \{(g(x,\pi), f(x,\pi)) : \pi \in \Pi\} \tag{3.20}$$

This is the region spanned in the plane by the pairs of total resource consumption-reward performance points under all admissible polices, and it is a closed convex polygon due to the optimality of the stationary deterministic policies for this kind of problems (see subsection 2.1.4). The z-threshold PCL-indexability property previously discussed can be assessed in terms of this region by analyzing the structure of the upper boundary of this region, which is defined as

$$\bar{\partial}\mathbb{H}_x \triangleq \{(g(x,\pi), f(x,\pi)) \in \mathbb{H}_x : f(x,\pi) \leq f \text{ for any } (g(x,\pi), f(x,\pi)) \in \mathbb{H}_x : g(x,\pi) = g\} \tag{3.21}$$

Whenever $\bar{\partial}\mathbb{H}$ is characterized by a *nested active set family* of z-threshold type in the following sense:

$$g(x,\infty) \leq g(x,z_i) \leq g(x,z_j) \leq g(x,-\infty) \quad \forall z_i, z_j \in \mathbb{X}^{0,1} : z_i > z_j \tag{3.22}$$

with $g(x,z_i)$ and $g(x,z_j)$ computed as in (7.7), and letting $g(x,\infty) = 0$ and $g(x,-\infty) = \mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t\right] = \frac{1}{1-\beta}$ then the project is PCL-indexable with respect to threshold polices.

To illustrate the idea, consider a restless bandit whose controllable state space is some interval of the real line and whose corresponding achievable work-reward region is given by Figure 3.1. The slope of the upper boundary represents the infinitesimal change in total rewards per unit of infinitesimal change in total work level, being thus a measure of *productivity* or *return* of the resources invested in the margin, before paying the current λ-charge. Given a particular λ-charge, e.g. $\lambda = \lambda_0$, we can determine the optimal *total resource-reward* pair as follows: provided that the marginal rate of productivity exceeds the marginal cost of investment λ_0 there is a net profit of allocating resources to this project. Thus, the optimal investment level for such a project, given $\lambda = \lambda_0$, corresponds to the pair $(g(x, z^*), f(x, z^*))$ Figure 3.1, in which the slope of the achievable work-reward region equals the current the λ-charge, achieving the best possible value (4.6).

As shown by Figure 3.2, for such subproblem the slope of the achievable work-reward region, representing the *marginal productivity* rate of resources invested, defines a monotone nondecreasing function of the subproblem's state x, $\lambda^*(x)$, which can be thus used to describe the subproblem's optimal investment policy: invest/work in the project provided in its current state the *marginal productivity* rate $\lambda^*(x)$ is greater or equal to the λ-charge. Furthermore, such a rule induces a natural monotone ordering of the bandit's controllable states.

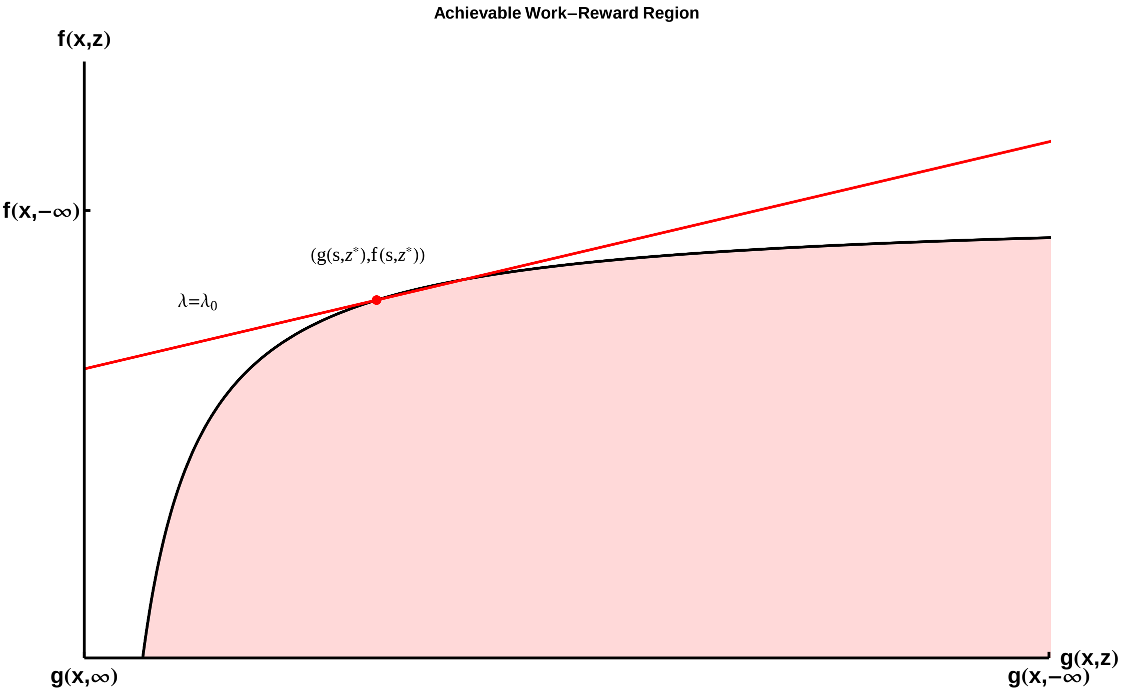


Figure 3.1: An illustration of the achievable work-reward region leading to an optimal family of z-threshold polices

In view of these results, it can be concluded that *indexable* subproblems are strongly connected to two fundamental ideas in traditional microeconomics: the *law of diminishing marginal returns* and the *profit maximization principle* of investment. The former, as already pointed out in Niño-Mora (2006b), derives from the fact that *indexable* projects are those in which as resource consumption increases, its marginal return rate dimini-

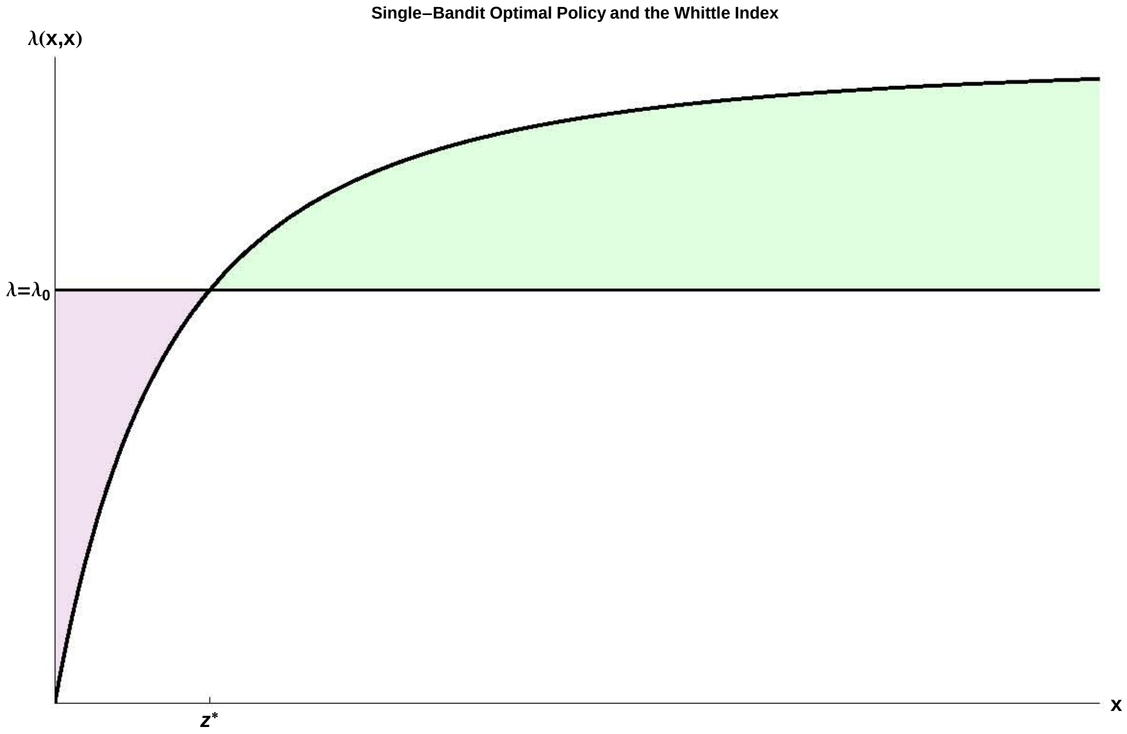


Figure 3.2: An indexable project, z-threshold optimal polices and the Whittle MP index

shes. Whereas the latter, which is a widely used principle for economic-financial evaluation of real-investment projects, prescribes to exploit a resource up the point in which the *marginal profit* of employing an extra unit of it equals zero. From definition (3.19), resources in indexable projects are allocated to work in a time slot only as long as the *marginal revenue* of investing them when project n occupies state x: $\lambda^*(x)$, exceeds the *marginal cost*: the charge λ. Thus, the optimal solution of indexable subproblems prescribes to engage in the project up to the point in which the *marginal profit* of investment is 0, i.e. $\lambda^*(x) - \lambda \geq 0$.

3.5 Applications

The exposition of the indexation literature we have done so far has focused on the intrinsic mathematical challenges raised by research on restless bandits. Yet, the wide range of applied problems falling within its scope has motivated a fast-growing attention of researchers. Among the most attractive features of the framework, two stand out: it yields both an intuitively appealing and economically sound heuristic index policy of low computational complexity, and it provides a practical way to asses the policy's suboptimality gap, through a bound on the optimal problem value.

Accompanying the work on theoretical aspects of bandit indexation, there have been many relevant and disparate applications of the methodology, unified in someway by this indexation framework. To give a glimpse of them, we refer to the following non-exhaustive list.

Related to queuing theory and optimal scheduling literature, we can mention: the dynamic control problem of customer admission and routing to parallel queues (Niño-Mora, 2002, 2007b) or the dynamic scheduling problem of a multiclass queue with finite buffers Niño-Mora (2006a), the scheduling of a multi-class make-to-stock queue (Niño-Mora, 2006b; Veatch and Wein, 1996; Dusonchet and Hongler, 2003). The indexation approach deployed to them yielded new insights and connections with routing problems.

Some applications arising in modern computer-communication networks: the problem of broadcast scheduling in information delivery systems (Raissi-Dehkordi and Baras, 2002); the dynamic bandwidth allocation in a communication channel with delays (Ehsan and Liu, 2004); and the dynamic scheduling of multiclass wireless transmissions (Ehsan and Liu, 2004; Niño-Mora, 2006a) Of special interest to this thesis, are the concrete applications of real-state MARBP: on opportunistic spectrum access, based on partial information (Niño-Mora, 2008; Liu and Zhao, 2008) and including sensing errors (Niño-Mora, 2009); multitarget tracking (Niño-Mora and Villar, 2009) and smart target hunt (Niño-Mora and Villar, 2011).

All these works form part of the large and still growing body of experimental evidence on the frequent near optimality of the resulting index policies and on their superior performance with respect to previously proposed heuristic index policies devised via ad hoc arguments.

Part II

Hunting Elusive Hiding Targets

But I still haven't found what I'm looking for
Bono.

Chapter 4

MARBP Formulation for Hunting Elusive Hiding Targets

In this chapter we formulate problem 1, stated in subsection 1.2.1, as a POMDP with special structure, which further fits into the frame of the real-state MARBP. We deploy the indexation methodology reviewed in Chapter 3 to propose a tractable heuristic search policy of priority-index type based on the Whittle index for RBs.

4.1 Background and Motivation

In recent years, the investigation of effective dynamic policies for operating wireless sensor networks has become an active research area. An issue that has received much attention is the design of scheduling policies to allocate over time a relatively small set of sensor resources to extract the required information about a scene containing a larger set of targets of interest, in order to optimize a system-wide performance objective. See, e.g., the survey Moran et al. (2008).

The sensors provide error-prone measurements of the sensed targets, such as their location, or their presence (or absence) at a given location. The current knowledge on each target is represented by its information state, which evolves via Bayesian updates depending on whether or not the target is sensed at each time slot. This allows for the formulation of a variety of optimal sensor scheduling problems as a POMDP with special structure, which often fit into the framework of the real-state MARBP, either in its classic version or, more often, in its restless variant. See, e.g., Washburn (2008). Although the restless variant is, generally, computationally intractable, formulating a sensor scheduling problem in such a framework allows for the use of the indexation methodology reviewed in the previous chapter. Such an approach, further provides with a bound on the optimal problem value that can be used to assess the deviation from optimality of a given policy.

In certain situations, sensing actions do not only affect the system's information state (e.g. in terms of its precision) but also alter targets' behavior. This is the case when objective targets are *smart*, in the sense that they react to being sensed by changing their dynamics, so as to hinder their detection or tracking. Sensor scheduling problems complicate substantially when targets under surveillance are able to detect and respond to sensing activities yet, it is natural to expect that different types of reaction would require a different operating rule to optimize system's performance.

Specifically, sensor scheduling to detect (and/or track) *smart* targets is an application that would strongly benefit from non-myopic decision rules, indicating the controller when it is better not to sense a site for the sake of the possible future gains obtained by influencing the target located at it accordingly. On the contrary, tractable myopic rules, of the type defined in (2.12) (Chapter 2), do not inform when a target should not be searched (specially in the case in which there are enough sensing resources available to do so). This is clearly undesirable if targets are elusive, as constantly searching for them makes them more and more elusive, resulting in larger use of system resources (especially in time) to successfully find them.

Despite all these problems, few papers have considered sensor scheduling problems with such reactive targets. Instead targets are typically assumed to follow dynamics that are unaffected by sensing decisions. In the recent literature, some sensor management models have been proposed for smart object localization disregarding such an unrealistic assumption. For instance, in Kreucher et al. (2006) reinforcement learning is used to obtain a non-myopic policy for detection and tracking of smart targets, while Liu et al. (2009) uses particle filter methods, and Savage and La Scala (2009) presents a game theoretic analysis.

The model presented in this chapter extends such a line of work by investigating a sensor scheduling model where a set of identical sensors are used to hunt a larger (or at least equal) set of heterogeneous targets, each of which is located at a corresponding site. As in Kreucher et al. (2006), target states change randomly over discrete time slots between *exposed* and *hidden*, according to Markovian transition probabilities that depend on whether sites are searched or not, so as to make the targets elusive. Sensors have a binary mode, so they can be either active or passive at a site, and they are imperfect, failing to detect an exposed target when searching its site with a positive misdetection probability

As a specific motivating application for such a model, we propose the problem investigated in Rucker (2006), where the targets are mobile platforms (transporter-erectorlaunchers) for launching short-range ballistic missiles (known as Scuds), and the sites are areas where it is known that such platforms are hidden. In this setting, the sensors can be mounted on unmanned aerial vehicles (UAV). A metric frequently used to measure the effectiveness of such operations is the time to detect all targets. Hence, an effective sensor scheduling rule may be derived by designing a a search policy that aims at maximizing the expected discounted rewards of detecting and destroying all missile launchers, where the discount factor represents how future detections are penalized in a given mission.

4.1.1 Goals and Contributions

It is the goal of this work to propose a dynamic and readily implementable index policy for a hunting elusive target model of POMDP type which exhibits a near-optimal performance both under the discounted and the total criterion.

We accomplish this by formulating the resulting POMDP as a real-state MARBP and deploying the recent extensions of the existing theoretical and algorithm results on discrete-state restless bandit indexation to the continuous-state case.

This work makes the following contributions: it successfully deploys the methodology announced in Niño-Mora (2008) to obtain a novel and dynamic index policy for the model of concern. The PCL-indexability of the model is shown for the ETD problem for discount factors smaller than a critical value. All this is done despite the lack of closed form expressions for the required performance measures, which is a severe technical difficulty introduced by considering a real state variable.

These contributions will be presented in this chapter in the following order: first we describe the model and we formulate it as a real-state MARBP. Next, in the remainder of the chapter we discuss the indexability analysis and the resulting index computation.

In the subsequent chapter, we present the empirical results which illustrate the indexability ideas discussed in this chapter as well as some interesting instances in which the proposed index policy not only outperforms alternative heuristic policies, but is shown to be near optimal.

4.2 Model description and MARBP Formulation

We consider a model where M sensors are available to hunt $N \geq M$ elusive hiding targets, where each target n is known to hide at a corresponding site $n = 1, \ldots, N$. We assume that the target present at site n alternates its visibility state $s_{n,t}$ at discrete time periods $t = 0, 1, \ldots$ over an infinite horizon between the *hidden* state ($s_{n,t} = 0$), in which it is invisible to sensors but cannot perform its tasks, and the *exposed* state ($s_{n,t} = 1$), in which it can perform its tasks but can be detected by a sensor surveying the site.

The visibility state $s_{n,t}$ evolves according to Markovian transition probabilities depending on whether or not its site is searched. We assume that only one sensor can search a site at each time slot, and model sensing decisions by binary actions processes $a_{n,t}$, where $a_{n,t} = 1$ if site n is sensed at time t, and $a_{n,t} = 0$ otherwise. When the action taken on site n is $a_{n,t} = a$ the target moves from the hidden to the exposed state (resp. from the exposed to the hidden state, in case the target is not detected) with probability $p_n^{(a)}$ (resp. $q_n^{(a)}$). Those transitions probabilities are such that after a site is searched and the *unhunted* target on it is not detected, it is more likely that the target moves into or remains in the hidden state than if the site had not been searched, i.e., $q_n^{(1)} > q_n^{(0)}$ and $p_n^{(1)} > p_n^{(0)}$. Notice that such condition ensures also that after a site is not searched, it is more likely that the target moves into or remains in the exposed state than if the site had been searched. We further assume that the visibility state processes have positive autocorrelation or memory, so $\rho_n^{(a)} \triangleq 1 - p_n^{(a)} - q_n^{(a)} > 0$.

The target at site n can only be hunted if it is exposed when searched, yielding a reward r_n for completing the site's mission. Information on target n's visibility state is gained by sensing it, which provides a *sensor outcome* $o_{n,t} \in \{0, 1\} : o_{n,t} = 1$ if the target is detected and hunted, and $o_{n,t} = 0$ otherwise. Sensing is imperfect in that the target at site n will not be detected when it is exposed and its site was sensed with a positive *misdetection* probability of $\alpha_n = P(o_{n,t} = 0|s_{n,t} = 1)$. Hence, target n's visibility state $s_{n,t}$ is not directly observable, but it is tracked by the *information state* $X_{n,t} \in \mathbb{X} \triangleq [0, 1]$, giving the posterior probability that the target is *exposed* in period t conditioned on the history $\{X_{n,s}, a_{n,s} : 0 \leq s < t\} \cup X_{n,t}$.
Since successfully hunting a target completes the mission at its site, we assume that a site n whose target has been hunted ($x_n = 0$) is removed from further search. Hence, we partition a target state space $\mathbb{X}$ into the set $\mathbb{X}^{0,1} \triangleq (0, 1]$ of controllable states, where both actions $A \triangleq \{0, 1\}$ are available, and the uncontrollable state 0, where only action $a_n = 0$ is available.

The dynamics of the information state for target n under each sensing action are obtained via Bayesian updates as follows. If the target at site's n has not yet been hunted at the beginning of period t, i.e. $X_{n,t} > 0$, and the site is searched ($a_{n,t} = 1$), then its next state will depend on wether the search was successful or not. Thus, if the sensor outcome is positive ($o_t^n = 1$), which happens with probability $(1 - \alpha_n)X_{n,t}$, and the target is detected $o_{n,t} = 1$, which happens with probability $(1 - \alpha_n)X_{n,t}$, then the target has been hunted and hence site n is removed from the search objectives. We model such a situation by letting the target's information state drop to zero, i.e. $X_{n,t+1} = 0$.

On the other hand, if the target is not detected $o_{n,t} = 0$, which happens with probability $1-(1-\alpha_n)X_{n,t}$, it is readily calculated that the information state changes to

$$X_{n,t+1} = p_n^{(1)} + \rho_n^{(1)} \left(\frac{\alpha_n X_{n,t}}{1-(1-\alpha_n)X_{n,t}} \right). \tag{4.1}$$

Hence, when site n is searched, its next information state is obtained in a randomized fashion depending on the sensing outcome.

Finally, if site n is not sensed ($a_{n,t} = 0$) in period t, with its information state being $X_{n,t} > 0$, i.e., as long as the target has not been hunted yet, its next information state is determined by

$$X_{n,t+1} = p_n^{(0)}(1-X_{n,t}) + (1-q_n^{(0)})X_{n,t}. \tag{4.2}$$

Yet, if the target has already been hunted $X_{n,t} = 0$, then its information state remains at 0 under both sensing actions. Thus, we summarize the information state dynamics for all controllable states $X_{n,t} \in \mathbb{X}^{0,1}$ as

$$X_{n,t+1} = \begin{cases} p_n^{(0)}(1-X_{n,t}) + (1-q_n^{(0)})X_{n,t}, & \text{if } a_{n,t} = 0 \quad \text{w.p } 1, \\ \\ 0, & \text{if } a_{n,t} = 1 \quad \text{w.p } (1-\alpha_n)X_{n,t}, \\ p_n^{(1)} + \rho_n^{(1)} \left(\dfrac{\alpha_n X_{n,t}}{1-(1-\alpha_n)X_{n,t}} \right), & \text{if } a_{n,t} = 1 \quad \text{w.p } 1-(1-\alpha_n)X_{n,t}, \end{cases}$$

Sensing actions are prescribed by a *scheduling policy* drawn from the class of admissible policies $\mathbf{\Pi}(M)$, consisting of the nonanticipative policies (i.e., based on the history of states and actions) that search at most M sites per slot:

$$\sum_{n=1}^{N} a_{n,t} \leq M, \quad t = 0, 1, \ldots. \tag{4.3}$$

As for the economic results of the sensing actions, taking action a_n on site n when it occupies the information state x_n yields the *expected one-slot net reward* $R_n(x_{n,t}, a_{n,t}) \triangleq (r_n\,(1-\alpha_n)\,x_n - c_n)\,a_n$, where $c_n \geq 0$ is a fixed site/target specific sensing cost.

The sensing system described by this model operates over time slots of equal length, assuming sensors are synchronized to operate over discrete time slots. The sequence of events within each slot is described in Figure 4.1. At the beginning of each slot, the system's manager given site's n current *information state* $X_{n,t}$, decides whether to sense that site or not, afterwards earning an expected reward $R_n(x_{n,t}, a_{n,t})$ which depends on the selected action and the current belief state. Afterwards target's n, if not hunted, changes its visibility state depending on the selected sensing action and hence by the end of the slot, site's n *belief state* is updated accordingly.

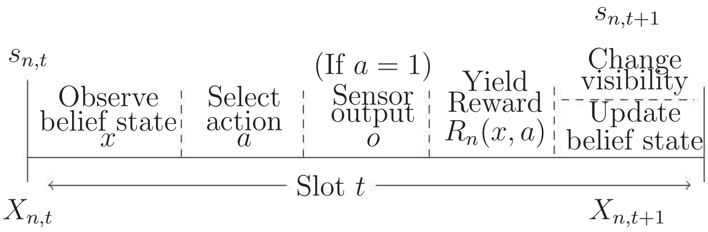


Figure 4.1: The sequence of events within a time slot for the elusive target hunt model.

4.2.1 Performance Objectives

We will consider the following dynamic optimization problem: find a *β-discounted reward optimal* policy, i.e.,

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(\boldsymbol{M})} \mathsf{E}_{\mathbf{x}_0}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty}\sum_{n=1}^{N}\beta^t R_n\big(X_{n,t}, a_{n,t}\big)\right], \tag{4.4}$$

where $0 < \beta \leq 1$ is the discount factor, $\mathbf{x}_0 = (x_{n,0})_{n=1}^{N}$ is the initial joint belief state, for n in $\{1, 2, \ldots, N\}$, and $\mathsf{E}_{\mathbf{x}_0}^{\boldsymbol{\pi}}[\cdot]$ denotes expectation under policy $\boldsymbol{\pi}$ conditioned on the initial joint state being equal to $\mathbf{x}_0$. Note that the undiscounted case $\beta = 1$, which corresponds to the *total expected reward* criterion, is well defined in the present setting given that the search plan terminates after a finite number of slots with probability one (but the number of slots until termination, i.e., the horizon, is uncertain and unbounded). Furthermore, when considering a discount factor $\beta = 1$ we may analize the case in which there is interest in finding targets regardless of how long it takes to do so. When there are reasons to penalise finding targets in a later futute, such as system's lifetime constraints, it makes sense to consider some $\beta < 1$.

As discussed in Chapter 2, problem (4.4) is a POMDP of restless MABP type, thus being notoriously hard to solve exactly. In the following section we shall present the results of deploying the real-state restless bandit Whittle MP indexation approach to the model of concern.

4.3 Real State Restless Bandit Indexation Analysis

4.3.1 Verification of PCL-indexability

As reviewed in Chapter 3, establishing Whittle's indexability Definition 3.1 by means of deploying sufficient indexability conditions 3.2 we focus on an individual site's subproblem as

$$\max_{\boldsymbol{\pi_n \in \Pi_{(n)}}} \mathsf{E}^{\boldsymbol{\pi_n}}_{\mathbf{x_{n,0}}} \left[\sum_{t=0}^{\infty} \beta^t \{ R_n \left(X_{n,t}, a_{n,t} \right) - \lambda a_{n,t} \} \right], \tag{4.5}$$

where (4.5) is a single-project restless bandit subproblem, consisting of a hunting problem considered for some site n in isolation. Π_n denotes the class of admissible policies for operating a single sensor on such site, i.e., deciding when it should be active ($a_{n,t} = 1$) and passive ($a_{n,t} = 0$) and with λ being a constant parameter representing an extra cost incurred per unit of time the sensor is active.

Next, we would like to establish that each subproblem (4.5) has the key structural *indexability* property defined by Definition 3.1. For such a purpose, we will deploy conditions 3.2 to establish that the problem is indexable with respect to the family of z-threshold policies, and thus we start by computing the performance measures under such a class of policies. In the remainder of this section we focus on a generic single site/target subproblem as (4.5), and hence drop the superscript n from the above notation.

We recall from Chapter 3 that we can evaluate the performance of any admissible sensing policies $\pi \in \Pi$ along two dimensions: the *work measure* $g(x, \pi)$, giving the ETD number of times a site is sensed under policy π starting at $X_0 = x$; and the *reward measure* $f(x, \pi)$, giving the corresponding ETD reward earned. Thus,

$$g(x, \pi) \triangleq \mathsf{E}^{\pi}_{x} \left[\sum_{t=0}^{\infty} \beta^t a_t \right], \quad f(x, \pi) \triangleq \mathsf{E}^{\pi}_{x} \left[\sum_{t=0}^{\infty} \beta^t R(X_t, a_t) \right].$$

So that we can formulate the single-site's optimal target hunting subproblem (4.5) as

$$\max_{\pi \in \Pi} f(x, \pi) - \lambda g(x, \pi). \tag{4.6}$$

Problem (4.6), is a real-state MDP, whose optimal policy, under certain assumption on the reward function $R(x, a)$[1], belongs to the family of *deterministic stationary policies* Π^{SD}, naturally represented by their *active (state) sets* (in this case, that is the set of information states where sensing the site is prescribed). For an active set $B \subseteq \mathbb{X}^{0,1}$, we shall refer to

[1] Assuming, for instance, that $R(x, a)$ is bounded and measurable ensures this fact by Blackwell Sufficient Conditions.

the B-*active policy*. We will further focus attention on the family of *threshold policies*. For a given *threshold level* $z \in \mathbb{R}$, the z-*threshold policy* senses the site in information state x iff $x > z$, so its active set is $B(z) \triangleq \{x \in \mathbb{X}^{0,1} \colon x > z\}$. Note that $B(z) = (z, 1]$ for $0 \leq z < 1$, $B(z) = \mathbb{X}^{0,1} = (0, 1]$ for $z < 0$, and $B(z) = \emptyset$ for $z \geq 1$. We denote by $g(x, z)$ and $f(x, z)$ the corresponding work and reward measures under a z-threshold policy.
In the following we will use the notation to stand for the functions in (4.1) and (4.2):

$$\phi^{(0)}(x) \triangleq (p^{(0)} + \rho^{(0)} x), \quad \phi^{(1)}(x) \triangleq p^{(1)} + \rho^{(1)} \frac{\alpha x}{1 - (1 - \alpha) x}. \tag{4.7}$$

For some fixed z, the total work measure $g(x, z)$ for any $x \in \mathbb{X}^{0,1}$ is characterized as the unique solution in the Banach space of bounded measurable real-valued functions on $\mathbb{X}$ endowed with the sup norm (See Hernández-Lerma and Lasserre, 1999) to

$$g(x, z) = \begin{cases} 1 + \beta \ [1 - (1 - \alpha)\, x\,]\, g\big(\phi^{(1)}(x), z\big), & x \in (z, 1] \\ \beta g\big(\phi^{(0)}(x), z\big), & x \in (0, z], \end{cases} \tag{4.8}$$

whereas the total reward measure $f(x, z)$ for any $x \in \mathbb{X}^{0,1}$ is characterized as the unique solution in the Banach space of bounded measurable real-valued functions on $\mathbb{X}$ endowed with the sup norm to

$$f(x, z) = \begin{cases} R(x, 1) + \beta \ [1 - (1 - \alpha)\, x\,]\, f\big(\phi^{(1)}(x), z\big), & x \in (z, 1] \\ \beta f\big(\phi^{0}(x, z\big), & x \in (0, z]. \end{cases} \tag{4.9}$$

Notice that for deriving expressions (4.8) and (4.9) we have used the fact that at the uncontrollable state the system does not operate, i.e., we let $g(0, z) = f(0, z) = 0$ for any possible threshold value z.

We will further use the marginal counterparts of such total evaluation measures. For any fixed threshold z and action a, denote by $\langle a, z\rangle$ the policy that takes action a in the initial period and adopts the z-threshold policy thereafter. Define the *marginal work measure* $w(x, z)$ and the *marginal reward measure* $r(x, z)$ as

$$\begin{aligned} w(x, z) &\triangleq g(x, \langle 1, z\rangle) - g(x, \langle 0, z\rangle), \\ &= 1 + \beta \ [1 - (1 - \alpha)\, x\,] \ g(\phi^{(1)}(x), z) - \beta\, g(\phi^{(0)}(x), z) \end{aligned} \tag{4.10}$$

$$\begin{aligned} r(x, z) &\triangleq f(x, \langle 1, z\rangle) - f(x, \langle 0, z\rangle), \\ &= R(x, 1) + \beta \ [1 - (1 - \alpha)\, x\,] \ f(\phi^{(1)}(x), z) - \beta\, f(\phi^{(0)}(x), z) \end{aligned} \tag{4.11}$$

If $w(x,z) \neq 0$, we define the MP measure

$$\lambda^{MP}(x,z) \triangleq \frac{r(x,z)}{w(x,z)}. \tag{4.12}$$

Total and Marginal Evaluation Measures

In order to analyze the PCL-indexability of (4.6) by means of the sufficient indexability conditions (SIC) stated in Definition 3.2 we must first solve the evaluation measures $g(x,z)$ and $f(x,z)$ for any fixed threshold $z \in \mathbb{R}$ and any $x \in \mathbb{X}^{0,1}$.

In order to do so, we must successfully address the problem posed by the fact that possible information state trajectories $\{X_t\}$ are naturally infinite, since X_t can take any value in $\mathbb{X}$ at each t. To do so, we will take advantage of the fact that under a z-threshold policy for any initial state x, possible information state trajectories $\{X_t\}$ are infinite but numerable, as they exhibit recurrent cyclical patterns depending on the threshold level. Yet, as we will next show, the total performance measures do not converge to a simple closed-form expression. In the cases in which such measures can be solved in closed form, as e.g. Niño-Mora (2008), both direct verification of the SIC and obtaining a closed-form index formula are possible. Yet, in both models addressed in this dissertation, a significant challenge is establish indexability and to derive an index policy, is to do so despite the fact that the evaluation equations do not admit a straightforward algebraic manipulation.

Next, we outline how to solve the evaluation measures to perform an indexability analysis and further shows how to use such solutions to evaluate the index $\lambda^{MP}(x)$ and to establish the PCL-indexability of the model.

To solve for (4.8) and (4.9) in closed form we further define $\phi_t^{(a)}(x)$ for $a = 0,1$ as the recursion generated by letting $\phi_0^{(a)}(x) \triangleq x$ and $\phi_t^{(a)}(x) \triangleq \phi_0^{(a)}(\phi_{t-1}^{(a)}(x))$ for $a = 0,1$. Note that for any $x \in \mathbb{X}^{0,1}$, both recursions $\phi_t^{(a)}(x)$ converge as $t \to \infty$ to the respective limits

$$\phi_\infty^{(0)} \triangleq \frac{p^{(0)}}{1-\rho^{(0)}} \qquad \phi_\infty^{(1)} \triangleq \frac{\gamma - \sqrt{\gamma^2 - 4p^{(1)}(1-\alpha)}}{2(1-\alpha)}$$

with $\gamma \triangleq 1 - \rho^{(1)} + (p^{(1)} + \rho^{(1)})(1-\alpha)$.

Most importantly, notice that both functions (4.1) and (4.2) and their resulting iterated mappings can be seen as (non-linear) functions known as Möbius transformations (or also as Linear Fractional Transformations). This observation was crucial to deriving the results of the subsequent indexability analysis. The Möbius transformations properties and results deployed in this dissertation are reviewed in Appendix A.

The definitions of (4.1) and (4.2) ensure that that $\phi_\infty^{(1)} < \phi_\infty^{(0)}$ (as shown in Lemma B.1

in Appendix B), and both limits are attractive fixed points of the active and passive dynamics respectively. This naturally divides the state space into three parts as portrayed in Figure 4.2 where the active and passive actions (depending on the initial information state x and the threshold z) produce movements in the state space, which are either both increasing (if $x, z \in (0, \phi_\infty^{(1)})$) or both decreasing (if $x, z \in [\phi_\infty^{(0)}, 1]$) or moving in opposite directions (if $x, z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$). Hence, we exploit this knowledge to solve the evaluation equations by distinguishing among three $z-$threshold cases, as discussed below.

In the sequel we assume, without loss of generality, that $c = 0$.

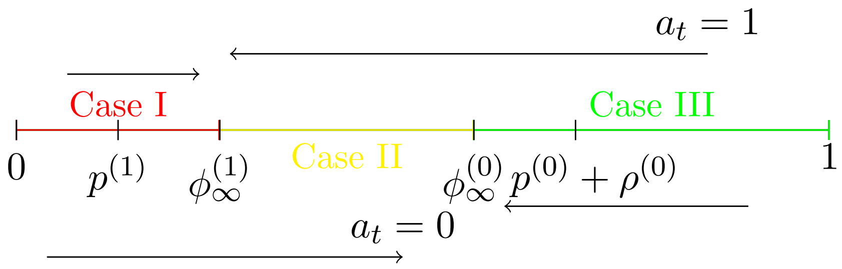


Figure 4.2: The state space and the fixed points of the active and passive dynamics for the single elusive target hunt model

Case I: Threshold $z \in [0, \phi_\infty^{(1)})$ (*Low thresholds*)

In this case, the active set $B(z) = (z, 1]$ contains the attractive fixed points of the recursions associated to both actions, i.e. $\phi_\infty^{(0)}, \phi_\infty^{(1)}$. This implies that once the state reaches the active set $B(z)$ it stays in $B(z)$ as long as the target remains unhunted. Such a result follows from Lemma B.2 in Appendix A by which for any $x \geq \phi_\infty^{(1)}$ then $\phi_t^{(1)}(x) \geq \phi_\infty^{(1)} > z$ for all $t \geq 0$. Further, for $z \in B^c(z) \triangleq [0, z]$: $\phi_t^{(0)}(x) \nearrow \phi_\infty^0$. Hence, after a finite number of passive slots $t_0^*(x, z) < \infty$: $\phi_{t_0^*(x,z)}^{(0)}(x) > z$, where we define the first (deterministic) hitting time to the active set as $t_0^*(x, z) \triangleq \min\{t \geq 1 : \phi_t^{(0)}(x) > z\}$
Also, denote by $\theta(x, z, t)$ the survival probability representing the probability that the target has not been hunted before time slot t under the z-threshold policy, starting from

state x. Note that, for $x > z$

$$\theta(x,z,t) \triangleq \prod_{s=0}^{t-1} \left[1 - (1-\alpha)\, \phi_s^{(1)}(x)\right] \tag{4.13}$$

where we let $\theta(x,z,0) = 1$. Thus, the total work measure has the following evaluation:

$$g(x,z) = \begin{cases} \sum_{t=0}^{\infty} \beta^t\, \theta(x,z,t) & x \in (z,1] \\ \beta^{t_0^*(x,z)} \left[\sum_{t=0}^{\infty} \beta^t\, \theta(y,z,t)\right] & x \in (0,z] \end{cases} \tag{4.14}$$

where $y \triangleq \phi^{(0)}_{t_0^*(x,z)}(x)$.
Similarly, we obtain the total reward evaluation

$$f(x,z) = \begin{cases} \sum_{t=0}^{\infty} \beta^t\, \theta(x,z,t)\; R(\phi_t^{(1)}(x,z),1) & x \in (z,1] \\ \beta^{t_0^*(x,z)} \sum_{t=0}^{\infty} \beta^t\, \theta(y,z,t)\; R(\phi_t^{(1)}(y,z),1) & x \in (0,z] \end{cases} \tag{4.15}$$

The above infinite series are convergent, yet they do not admit closed form formulae. Hence, they must be truncated in practice to evaluate $w(x,z)$ and $r(x,z)$ via (4.10) and (4.11), and hence also for establishing that SIC conditions i) and ii) in Definition 3.2 hold.

In the following we list our main results, drawing on the technical analysis of the marginal work measure presented in Appendix B. As explained in detail in that Appendix, β^* is defined as the discount factor β such that:

$$\left(\sum_{t=0}^{\infty}(\beta^*)^t \theta(x,z,t) - \beta^* \sum_{t=0}^{\infty}(\beta^*)^t \theta(\phi^{(0)}(x),z,t)\right) = 0 \tag{4.16}$$

We further define $\beta^{(1)}$ as:

$$\beta^{(1)} \triangleq \frac{1}{1 + \left[1 - (1-\alpha)(1-\phi_\infty^{(1)})\right]}.$$

The strategy deployed for proving the positivity of marginal work measures in all the threshold cases of concern, despite the lack of a closed form formulae, is the following: for each z-threshold case and every possible initial state x, based on properties of the active and passive recursions as Möbius transformations, we derive a lower bound on $w(x,z)$ and then study its positivity (or the conditions under which its positivity it is ensured).

The lower bounds on $w(x,z)$ in this case are given in the following lemma.

Lemma 4.1. *For all $z < \phi_\infty^{(1)}$,*

(a) $w(x,z) > \min\{1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z}, 1-\beta\} \geq 0$

for any $x \in (0,z], \quad 0 \leq \beta \leq 1.$

(b) $w(x,z) > (1-\beta) \geq 0$ *for any* $x \in (z, \phi_\infty^{(0)}], \quad 0 \leq \beta \leq 1.$

(c) $w(x,z) > 0$ *for any* $x \in (\phi_\infty^{(0)}, 1]$*, only if* $\beta < \beta^*$.

The following proposition, based on the above lemmas, provides the conditions under which positivity of the marginal work measures is ensured.

Proposition 4.1. *The marginal work measure $w(x,z)$ in problem (4.5) with $x \in \mathbb{X}^{0,1}$ and $z \in [0, \phi_\infty^{(1)})$ is strictly positive for $\beta < \beta^*$ with $\beta^* > \beta^{(1)}$.*

The complete derivation of these bounds which proves Proposition 4.1 in shown in Appendix B.

As there is no closed form expression for those infinite sums, β^* cannot be computed exactly. $\beta^{(1)}$ is a lower bound on it obtained by imposing that the lowest bound on $w(x,z)$ for $x = 1$ is strictly positive. Further bounds can be obtained by truncation of the infinite sums in (4.16).

Proposition 4.1 ensures that condition (i) in the SIC holds for this case. Regarding the monotonicity condition of the index, first notice that it follows from the definition of $t_0^*(x,z)$ that: $t_0^*(\phi^{(1)}(x),x) = t_0^*(\phi^{(0)}(x),x) = 0$, since $\phi^{(0)}(x) > x$ and $\phi^{(1)}(x) > x$, which allows us to compute the index (4.12) for case I as follows:

$$\lambda^{MP}(x) = \frac{R\,(1-\alpha)\left[\sum_{t=0}^{\infty}\beta^t\left[\phi_t^{(1)}(x)\,\theta(x,x^-,t) - \beta\phi_t^1(\phi^{(0)}(x))\,\theta(\phi^{(0)}(x),x,t)\right]\right]}{\sum_{t=0}^{\infty}\beta^t\left[\theta(x,x^-,t) - \beta\,\theta(\phi^{(0)}(x),x,t)\right]}, \quad (4.17)$$

for $x \in (0, \phi_\infty^{(1)})$, where x^- stands for the sensing policy with active set equal to $B(x^-) = [x,1]$.

Notice, that for all $x \in [0, \phi_\infty^{(1)})$ the $\lambda^{MP}(x)$ is an infinite sum of continuous functions of the state. Next, to ensure indexability we must prove that this index is nondecreasing with respect to the information state. For such a purpose, we must take the derivative of the two infinite sums defining the index with respect to x. Since, there is no closed form formulae for those sums to manipulate it algebraically, the strategy to accomplish such a goal is to show that it holds that a) $\frac{\partial w(x,x)}{\partial x} < 0$ and b) $\frac{\partial r(x,x)}{\partial x} > 0$ by manipulating the derivative of each term in the infinite sum.

Lemma 4.2. *For all $x < \phi_{\infty}^{(1)}$, it holds that:*

$$\sum_{t=1}^{\infty} \beta^t \left[\frac{\partial \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \theta(\phi^{(0)}(x), x, t)}{\partial x} \right] < 0 \qquad (4.18)$$

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial \phi_t^{(1)}(x) \, \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \phi_t^{(1)}(\phi^{(0)}(x)) \, \theta(\phi^{(0)}(x), x, t)}{\partial x} \right] > 0 \qquad (4.19)$$

The following proposition, based on the above lemma, provides the conditions under which MP index is monotone. Proof of Lemma 4.2 is included in the Appendix B.

Proposition 4.2. *The index $\lambda^{MP}(x)$ as defined in (4.17) for problem (4.5) is monotone increasing and continuous in the information state x for $x \in (0, \phi_{\infty}^{(1)})$..*

Case II: Threshold $z \in [\phi_{\infty}^{(1)}, \phi_{\infty}^{(0)})$ (*Intermediate thresholds*)

In this case, the passive set $B^c(z)$ contains the attractive fixed point of the recursion associated to the active action, i.e. $\phi_{\infty}^{(1)}$, whereas the active set $B(z)$ contains the attractive fixed point of the recursion associated to the passive action, i.e. ϕ_{∞}^{0}. Hence, the state X_t jumps above and below the threshold z, until the target is found. Following the argument introduced in Niño-Mora (2009), define the map $\phi(x, z) \triangleq 1_{(x>z)}\phi^{(1)}(x) + 1_{(x \leq z)}\phi^{(0)}(x)$, and let $\phi_0(x, z) = x$, $\phi_t(x, z) = \phi(\phi_{t-1}(x, z), z)$ for $t \geq 1$. Then, writing $a_t(x, z) \triangleq 1_{(\phi_t(x,z)>z)}$, $(\phi a)_t(x, z) \triangleq \phi_t(x, z)a_t(x, z)$. In this case, the survival probability has evaluation

$$\theta(x, z, t) \triangleq \prod_{s=0}^{t-1} \left[1 - (1 - \alpha) \, (\phi a)_s(x, z)\right] \qquad (4.20)$$

with $\theta(x, z, 1) = 0$. Thus, total evaluation measures admit the following expressions

$$g(x, z) = \sum_{t=0}^{\infty} \beta^t \, a_t(x, z)\theta(x, z, t) \qquad (4.21)$$

$$f(x, z) = \sum_{t=0}^{\infty} \beta^t \, R\left(\, (\phi a)_t(x, z), 1 \, \right) \, \theta(x, z, t) \qquad (4.22)$$

In this case also, since the expressions (4.21) and (4.22) cannot be calculated in a closed form, truncation is neccesary for evaluating them numerically. However, we are able to describe a *recurrent cyclical* pattern in the resulting information state X_t process under a z-threshold policy, which allows us to describe the possible trajectories of the information state to be considered. Specifically, using properties of the Möbius Transformations we are able to establish regularities , in terms of the sequence of active and passive slots until a target is hunted, that allow us to derive the corresponding lower bounds on $w(x, z)$ for this case, which is the most complex of the three threshold cases. Those regularities are summarized in Lemma B.10, Lemma B.11 and Lemma B.12 of the Appendix B. In the following we list the main results for this case.

Lemma B.10, Lemma B.11 and Lemma B.12 are used to derive the lower bound on $w(x, z)$ for this case.

The lower bounds on $w(x, z)$ in this case are given in the following lemma.

Lemma 4.3. *For all* $z \in [\phi_\infty^{(1)}, \phi_\infty^0)$,

(a) $w(x, z) > \min\{1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,\phi_\infty^{(1)}}, 1 - \beta\} \geq 0$
for any $x \in (0, z]$, $\quad 0 \leq \beta \leq 1$.

(b) $w(x, z) > (1 - \beta) \geq 0 \quad$ *for any* $x \in (z, \phi_\infty^0]$, $\quad 0 \leq \beta \leq 1$.

(c) $w(x, z) > 0 \quad$ *for any* $x \in (\phi_\infty^0, 1]$, *only if* $\beta < \beta^*$

The following proposition, based on the above lemma, provides the conditions under which positivity of the marginal work measures is ensured. The complete derivation of these bounds in shown in Appendix B.
Proposition 4.3 ensures that condition (i) in the SIC holds for this case.

Proposition 4.3. *The marginal work measure* $w(x, z)$ *in problem* (4.5) *with* $x \in \mathbb{X}^{0,1}$ *and* $z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$ *is positive for* $\beta < \beta^*$ *with* $\beta^* > \beta^{(1)}$.

Further, Proposition 4.3 implies that Proposition 4.1 holds in this threshold case also. Next, we compute the index (4.12) in this case, using the fact that $t_0^*(x, x) = 1$ and, given that $\phi^{(1)}(x) < x < \phi^{(0)}(x)$, as follows:

$$\lambda^{MP}(x) = \frac{\sum_{t=0}^{\infty} \beta^t R\,(1-\alpha)\left[(\phi a)_t(x, x^-)\theta(x, x^-, t) - \beta(\phi a)_t(\phi^{(0)}(x), z)\,\theta(\phi^{(0)}(x), x, t)\right]}{\sum_{t=0}^{\infty} \beta^t\left[\,\theta(x, x^-, t)a_t(x, x^-) - \beta\,\theta(\phi^{(0)}(x), x, t)a_t(\phi^{(0)}(x), z)\right]}, \tag{4.23}$$

for $x \in [\phi_\infty^{(1)}, \phi_\infty^0)$, where where x^- stands for the sensing policy with active set equal to $B(x^-) = [x, 1]$.

Such an index can be expressed as an infinite sum of functions defined by a composition of the two the Möbius transformations describing the active and passive dynamics, depending on the concrete cycle that a given threshold x generates. Showing continuity of the index in this case calls for further research, but the experimental evidence suggests it holds.

Lemma 4.4. *For all $x < \phi_\infty^{(1)}$, it holds that:*

$$\sum_{t=1}^{\infty} \beta^t \left[\frac{\partial \theta(x, x^-, t) a_t(x, x^-)}{\partial x} - \beta \frac{\partial \theta(\phi^{(0)}(x), x, t) a_t(\phi^{(0)}(x), z)}{\partial x} \right] < 0 \tag{4.24}$$

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial (\phi a)_t(x, x^-) \theta(x, x^-, t)}{\partial x} - \beta \frac{\partial (\phi a)_t(\phi^{(0)}(x), x)\; \theta(\phi^{(0)}(x), x, t)}{\partial x} \right] > 0 \tag{4.25}$$

The following proposition, based on the above lemma, provides the conditions under which MP index is monotone. roof of Lemma 4.3 and Lemma 4.4 are included in the Appendix B and also follow from the application of Lemma B.10, Lemma B.11 and Lemma B.12 and further properties of the Möbius Transformations.

Proposition 4.4. *The index $\lambda^{MP}(x)$ as defined in (4.23) for problem (4.5) is monotone increasing and continuous in the information state x for $x \in [\phi_\infty^{(1)}, \phi_\infty^{(0)}]$.*

Case III: Threshold $z \in [\phi_\infty^{(0)}, 1]$ (*High thresholds*)

In this case, the passive set $B^c(z)$ contains the attractive fixed points of the recursions associated to both actions, i.e. $\phi_\infty^{(0)}, \phi_\infty^{(1)}$. This, in turn, implies that once the information state reaches the passive set $B^c(z)$, it remains in it, regardless if the target has been hunted or not at that moment of time. Such a result follows from Lemma B.2 by which for all $x > z$, $z \geq \phi_\infty^{(0)} \geq \phi_t^{(0)}(x)$ for all $t \geq 0$. Further, for $z \in [\phi_\infty^{(0)}, 1]$ and $x > z$ by Lemma B.2: $\phi_t^{(1)}(x) \swarrow \phi_\infty^{(1)}$. Hence, after a finite number of active slots $\tau^*(x, z) < \infty$, with $\tau^* \triangleq \min\{t \geq 1 : X_t \leq z\}$,: $\phi^1_{\tau^*}(x) \leq z$. Notice that $\tau^*(x, z)$ for some $x > z$ is a random variable with maximum value $t_1^*(x, z) \triangleq \min\{t \geq 1 : \phi_t^{(1)}(x) \leq z\}$.
Then, we have that

$$g(x, z) = 1_{\{x > z\}} \left[\sum_{t=0}^{t_1^*(x,z)-1} \beta^t \theta(x, z, t) \right], \tag{4.26}$$

$$f(x, z) = 1_{\{x > z\}} \left[\sum_{t=0}^{t_1^*(x,z)-1} \beta^t \, R(\, \phi_t^{(1)}(x), 1\,) \theta(x, z, t) \right]. \tag{4.27}$$

where $\theta(x, z, t)$ is the survival probability as defined in Case I. For $x > z$, equations (4.26) and (4.27) are readily computed by evaluating finite sums with $t_1^*(x, z) - 1$ terms.

The lower bounds on $w(x, z)$ in this case are respectively given by Lemma 4.5.

Lemma 4.5. *For all $z \geq \phi_\infty^0$,*

(i) $w(x, z) = 1$ *for any* $x \in (0, z], \quad 0 \leq \beta \leq 1.$

(ii) $w(x, z) > 0 \quad$ *for any* $x \in (z, 1]$ *for* $\beta < \beta^*$.

The following proposition, based on the above lemma, provides the conditions under which positivity of the marginal work measures is ensured.

Proposition 4.5. *The marginal work measure $w(x, z)$ in problem* (4.5) *with $x \in \mathbb{X}^{0,1}$ and $z \in [\phi(0)_\infty, 1)$ is positive for $\beta < \beta^*$ with $\beta^* > \beta^{(1)}$.*

Hence, for $x \leq z$ it is readily seen that $w(x, z) = 1$ and $r(x, z) = R(x, 1)$. Therefore, the index in (3.19)

$$\lambda^{MP}(x) = R(x, 1), \quad \phi_\infty^{(0)} \leq x \leq 1 \tag{4.28}$$

The following proposition, based on the above lemma, provides the conditions under which MP index is monotone.

Proposition 4.6. *The index $\lambda^{MP}(x)$ as defined in* (4.28) *for problem* (4.5) *is monotone increasing and continuous in the information state xfor $x \in (\phi_\infty^{(0)}, 1]$.*

Proof. Given the MP index is, in this case, a continuous function of the information state, we take partial derivative to index (4.28), it follows that:

$$\frac{\partial \lambda^{MP}(x)}{\partial x} = \frac{\partial R(x, 1)}{\partial x} = \frac{\partial r(1-\alpha)x}{\partial x} = r(1-\alpha) > 0.$$

□

Notice that in case III, the MP index $\lambda^{MP}(x)$ coincides with the myopic index $\lambda^{myopic}(x)$, as defined in (2.12).

Verification of PCL-indexability Sufficient Conditions

Based on Proposition 4.1-Proposition 4.6, we conclude:

Theorem 4.1. *The single-site elusive target hunt ETD problem* (4.5) *is PCL indexable for $\beta \in [0, \beta^*)$, with*

$$\beta^* > \frac{1}{1 + \left[1 - (1-\alpha)(1 - \phi_\infty^{(1)})\right]}.$$

Therefore, it is indexable for $\beta \in [0, \beta^*)$, and the MP index $\lambda^{MP}(x)$ calculated above is its Whittle index $\lambda^*(x)$.

Short Proof: Proposition 4.1, Proposition 4.3 and Proposition 4.5 ensure the positivity of the marginal work measure for $\beta \in [0, \beta^*)$. Proposition 4.2, Proposition 4.4 and Proposition 4.6 ensure monotonicity and continuity.
Remark:
Once the information state process X_t reaches the set $[\phi_\infty^{(1)}, \phi_\infty^{(0)}]$, it never leaves it. Then, for $x \in [0, \phi_\infty^{(0)}]$ the ETD problem (4.5) is PCL-indexable for all discount values $\beta \in [0, 1]$, as shown by Proposition 4.1, Proposition 4.3 and Proposition 4.5. Hence, the set of information states for which PCL-indexability in ensured only if $\beta < \beta^*$, i.e. $x \in (\phi_\infty^{(0)}, 1]$, applies only to a set of states which the system will, with certainty, leave and never return to, since the subset $[\phi_\infty^{(1)}, \phi_\infty^{(0)}]$ contains the absorbing set of states of the system operated under any z-threshold policy (See Lemma B.10).

4.3.2 Index Computation

The Whittle index has evaluation given by (4.17), (4.23) and (4.28). As already mentioned during the indexability analysis, the index $\lambda^*(x)$ for the information states $0 \leq x < \phi_\infty^{(0)}$ in practice must be computed by truncating the infinite series defining them to a finite number of terms.

Performance Bound Computation

Once the indexability of subproblem (4.5) is ensured by Theorem 4.1 and having proposed a tractable procedure to compute its optimal value given any λ (i.e. the optimal active set $B^*(z)$ contains those information states x such that $\lambda^*(x) - \lambda \geq 0$), we can solve the Lagrangian dual problem (3.7) stated as

$$V_D^d(\boldsymbol{x_0}) = \min_{\lambda \geq 0} \sum_{n=1}^{N} \left[\max_{\pi_n \in \Pi_n} f(x_{n,0}, \pi) - \lambda g(x_{n,0}, \pi) \right] + \lambda \frac{M}{(1-\beta)} \tag{4.29}$$

Hence, we may use $V_D^d(\boldsymbol{x_0})$ as a upper bound on the best attainable performance for problem (4.4). In the next section we will compute such a bound for the simulated scenarios considered and use it to evaluate the suboptimality gap of our proposed policy and other possible heuristics.

Life was always a matter of
waiting for the right moment to act.
Paulo Coelho

Chapter 5

Computational Experiments

In this chapter we clarify and extend the ideas on the elusive target hunt MARBP presented in Chapter 4. First, we discuss, through a series of computational experiments, index tractability, the validity of PCL-indexability conditions and of theorem Theorem 4.1, and relative and absolute performance of Whittle's MP index policy. Throughout the analysis, we will seek to draw insightful interpretations of the results in terms of the search problem of concern.

5.1 Index Evaluation

As an example of the use of our index computation method, we have simulated 10^3 runs of a scenario involving a target instance with the following parametric specification: $q^{(0)} = 0.1$, $p^{(0)} = 0.5$, $\rho^{(0)} = 1 - p^{(0)} - q^{(0)}$, $q^{(1)} = 0.5$, $p^{(1)} = 0.3$, $\rho^{(1)} = 1 - p^{(1)} - q^{(1)}$, $R = 1$, and $\alpha = 0.05$. The fixed points dividing the state space $\mathbb{X}^{0,1} \triangleq (0, 1]$ into the three analyzed threshold cases are $\phi^{(1)}_{\infty} = 0.3043$ and $\phi^{(0)}_{\infty} = 0.8333$. The discount factor β varied over the range $\beta \in \{0, 0.1, 0.2, \dots, 0.9, 0.99\}$ and the critical discount factor is in this case $\beta^{(1)} = 0.7468$.

The index was computed using a MATLAB script for index evaluation based on the expressions (4.17), (4.23), (4.28). For each β, the index $\lambda^*(x)$ was evaluated on a grid of x information state values of width 10^{-2} and the infinite sums of cases I and II were approximately evaluated by truncating them to $T = 10^4$.

Figure 5.1 plots the results. As required by the PCL-indexability conditions, in each case the index $\lambda^*(x)$ is monotone nondecreasing in x. Note that the index is continuous in x and piecewise differentiable and it converges as $\beta \nearrow 1$ to a limiting index that can be used for the expected total criterion. For each x the time required to compute the index is negligible.

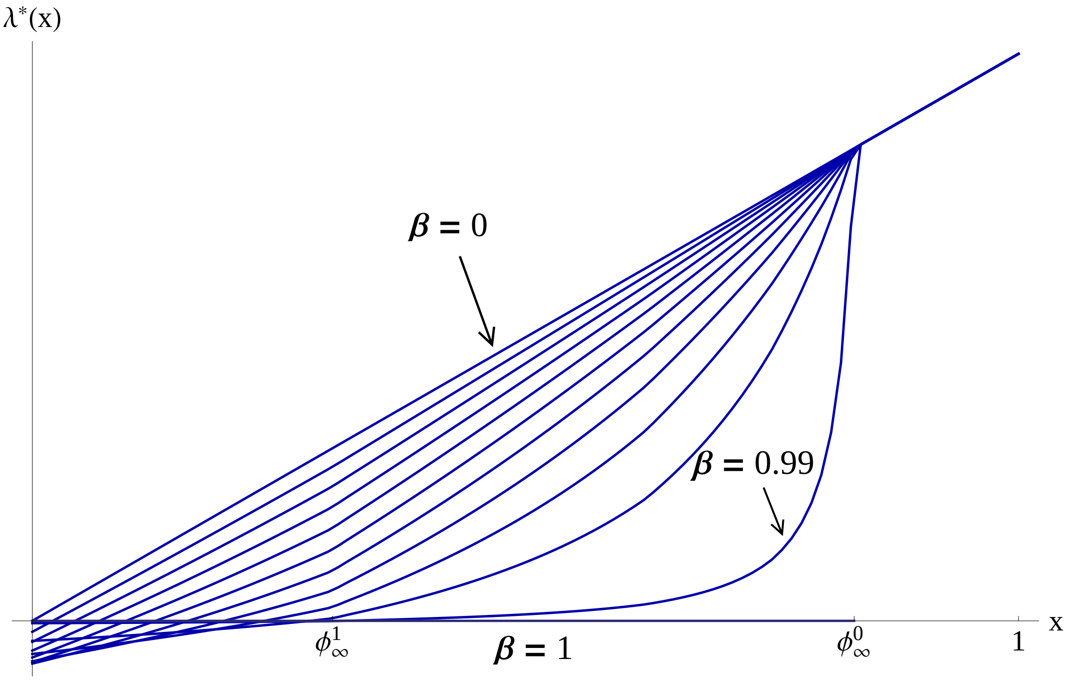


Figure 5.1: Whittle MP index for different discount factors β

From Figure 5.1 we derive the following relevant conclusions regarding the intuition of the optimal search policy for one elusive target in isolation.
For small enough x, (i.e., for $x \le \phi^{(1)}_\infty$) the index $\lambda^*(x)$ may be negative for large values of β, reflecting the fact that it is unproductive to search a site when it is very unlikely that the target is visible, as both actions result in an increased probability that it is exposed (further, this increase is larger if we do not search for it).

For x within the *absorbing* set of states ($\phi^{(1)}_\infty \le x \le \phi^{(0)}_\infty$), as $\beta \nearrow 1$, the marginal profit of searching the target practically vanishes. This reflects the fact that as the system's lifetime grows, it becomes counterproductive to try to hunt a target which is unlikely to be exposed, as doing so will only drive the target into hiding, delaying the hunt.
By the same reasoning, the fact that the $\lambda^*(x)$ is decreasing in the discount factor β within the *absorbing* set $\phi^{(0)}_\infty$, suggests that as the moment in which the target is hunted is less important, then the best search strategy is to let the target be unsensed so that its probability of being exposed raises (up to its maximum value if $\beta = 1$), and only then attempt to hunt it. In simpler terms, if we have enough time to hunt the target, it is best to wait for the moment in which it becomes the most likely to be exposed, and only then try to hunt it. For larger values of x (i.e., for $x > \phi^{(0)}_\infty$), it is optimal to behave myopically, since in those states the target is most likely to be exposed, yet those states are only *transient*. (See, Figure 4.2).

5.2 PCL-indexability

As required by the PCL-indexability condition (ii), Figure 5.1 shows that in each case the index $\lambda^*(x)$ is monotone nondecreasing in x (in fact, it is strictly increasing in x). This section reports some computational evidence on the validity of condition (i), regarding the positivity of the marginal work measures, considering 10^3 runs of the target instance analyzed in Section 5.1.

Figure 5.2 shows the results of computing the marginal work measure $w(x,z)$ fixing the z threshold value in $\{0.05, 0.5, 0.85\}$ and letting x vary in $\mathbb{X}^{0,1}$, analyzing a z value for each of the possible threshold cases described in Section 4.3.1. The discount factor β varied over the range $\beta \in \{0, 0.1, 0.2, \ldots, 0.9, 0.99, 0.999\}$. For each β and z, the index $w(x,z)$ was evaluated on a grid of x values of width 10^{-2} and the infinite sums of cases I and II were approximately evaluated by truncating them to $T = 10^4$. Figure 5.2 illustrates how $w(x,z)$ differs for each threshold case considered. Further, notice that in these examples of case I ($z = 0.05$) and case II ($z = 0.5$), the marginal work measure positivity condition only holds for $\beta \leq 0.8$.

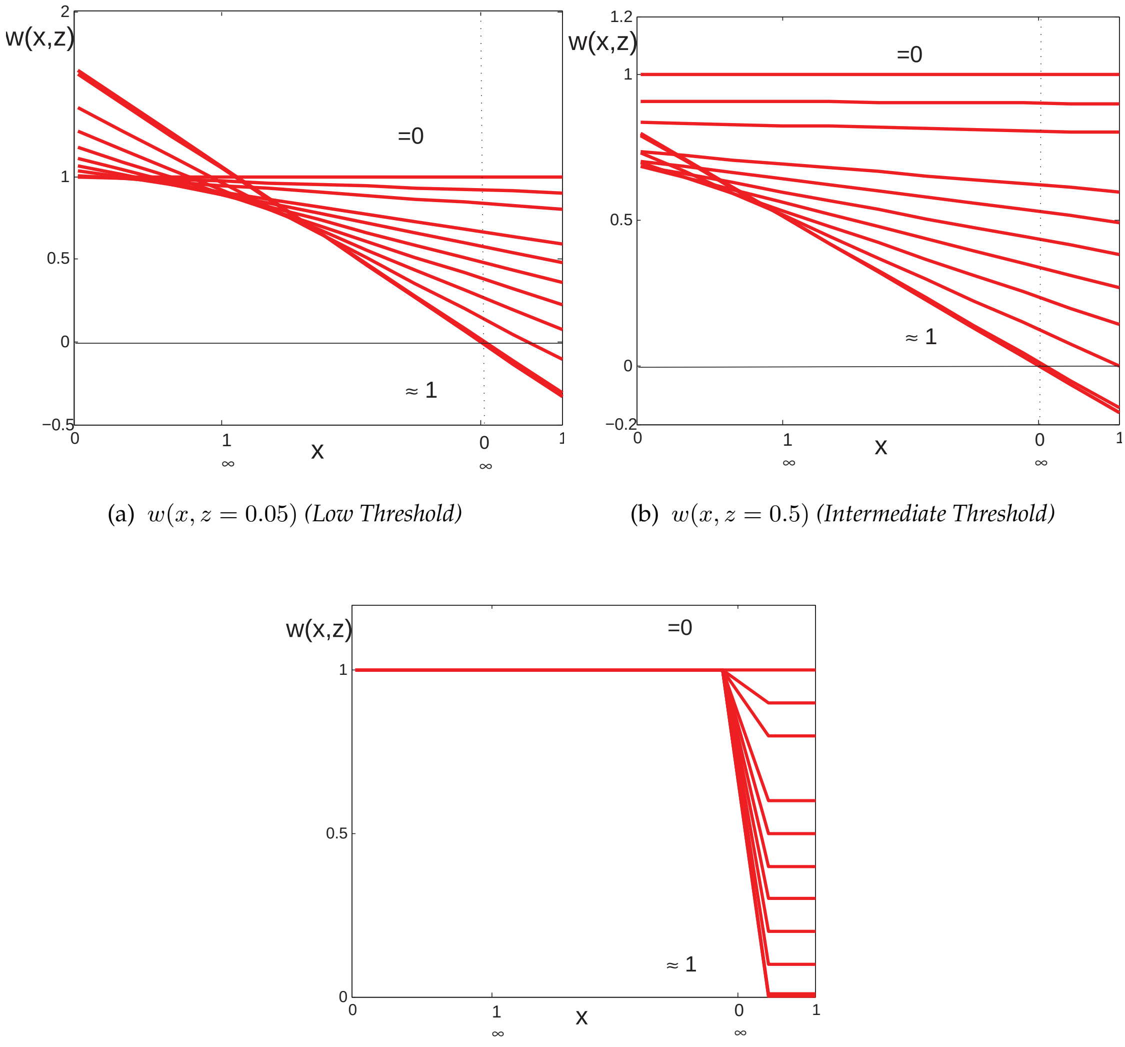
Notice that these simulation results are in accordance with the indexability analysis described in Chapter 4 and Appendix B. Also, in light of the interpretations provided in Section 5.1, note that since the target never returns to the largest values of x (i.e., $x > \phi_\infty^{(0)}$), then the total expected search effort to hunt it will be larger (in time) if we miss the opportunity to hunt it in those states than if we do not. Hence, the marginal work measure becomes negative for this range of x as the time horizon of the search increases.

5.3 Alternative Index Policies

In this section we define some alternative heuristics for the elusive target hunt MARBP (4.4) as stated in subsection 4.2.1. In the following section we will report simulation studies that compare the performance of Whittle index policy against these simpler alternatives.

The Myopic Index Policy

The *myopic* policy is based on index $\lambda^{Myopic}(x) = R(x,1)$ for all $x \in \mathbb{X}^{0,1}$, as defined in (2.12). Notice from Figure 5.1 that this index also corresponds with Whittle index $\lambda^*(x)$ for the case $\beta = 0$ and also for all discount factors β when x is in the range $(\phi_\infty^{(0)}, 1]$.



(a) $w(x, z = 0.05)$ *(Low Threshold)*

(b) $w(x, z = 0.5)$ *(Intermediate Threshold)*

(c) $w(x, z = 0.85)$ *(High Threshold)*

Figure 5.2: Marginal work measure for the z-threshold cases

The Belief State Index Policy

The *belief state* policy is based on index $\lambda^{B}(x) = x$, for all $x \in \mathbb{X}^{0,1}$, as defined in (2.13). At this point, it is worth pointing out that since $\lambda^{Myopic}(x)$, $\lambda^{B}(x)$ and $\lambda^{*}(x)$ are monotone increasing functions of the information state x, in instances of identical targets the three policies result in equivalent sensing decisions, as the higher the information state the greater the priority a target receives under all search rules.

The Random Search Policy

The *random* selection policy is based on picking a site to search (among the ones that contain an uhunted target) at random, with each site having the same probability of being selected.

5.4 Bechmarking the Whittle Index Policy

In this section we report on some small-scale preliminary simulation studies we have performed. The studies are based on MATLAB implementations we have developed to compare the performance of the proposed Whittle index policy against the the *myopic* policy, the *belief state* policy, and the *random* selection policy for the elusive target hunt MARBP model proposed in Chapter 4.

Further, we have computed an upper bound on the optimal value (4.4) based on the ideas discussed in Section 3.1.

5.4.1 Cautious and Reckless targets

In this experiment we assess the relative performance of the Whittle index policy against the other heuristics distinguishing target instances between *reckless* and *cautious*. We call reckless those targets which "*after not being searched, are highly likely to expose themselves*", i.e. with $p^{(0)} \approx 1$, while cautious targets display the opposite behavior, i.e. with $p^{(0)} \approx 0$ (while having $p^{(0)} > p^{(1)}$).

Each base instance has a single sensor $M = 1$ for searching within $N = 30$ sites, in one instance all targets are reckless with $p_n^{(0)} = 0.95$, while in the other instance all targets are cautious with $p_n^{(0)} = 0.35$. In both instances, $p_n^{(1)} = 10^{-3}$, $q_n^{(1)} = 0.97$, $q_n^{(0)} = 0.003$, $\alpha_n = 0.30$ and $R_n = 1$ for all n. Thus, for both targets $\phi_{\infty}^{(1)} = 0.0010$ while for reckless $\phi_{\infty}^{(0)} = 0.9694$ and for cautious $\phi_{\infty}^{(0)} = 0.9211$.

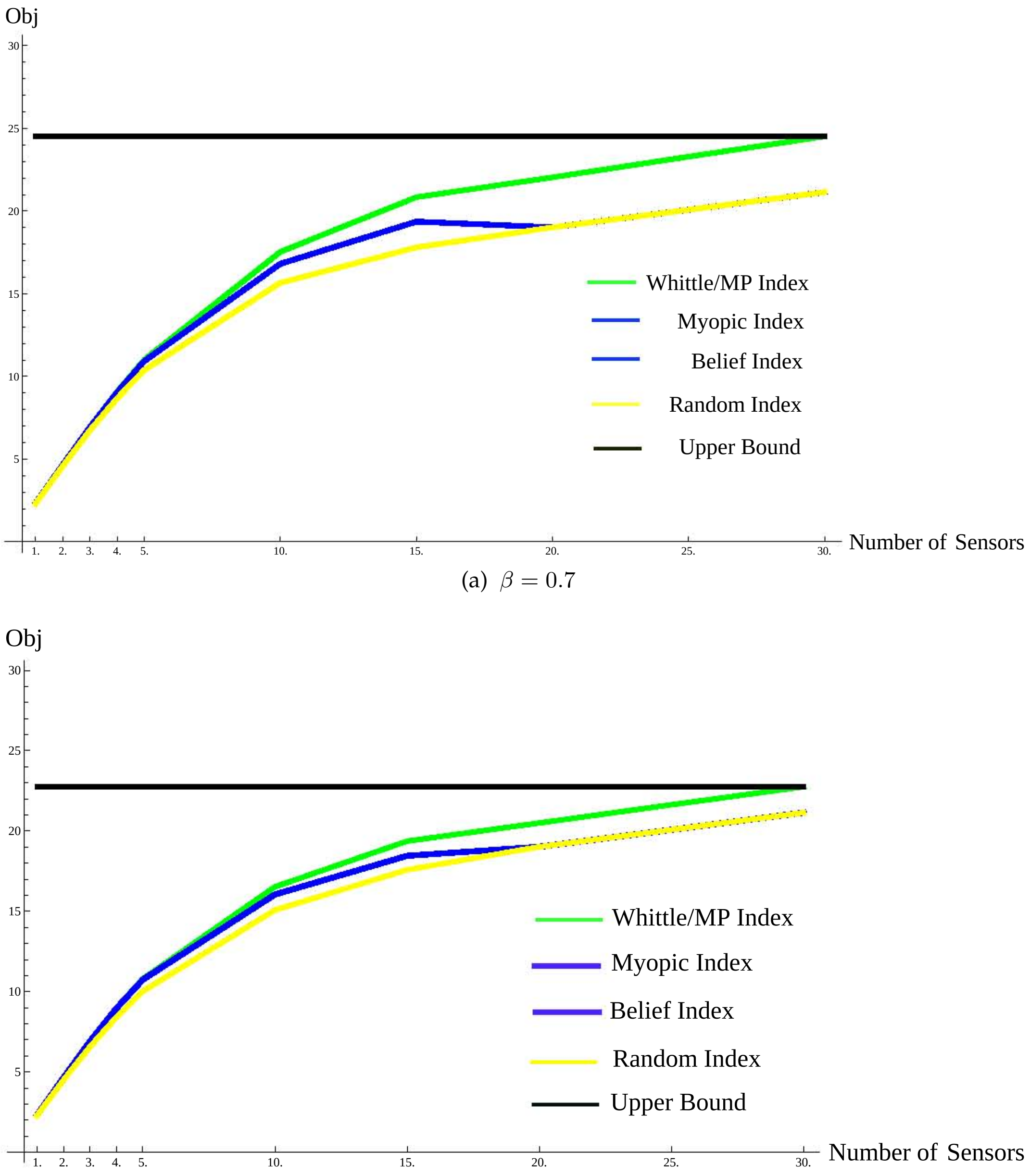
We take the initial state $x_n = 1$, which corresponds to exact knowledge of N exposed targets at the start of the search. Sensing costs were taken to be zero and we consider two possible discount factors $\beta \in \{0.7, 0.99\}$, where β^* is equal to 0.7688 both for the reckless and cautious instance. Both base instances were modified, letting the number of sensors increase from $M = 1$ up to $M = N = 30$. For each instance, 10^3 independent runs were performed on a horizon of $T = 10^4$ time slots.

Figure 5.4 shows the ETD net rewards under each policy as the number of sensors in the network grows. The upper bound from the relaxation for all the instances with reckless targets was of 24.510 and 29.735 for discount factors 0.7 and 0.99, respectively, whereas for cautious targets those values were 22.767 and 29.374. Note that the bound on the best result of the search is always less when targets are cautious, since they are harder to hunt.

As depicted by Figure 5.4, the Whittle policy outperforms other heuristic policies for any number of sensors with the performance improvement increasing as $M \nearrow N$. In fact, as the number of sensors grows all policies perform worse, except for the Whittle policy for which the opposite occurs. The explanation of this results is that all other heuristics *overuse* the network resources as they become available, searching more and more sites possibly containing a target, thus making targets more elusive and hence, more difficult to hunt. This is a salient result, since it points out a severe drawback that myopic or simpler heuristic have for allocating resources in cases in which idling is expected to have a greater impact on the system's expected returns.

Another interesting result is that the Whittle index policy's suboptimality gap tends to 0 for a relatively small number of sensors when $\beta \approx 1$, while the largest sensor network size is required for the Whittle policy to be nearly optimal for smaller β (i.e. when hunting targets is urgent). Such a result is related to the fact that all policies successfully find the N targets, yet they differ significantly in the time they take to do so. Thus, if the hunt mission is urgent a large sensor network (operated under the Whittle index policy) will result nearly optimal whereas if the mission is just to find the objects but not urgently a relatively small sensor network is required.

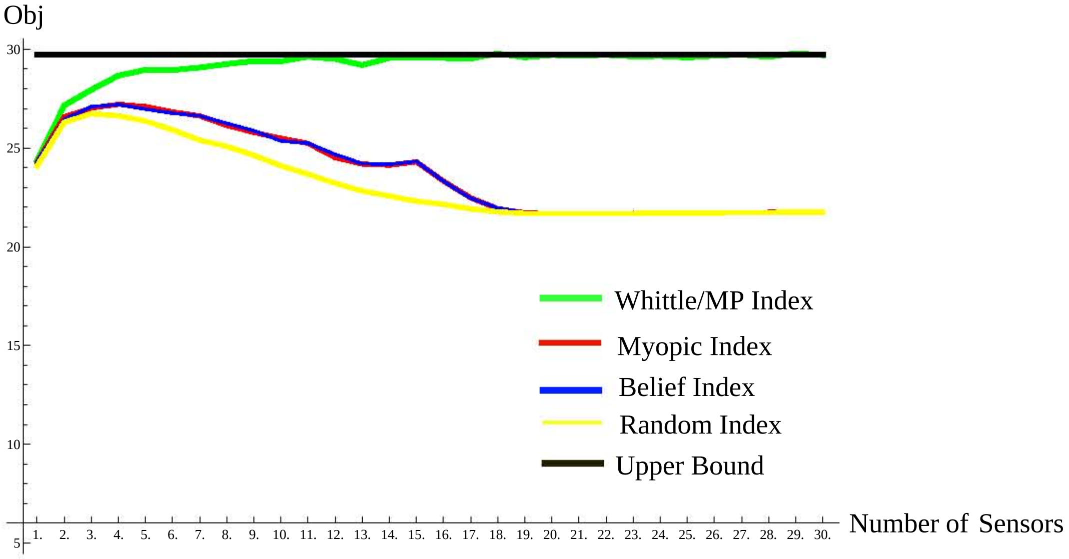
Table 5.1 shows the average time that the system takes to hunt all targets operated under each policy. Such results illustrate the fact that a large sensor network which is constantly searching will spend a larger period of time to hunt targets. However, all policies succeed at finding the N targets at some period. The Whittle index policy takes significantly less time to hunt targets than the alternative polices for both Reckless and Cautious targets, yet hunting the Cautious targets naturally takes longer for all policies. These results also show the overuse under other heuristics since their average operating time substantially increases as the number of sensors grows. Results in Table 5.1 are of particular relevance in terms of the specific motivating application proposed in Section 4.1 and investigated in Rucker (2006). The main goal in that case was to have



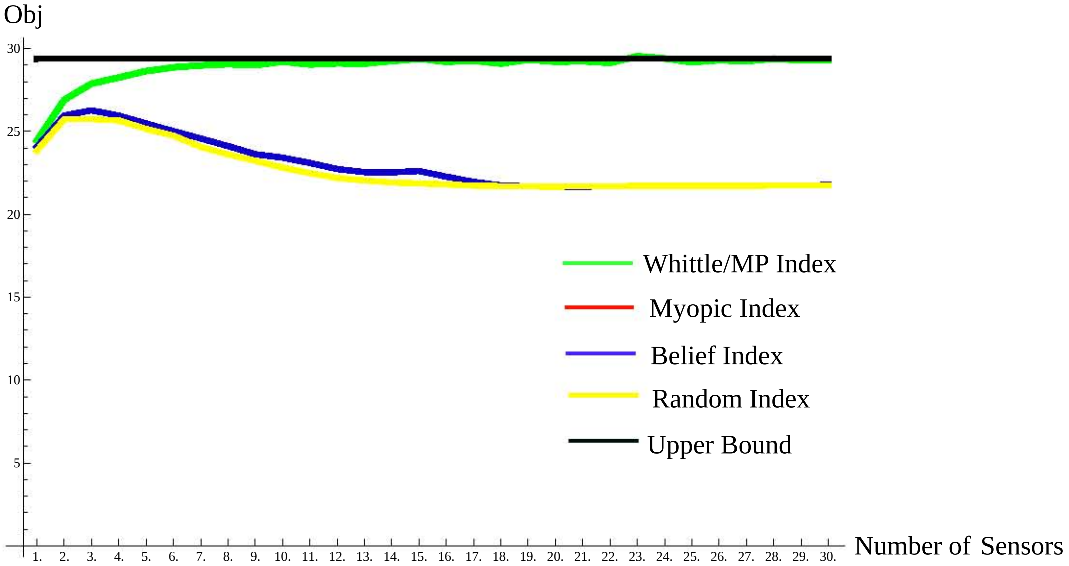
(a) $\beta = 0.7$

(b) $\beta = 0.7$

Figure 5.3: Experiment 1 - (5.3a) & (5.3b), *Reckless and Cautious Targets* instances



(a) $\beta = 0.99$



(b) $\beta = 0.99$

Figure 5.4: Experiment 2 - (5.4a) & (5.4b), *Reckless and Cautious Targets* instances

Table 5.1: Average System's time to hunt all targets

M / *Reckless*	$\bar{T}^{MP}$	$\bar{T}^{My}$	$\bar{T}^{B}$	$\bar{T}^{R}$
1	6.175	9.768	7.083	31.039
2	5.778	18.994	12.021	42.475
3	2.405	49.074	45.718	95.318
4	3.344	36.678	33.071	70.652
5	3.034	90.074	76.949	102.643
15	1.928	122.554	155.053	371.901
30	1.924	373.586	366.239	458.258
M / *Cautious*	$\bar{T}^{MP}$	$\bar{T}^{My}$	$\bar{T}^{B}$	$\bar{T}^{R}$
1	10.770	43.833	37.630	44.864
2	6.384	29.845	33.414	80.638
3	4.841	66.301	55.581	87.306
4	3.828	74.822	76.426	138.993
5	3.970	135.726	86.134	182.447
15	3.277	281.945	264.044	410.706
30	3.073	448.593	465.127	423.586

a scheduling policy which minimizes the average time until all missile launchers are detected and destroyed. As Table 5.1 shows, the Whittle policy is the heuristic that manages to find all targets in the least time.

To sum up, the proposed policy is always as good as the other heuristics, yet in many instances it does yield important performance improvements. These performance improvements of the Whittle policy are significant from a statistical point of view,and from a practical point of view (performance gains can be up to 36,48%). Further, the performance improvements become more important as the size of the sensor network increases. In fact, the Whittle policy is even nearly optimal in both scenarios when $M \nearrow N$. Also, the Myopic and the Belief policy are not significantly different in these scenarios, nor do they improve significantly on the random policy. Further, all the policies produce the same results when $M = 1$, basically because they are all equally forced to not search the remaining unhunted targets. Performance differences are observed when the system has the possibility of searching a site and a index policy prescribes not to do so.

5.4.2 Sensing Costs

In this experiment we assess the relative performance of the Whittle index policy against the other heuristics as the sensing cost c increases. We consider two base instances of $N = 10$ sites with $M = 1$ and $M = 5$ sensors. In both instances targets parameters are: $p_n^{(1)} = 10^{-3}$, $q_n^{(1)} = 0.97$, $p_n^{(0)} = 0.05$, $q_n^{(0)} = 0.003$, $\alpha_n = 0.30$, $x_n = 1$, $\beta = 0.99$ and $R_n = 1$ for all n. Both base instances were modified, letting sensing costs for all sites vary as $c \in \{0, 0.3, 0.5, 0.75\}$. For each instance, 10^3 independent runs were performed on a horizon of $T = 10^4$ time slots.

Figure 5.5 plots the ETD net rewards under each policy and the upper bound as c grows. Results show that the Whittle's MP index policy outperforms the other policies in all instances. The random policy performs significantly worse in this case, basically because it prescribes to search sites, provided there are enough sensors, regardless of the sensing cost, while the other two heuristics have been defined in such a way that they prescribe to search only if their index value exceeds c.

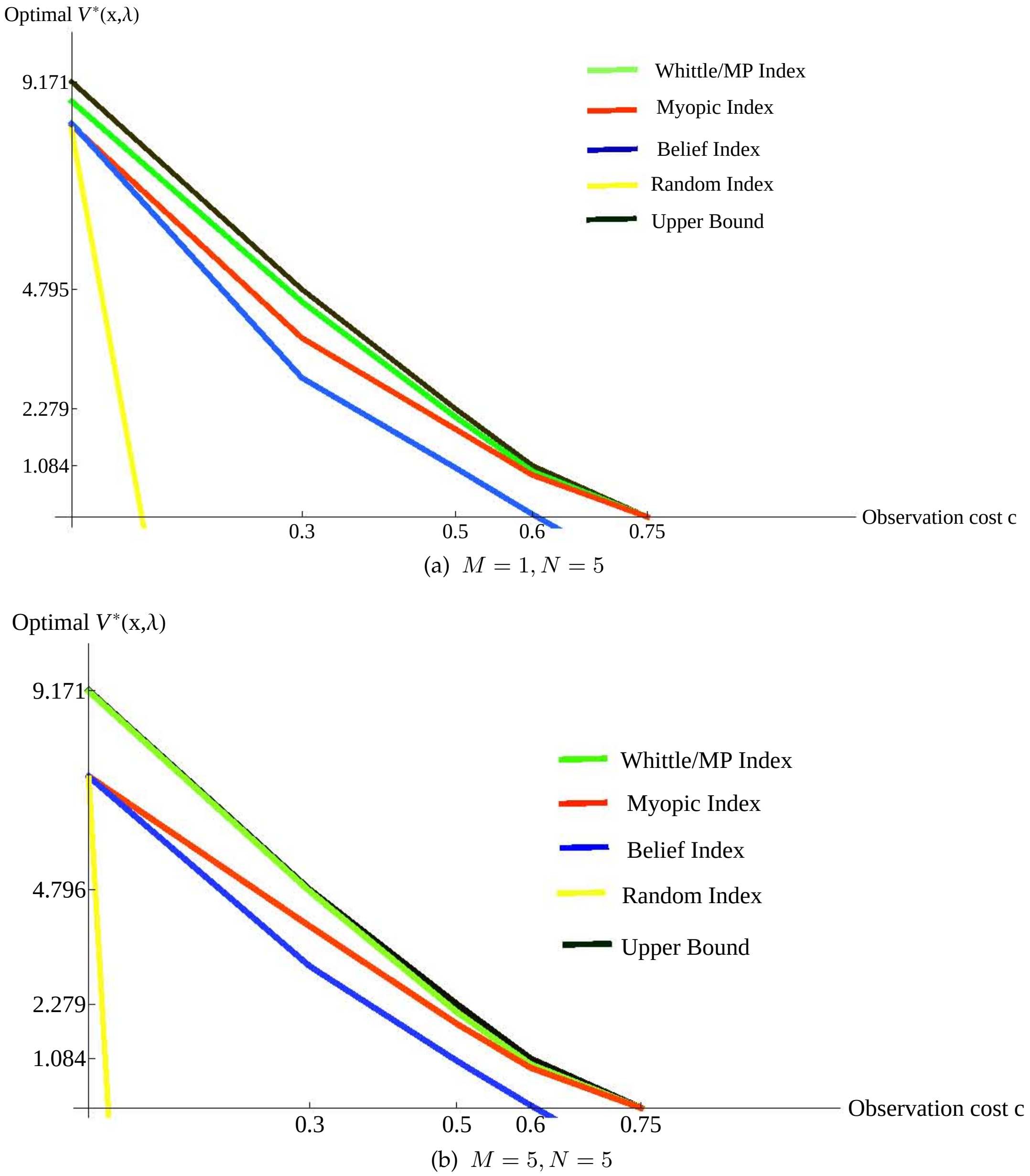
Naturally, as searching becomes expensive, both the resulting performance under all policies and its upper bound decrease. In the Figure 5.5 we observe that the system yields 0 rewards for $c > 0.75$. Actually, the optimal value function vanishes when $c = R(1, 1)$, which in this case is $c = 0.7$.

Notice that the Whittle policy is nearly optimal for all values of the sensing cost when $M = 5$ while the suboptimality gap of the other heuristics is larger for $M = 5$ than for $M = 1$, a result consistent with the overuse of the simpler heuristics pointed out before.

5.4.3 Sensor Network Size

Perhaps one of the most notorious results obtained, with special consequence for the design of sensing systems for hunting such elusive targets, is that if the horizon is long enough (i.e. as $\beta \nearrow 1$), operating a system's under the Whittle index policy requires a few sensors to optimally hunt a larger set of targets. In the instances plotted in 5.4a we observe that a sensor network of $M \approx 12$ or more sensors is enough to achieve the best possible expected reward under the Whittle'index policy provided that targets are reckless. If targets are cautious, as in 5.4b, a sensor network of $M \approx 8$ is enough to achieve optimality, as the system spends less time actively searching targets.

The results also suggest that if the optimal scheduling policy is not tractable, and we are forced to operate the system under a simple heuristics, if we define heuristics which do not advise the system to idle, it will take longer to find all targets. Thus, for this



(a) $M = 1, N = 5$

(b) $M = 5, N = 5$

Figure 5.5: Experiment 2: Sensing Cost Effect with: $M/N = 1/10$ (5.5a) and $M/N = 1/2$ (5.5b)

kind of problems it makes more sense to define heuristics of *round-robin* type, specifying how to alternate between searching and not searching a target, as the Whittle index does, than myopically operating the system.

Part III

Multitarget Tracking

Every step you take,
I'll be watching you.
The Police.

Chapter 6

MARBP Formulation for a Multitarget Tracking Kalman Filter Model

In this chapter we formulate problem 2, stated in subsection 1.2.2, as a real-state MARBP and we deploy the indexation methodology revised in Chapter 3 to propose a tractable heuristic search policy of priority-index type based on Whittle index for RBs.

6.1 Motivation and Prior Work

As reviewed in Section 1.2, recent advances in sensor technology have provided modern multi-sensor systems with an increased operating flexibility to achieve given performance objectives through the development of appropriate *scheduling algorithms*. The widespread adoption of these cutting-edge technologies has stimulated this demand and has ultimately matured into an emerging field of research: *sensor management*.

A concrete example of *sensor management* problems posed by the introduction of an advanced sensing technology is given by the active electronically scanned *phased-array* radar. Typical pulse radar systems operate by illuminating a scene with a short pulse of electromagnetic energy and collecting the energy reflected from the scene. In contrast to traditional radar systems, in which illumination parameters, such as beam direction and shape among others, are typically hard-wired, *phased-array* radars are capable of electronically controlling these parameters during system operation so as to best extract information from the scene. Naturally, efficient usage of these flexible sensing resources requires the scheduling of transmission parameters so as to optimize the system's performance. See Moran et al. (2008) for a survey of the substantial literature in the area.

Phased-array radars, operating in the *tracking or revisit* mode, maintain targets' location estimates by steering the radar's beam to point toward desired directions, as opposed to conventional track-while-scan radars, which track targets while the radar's antenna mechanically rotates at a constant rate. In this context, appropriately switching beam direction (which implies the monitoring of the total energy intercepted by a measured target) raises the possibility of improving tracking performance via the design of a suitable *scheduling control policy* adopted for dynamic prioritization of target track updates.

Early work on the subject of optimal scheduling of track updates in phased array radars dealt with the minimization of radar energy required for track maintenance, see, e.g., van Keuk and Blackman (1993), Stromberg (1996), Hong and Jung (1998). The design of optimal target track updates scheduling policies in highly idealized system models which ignore other relevant issues as target detection, waveform selection, and control of the Pulse Repetition Interval (PRI) is addressed in recent work. In Krishnamurthy and Evans (2001), a beam scheduling algorithm is derived from a discrete-time and discrete-state POMDP model which assumes that targets' motion from one PRI to the next is negligible (i.e. targets are *stationary*). Exploiting the special structure of the suggested POMDP as a classic MABP, the optimal policy is characterized in terms of an index policy.

A discrete-time finite horizon formulation for *non-stationary* targets in which targets' motions and targets' track measurements follow scalar, linear Gauss–Markov dynamics, and target Tracking Error Variances (TEVs) are updated via Kalman filter's equations is introduced in Howard et al. (2004). The authors seek to to optimize the sum of the targets' track error variances over a finite horizon and propose a *myopic* scheduling policy, which updates at each time a target of largest TEV, thus taking a target's current TEV as its *priority index* . They further claim such a *myopic-index* policy to be optimal in for the case of two symmetric targets, yet no full proof of such result is provided. The inadequacy of the classic model is also pointed out in La Scala and Moran (2006), where the authors extend the results in Howard et al. (2004) on optimality of the greedy-index scheduling policy for tracking two symmetric targets to more general linear dynamical systems under the same finite-horizon total TEV performance objective.

6.1.1 Goals and Contributions

It is the goal of this work to derive a multitarget track update heuristic scheduling policy for a model that extends the one formulated in Howard et al. (2004), in which targets' motions and targets' track measurements follow scalar, linear Gauss–Markov dynamics, and target TEVs are updated via Kalman filter's equations. Further, we aim at proposing a policy which is dynamic, readily implementable and further exhibits a near-optimal performance both under the discounted and the total criterion.

We accomplish this goal by formulating the proposed extended model as a real-state MARBP and deploying the recent extensions of the existing theoretical and algorithm results on discrete-state RB indexation to the real-state case.

This work makes the following contributions. It investigates a MARBP formulation of the dynamic problem of tracking multiple asymmetric targets with scalar linear Gauss–Markov dynamics, which incorporates both tracking-error and measurement (energy) costs, to obtain a tractable index policy that performs well based on restless bandit indexation.

It further successfully deploys the methodology announced in Niño-Mora (2008) to obtain a novel and dynamic index policy for the model of concern, which is both non-myopic and depends on the target's initial TEV as well as on its motion and measurement dynamics' specific parameters. The PCL-indexability of the model is shown for the ETD problem and it is extendable to the LRA problem. All this is done despite the lack of closed form expressions for the required performance measures, which is a severe technical difficulty introduced by considering a real state variable.

These contributions will be presented in this chapter in the following order: first we describe the model and we formulate it as a real-state MARBP. Next, in the remainder of the chapter we discuss the indexability analysis and the resulting index computation method. In the subsequent chapter, we present the computational results obtained which demonstrate the tractability of index evaluation, the substantial performance gains that the Whittle index policy achieves against myopic policies advocated in previous work as well as the resulting index policies suboptimality gaps.

6.2 Model description and MARBP Formulation

We consider the tracking of N moving targets labeled by $n \in \mathsf{N} \triangleq \{1, \ldots, N\}$ by means of a sensing system composed of M phased array radars. All radars in the system are synchronized to operate over time slots $t = 0, 1, \ldots$, where a time slot corresponds to a PRI. The system is controlled by a *central coordinator*, who at each slot t must decide to update the tracks of at most M targets by steering toward them the beams of as many radars to measure their positions.

As in Howard et al. (2004) and La Scala and Moran (2006), we assume that there are no clutter or false measurements, and that the probability of target detection is unity. For simplicity we also assume that targets move in one dimension. Let $x_{n,t}$ be the (unobservable) *position* of target n in the real line $\mathbb{R}$ at the beginning of slot t. If a radar measures target n's position in slot t, a noisy *measurement* $y_{n,t}$ is obtained. Decisions on which target tracks to update at each time are formulated by binary *action* processes $a_{n,t} \in \{0,1\}$, where $a_{n,t} = 1$ if target n is measured in slot t and $a_{n,t} = 0$ otherwise.

The targets move over $\mathbb{R}$ following independent linear Gauss–Markov dynamics

$$x_{n,t} = F_n x_{n,t-1} + \omega_{n,t}, \quad t \geq 1, \tag{6.1}$$

where the *position-noise process* $\omega_{n,t}$ is an i.i.d. zero-mean Gaussian white noise with variance q_n, and F_n is a fixed constant in $\mathbb{R}$.

At a slot t in which target n is measured, the corresponding measurement $y_{n,t}$ is generated by the following linear Gauss–Markov dynamics

$$y_{n,t} = H_n x_{n,t} + \nu_{n,t}, \tag{6.2}$$

which is target specific but independent of the radar being used (as radars are assumed homogenous), and where the *measurement-noise process* $\nu_{n,t}$ is an i.i.d. zero-mean Gaussian white noise with variance r_n, and $H_n \in \mathbb{R}$.

Although our approach applies to arbitrary parameters F_n and H_n, for simplicity of exposition we will focus the subsequent discussion on the case $F_n = 1$ and $H_n = 1$.

If an initial estimate of the position and of the Tracking Error Variance (TEV), denoted by $\widehat{x}_{n,0}$ and $p_{n,0}$, respectively, are given for each target n, then, as discussed in Section 1.3, the optimal minimum-variance predicted estimates are given by the Kalman filter. The TEV $p_{n,t}$, which describes the uncertainty in target n's track at the beginning of slot t, is recursively updated by the Kalman equations

$$p_{n,t} = \begin{cases} p_{n,t-1} + q_n, & \text{if } a_{n,t} = 0 \\ \dfrac{p_{n,t-1} + q_n}{p_{n,t-1}/r_n + q_n/r_n + 1}, & \text{if } a_{n,t} = 1 \end{cases} \tag{6.3}$$

To take actions $a_{n,t}$, the coordinator follows a *scheduling policy* $\boldsymbol{\pi}$, which is drawn from the class $\mathbf{\Pi}(M)$ of *admissible scheduling policies* that are nonanticipative (based on the history of states and actions) and measure at most M targets per time slot,

$$\sum_{n=1}^{N} a_{n,t} \leq M, \quad t = 0, 1, 2, \ldots \tag{6.4}$$

We assume that a radar which updates the target n's track in a time slot incurs a *measurement cost* $c_n \geq 0$, representing the cost of beam energy expended for the track's update. Further, we take the *tracking-error cost* at slot t to be $d_n\, p_{n,t+1}$, where $d_n > 0$ is a constant that may differ by target. The flexibility furnished by the d_n will be of use if the relative importance of tracking precision differs across targets. Hence, the one-slot cost incurred by taking action a on target n when it occupies *tracking error variance* $p_{n,t}$ is $C_n(p_{n,t}, a) \triangleq d_n\, p_{n,t+1} + h_n a$.

6.2.1 MARBP Formulation

First, we shall take the *state* of each target n to be its Scaled Tracking Error Variance (STEV) $s_{n,t} \triangleq p_{n,t}/r_n$, which follows the dynamics

$$s_{n,t} = \begin{cases} \phi_n^{(0)}\big(s_{n,t-1}\big), & \text{if } a_{n,t} = 0 \\ \\ \phi_n^{(1)}\big(s_{n,t-1}\big), & \text{if } a_{n,t} = 1, \end{cases}$$

where

$$\phi_n^{(0)}(s) \triangleq \theta_n + s, \quad \phi_n^{(1)}(s) \triangleq \frac{\theta_n + s}{1 + \theta_n + s} \tag{6.5}$$

and $\theta_n \triangleq q_n/r_n$ is the *position to measurement noise variance ratio* for target n.

Notice that such a STEV state, being a scaled variability measure of target's n current position estimate, naturally moves over a *state space* with is a subset of $\mathbb{R}_+$. In fact, it holds that $\mathbb{S} \triangleq [0, \infty)$, where $\phi_\infty^{(1)}$ is the minimum STEV achieved after uninterrupted measurement. Hence, for some $r_n \in (0, \infty)$, $s_{n,0} = 0$ corresponds to exact knowledge of the targets' initial positions and $s_{n,0} = \infty$ to complete uncertainty of the targets' initial positions. Thus, $\mathbb{S}$ results from the union of the transient state 0 and the absorbing set of states $[\phi_\infty^{(1)}, \infty)$.

As alleged by Whittle in Whittle (1988) when describing the *submarine surveillance* example, the passive and active dynamics, $\phi_n^{(0)}(s)$ and $\phi_n^{(1)}(s)$, result in contrary movements in the state space $\mathbb{S} \triangleq [0, \infty)$, which respectively correspond to loss and gain of precision on target n's location estimates.

Hence, the one-slot cost incurred by taking action a on target n when it occupies STEV state s is $C_n(s, a) \triangleq d_n r_n \phi_n^{(a)}(s) + h_n a$.

6.2.2 Performance Objectives

Given that there is no natural stopping time specified a priori for the tracking system, we will consider the following infinite horizon dynamic optimization problems:

(1) find a *β-discounted reward optimal* policy, i.e.,

$$\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(M)} \mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty}\sum_{n=1}^{N}\beta^t C_n\left(s_{n,t},a_{n,t}\right)\right], \tag{6.6}$$

where $0<\beta\leq 1$ is the discount factor, $\mathrm{s}_0=(s_{n,0})_{n=1}^N$ is the initial joint TEV state, for n in $\{1,2,\ldots,N\}$, and $\mathsf{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}}[\cdot]$ denotes expectation under policy $\boldsymbol{\pi}$ conditioned on the initial joint state being equal to s_0; and (2) find an *average-optimal* policy,

$$\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(M)} \limsup_{T\to\infty}\frac{1}{T}\mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{T-1}\sum_{n=1}^{N}C_n\left(s_{n,t},a_{n,t}\right)\right], \tag{6.7}$$

which minimizes the expected long-run average cost.

Problems (6.6) and (6.7) are discrete-time MARBP with real-state projects. Each project feeds on the limited sensing system's resources and it is modeled as a binary-action MDP whose STEV state $s_{n,t}$ lives on the Borel state space $\mathbb{S}$. Note that, taking action $a_{n,t}$ on target n, with $a_{n,t}=1$: a beam is steered toward target n to measure its position; $a_{n,t}=0$: no beam is steered toward target n to measure its position, leads to the following consequences: (i) the tracking of target n results in a system cost $C_n\left(s_{n,t},a_{n,t}\right)$ per PRI, which describes the tracking accuracy for a given resource consumption $a_{n,t}$; and (ii) the target's next state $s_{n,t+1}$ is given by (6.5), which implies that, given $a_{n,t}$, state transitions are deterministic and independent across projects.

As reviewed in chapter Chapter 2, the existence of an optimal solution for a MARBP such as (6.6) is ensured under appropriate conditions on C_n and $a_{n,t}$, (cf. Hernández-Lerma and Lasserre (1999)). Moreover, such a solution is a *deterministic stationary policy* taken from the class $\Pi(M)$ of *admissible scheduling policies* and it is characterized by the corresponding Dynamic Programming Equationss (DPEs). Nonetheless, exact numerical solution to such DPEs is generally intractable due to specific difficulties introduced by its continuous state space. This computational infeasibility is also the case for the average-cost MARBP (6.7). Hence, instead of attempting to solve such problems optimally, we shall pursue the more practical goals of designing and computing a *well-performing* heuristic policy of *priority-index* type.

6.3 Real State RB Indexation Analysis

As reviewed in Chapter 3, with the purpose of establishing Whittle's indexability 3.1 by means of deploying sufficient indexability conditions 3.2, we focus on an individual site's subproblem. Thus, in the sequel we shall focus for concreteness on ETD-cost problem (6.6), although our approach also applies to LRA-cost problem (6.7).

$$\min_{\boldsymbol{\pi_n} \in \boldsymbol{\Pi_{(n)}}} \mathsf{E}^{\boldsymbol{\pi_n}}_{\mathbf{s_{n,0}}} \left[\sum_{t=0}^{\infty} \beta^t \{ C_n \big(s_{n,t}, a_{n,t} \big) + \lambda a_{n,t} \} \right], \tag{6.8}$$

where (6.8) is a single-project restless bandit subproblem, consisting of tracking target n in isolation. Π_n denotes the class of admissible policies for tracking a single target, i.e., deciding when the radar tracking it should be active ($a_{n,t} = 1$) and passive ($a_{n,t} = 0$) and with λ being a constant parameter representing the extra cost incurred per unit of time the radar is active.

Next, we would like to establish that each subproblem (6.8) has the key structural *indexability* property defined by 3.1. For such a purpose, we will deploy conditions 3.2 to establish that the problem is indexable with respect to the family of z-threshold policies, and thus we must start by computing the performance measures such policies. In the remainder of this section we focus on a generic single target subproblem as (6.8), and hence drop the superscript n from the above notation.

We recall from Chapter 3 that we can evaluate the performance of any tracking policy $\pi \in \Pi$ along two dimensions: the *work measure* $g(s, \pi)$ giving the ETD number of times the target is measured under policy π starting at $s_0 = s$; and the *cost measure* $f(s, \pi)$ giving the corresponding ETD cost incurred. Thus,

$$g(s, \pi) \triangleq \mathsf{E}^{\pi}_{s} \left[\sum_{t=0}^{\infty} \beta^t a_t \right], \quad f(s, \pi) \triangleq \mathsf{E}^{\pi}_{s} \left[\sum_{t=0}^{\infty} \beta^t C(s_t, a_t) \right].$$

So that we can formulate the single-targets's optimal tracking subproblem (6.8) in terms of these measures as

$$\min_{\pi \in \Pi} f(s, \pi) + \lambda g(s, \pi). \tag{6.9}$$

Once again, we shall consider problem (6.9), which is a real-state MDP, as target's λ-*charge subproblem*. Thus, its optimal policy belongs to the family of *deterministic stationary policies* Π^{SD}, naturally represented by their *active (state) sets* (in this case, that is the set of STEV states where measuring the target is prescribed). For an active set $B \subseteq \mathbb{S}$, we shall refer to the B-*active policy*. As in Section 4.3, we shall focus attention on the family of *threshold policies*. For a given *threshold level* $z \in \overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$, the z-*threshold policy* measures the target in STEV state s iff $s > z$, so its active set is $B(z) \triangleq \{s \in \mathbb{S} \colon s > z\}$. Note that $B(z) = (z, \infty)$ for $s \geq 0$, $B(z) = \mathbb{S} = [0, \infty)$ for $z < 0$, and $B(z) = \emptyset$ for $z = \infty$. We denote by $g(s, z)$ and $f(s, z)$ the corresponding work and reward measures.

For fixed z, work measure $g(s, z)$ is characterized as the unique solution to the functional equation

$$g(s,z) = \begin{cases} 1 + \beta g\big(\phi^{(1)}(s), z\big), & s \in (z, \infty] \\ \beta g\big(\phi^{(0)}(s), z\big), & s \in (0, z], \end{cases} \tag{6.10}$$

in the Banach space of bounded measurable real-valued functions on $\mathbb{S}$ endowed with the sup norm. Whereas cost measure $f(s,z)$ is characterized by

$$f(s,z) = \begin{cases} C(s,1) + \beta f\big(\phi^{(1)}(s), z\big), & s \in (z, \infty], \\ C(s,0) + \beta f\big(\phi^{(0)}(s), z\big), & s \in (0, z]. \end{cases} \tag{6.11}$$

We shall use the marginal counterparts of such measures. For threshold z and action a, denote by $\langle a, z\rangle$ the policy that takes action a in the initial slot and adopts the z-threshold policy thereafter. Define the *marginal work measure*

$$\begin{aligned} w(s,z) &\triangleq g(s,\langle 1,z\rangle) - g(s,\langle 0,z\rangle), \\ &= 1 + \beta\, g(\phi^{(1)}(s), z) - \beta\, g(\phi^{(0)}(x), z) \end{aligned} \tag{6.12}$$

and the *marginal cost measure*

$$\begin{aligned} c(s,z) &\triangleq f(s,\langle 0,z\rangle) - f(s,\langle 1,z\rangle). \\ &= (C(s,0) - C(s,1)) + \beta \Big(\ f(\phi^{(0)}(s), z) -\ f(\phi^{(1)}(x), z)\Big) \end{aligned} \tag{6.13}$$

If $w(s,z) \neq 0$, define further the *MP measure*

$$\lambda^{MP}(s,z) \triangleq \frac{c(s,z)}{w(s,z)}. \tag{6.14}$$

Total and Marginal Evaluation Measures

In order to analyze the PCL-indexability of (6.8) by means of the Sufficient Indexability Conditions (SIC) stated in definition 3.2 we must first solve the evaluation measures $g(x,z)$ and $f(x,z)$ for any fixed threshold $z \in \mathbb{R}$ and any $s \in \mathbb{S}$.

As already mentioned in Section 4.3.1, the main challenge to do so is posed by the fact that possible STEV state trajectories $\{S_t\}$ are naturally infinite, since S_t can take any value in a infinite non-denumerable set $\mathbb{S} \subseteq \mathbb{R}_+$ at each t. Therefore, we will take advantage of the fact that under a z-threshold policy for any initial state s, possible STEV state trajectories $\{S_t\}$ are infinite but numerable, as they exhibit recurrent cyclical patterns depending on the threshold level. Further, these patterns (in terms of possible active-passive cycles) are the same as the ones described for the model analyzed in Chapter 4.

Yet, as we will next show, the total and marginal cost measures do not converge to a simple closed-form expression. Next, we outline how to solve the evaluation measures to perform an indexability analysis of present model (analogous to the analysis done in Section 4.3) and further show how to use such solutions to evaluate the index $\lambda^{MP}(x)$.

To solve for (6.10) and (6.11) in closed form we further define $\phi_t^{(a)}(s)$ for $a = 0, 1$ as the recursion generated by letting $\phi_0^{(a)}(s) \triangleq s$ and $\phi_t^{(a)}(s) \triangleq \phi_0^{(a)}(\phi_{t-1}^{(a)}(s))$ for $a = 0, 1$. Note that for any $s \in \mathbb{S}$, both recursions $\phi_t^{(a)}(s)$ converge as $t \to \infty$ to the respective limits

$$\phi_\infty^{(0)} \triangleq \infty \quad \phi_\infty^{(1)} \triangleq \frac{1}{2}\left(\sqrt{\theta(4+\theta)} - \theta\right),$$

We recall that $\theta_n \triangleq q_n/r_n$ is the *position to measurement noise variance ratio* for target n.

Both of the resulting iterated mappings can be seen as (non-linear) functions known as Möbius transformations (or also as Linear Fractional Transformations). This observation was crucial to deriving the results of the subsequent the indexability analysis. The Möbius transformations properties and results deployed in this work are revised in Appendix A.

The definitions in (6.5) ensure that, as long as $\theta < \infty$, that $\phi_\infty^{(1)} < \phi_\infty^{(0)} = \infty$ (as shown in Lemma B.1 in Appendix B), and both limits are attractive fixed points of the active and passive dynamics respectively. This naturally divides the state space in two parts as portrayed in figure Figure 6.1 where the active and passive actions (depending on the initial STEV state s and the threshold z) produce movements in the state space, which are either both increasing (if $s, z \in (0, \phi_\infty^{(1)})$) or moving in opposite directions (if $s, z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$). Hence, we exploit this knowledge to solve the evaluation equations by distinguishing among two $z-$threshold cases, as discussed below.

Notice that the subset of states $\mathbb{S} \triangleq [\phi_\infty^1, \infty)$ is *absorbing* for target n. Note further that $\phi_\infty^1 \leq \theta$ iff $\theta \geq 1/2$, which will be the case if, for instance, radar's measurements on target n are precise enough, while $\phi_\infty^1 \geq \theta$ iff $\theta \leq 1/2$.
In the sequel we assume, without loss of generality, that $h = 0$.

Case I: Threshold $z \in [0, \phi_\infty^1)$ (*Low thresholds*)

In this case, the active set $B(z) = (z, 1]$ contains the attractive fixed points of the recursions associated to both actions, i.e. $\phi_\infty^{(0)}, \phi_\infty^{(1)}$. This implies that once the STEV state reaches the active set $B(z)$ it stays in $B(z)$ for the rest of the tracking system's life. Such a result follows from Proposition B.2 in Appendix A by which for any $s \geq \phi_\infty^{(1)}$ then

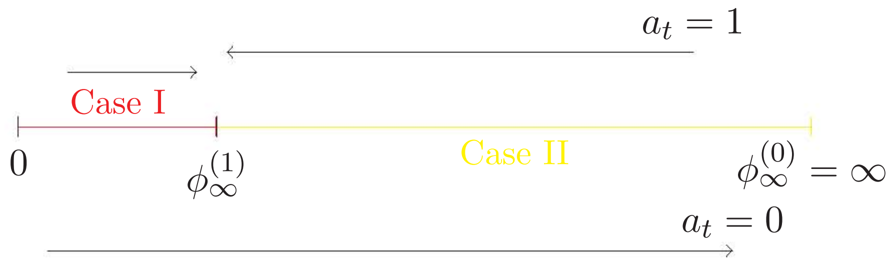


Figure 6.1: The state space and the fixed points of the active and passive dynamics for the single target tracking model

$\phi_t^{(1)}(s) \geq \phi_\infty^{(1)} > z$ for all $t \geq 0$. Further, for $z \in B^c(z) \triangleq [0, z]$: $\phi_t^{(0)}(s) \nearrow \phi_\infty^0$. Hence, after a finite number of passive slots $t_0^*(s, z) < \infty$: $\phi_{t_0^*(s,z)}^{(0)}(x) > z$, where we define the first (deterministic) hitting time to the active set as $t_0^*(s, z) \triangleq \min\{t \geq 1 : \phi_t^{(0)}(s) > z\}$

$$g(s, z) = \begin{cases} \dfrac{1}{(1-\beta)} & x \in (z, 1] \\ \beta^{t_0^*(x,z)} \dfrac{1}{(1-\beta)} & x \in (0, z] \end{cases} \tag{6.15}$$

Notice that the above total work measure admits a closed form expression as it can be seen as a special case of (4.14) (corresponding to the case in which $\theta(s, z, t) = 1$ for all t). Similarly, we obtain the total reward evaluation

$$f(s, z) = \begin{cases} d\, r \displaystyle\sum_{t=0}^{\infty} \beta^t \phi_t^{(1)}(s) & s \in (z, 1] \\ d\, r \displaystyle\sum_{t=0}^{t_0^*(s,z)-1} \beta^t \phi_t^{(0)}(s) + \beta^{t_0^*(s,z)} \sum_{t=0}^{\infty} \beta^t \phi_t^1(y) & s \in (0, z] \end{cases} \tag{6.16}$$

where $y \triangleq \phi_{t_0^*(x,z)}^{(0)}(x)$.

The infinite series in (4.15) are convergent, yet they do not admit closed form formulae. Hence, they must be truncated in practice to evaluate $r(x, z)$ via (6.13), and hence Möbius Transformations properties are required for establishing that SIC condition ii) holds.

The following proposition, based on the above lemma, provides the conditions under which positivity of the marginal work measures is ensured.

Proposition 6.1. *The marginal work measure $w(s,z)$ in problem (6.8) with $s \in \mathbb{S}$ and $z \in [0, \phi_\infty^{(1)})$ is strictly positive for $\beta \in [0,1)$.*

Proof.

$$w(s,z) = \begin{cases} 1 & s \in (z, \infty] \\ 1 - D\beta^M \dfrac{1-\beta^D}{(1-\beta)} & s \in (0, z] \end{cases} \tag{6.17}$$

where $D \triangleq \left(t_0^*(s,z) - t_0^*(\phi^{(1)}(s), z)\right)$ and $M = t_0^*(s,z)$.
For $s > z$, by Lemma C.1 in Appendix C, any initially active state s, it will hold that $\phi^{(0)}(s) > z$ and $\phi^{(1)}(s) > z$. Thus, $w(s,z) = 1$ for $s \in (z, \infty]$. As shown in Appendix C, Lemma C.2 ensures that $D \in \{0,1\}$, thus $w(s,z) > 0$ for $s \in (0,z]$.
Actually, it lowest bound $w(s,z) \geq 1 - \beta^M > 0$ given that $M \geq 1$ for all $s \leq z$ for $\beta < 1$. □

Regarding index computation, first notice that it follows from the definition of $t_0^*(s,z)$ that: $t_0^*(s,s) = 1$ and, given that $\phi^{(1)}(s) > s$ and $\phi^{(0)}(s) > s$, $t_0^*(\phi^{(1)}(s), s) = t_0^*(\phi^{(0)}(s), s) = 0$. This allows us to compute the MP index (6.14) for case I as follows:

$$\lambda^{MP}(s) = \frac{d\,r \left[\displaystyle\sum_{t=0}^{\infty} \beta^t \left[\, \phi_t^{(1)}(\phi^{(0)}(s)) - \phi_t^{(1)}(\phi^{(1)}(s)) \,\right]\right]}{(1-\beta)}, \quad s \in (0, \phi_\infty^{(1)}) \tag{6.18}$$

Notice, that for all $x \in [0, \phi_\infty^{(1)})$ the $\lambda^{MP}(x)$ is an infinite sum of continuous functions of the state. Next, to ensure indexability we must prove that this index is nondecreasing with respect to the STEV state. For such a purpose, we must take the derivative of the two infinite sums in the denominator of the MP ratio defining the index with respect to s. Since, there is no closed form formulae for those sums to manipulate it algebraically, the strategy to accomplish such a goal is to show that it holds $\frac{\partial c(s,s)}{\partial s} > 0$ by manipulating the derivative of each term in the infinite sum.

Lemma 6.1. *For all $s < \phi_\infty^{(1)}$, it holds that:*

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \right] \quad > \quad 0 \tag{6.19}$$

The following proposition, based on the above lemma, provides the conditions under which MP index is monotone. Proof of Lemma 6.1 is outlined in Appendix C, and it follows from Proposition A.5 of the Appendix A.

Proposition 6.2. *The index $\lambda^{MP}(s) = \frac{c(x,x)}{(1-\beta)}$ as defined in (6.18) for problem (6.8) is monotone increasing and continuous in the STEV state s for $s \in (0, \phi_{\infty}^{(1)})$.*

Case II: Threshold $z \in [\phi_{\infty}^{1}, \infty)$ (*Intermediate thresholds*)

In this case, the passive set $B^c(z)$ contains the attractive fixed point of the recursion associated to the active action, i.e. $\phi_{\infty}^{(1)}$, whereas the active set $B(z)$ contains the attractive fixed point of the recursion associated to the passive action, i.e. ϕ_{∞}^{0}. Hence, the state S_t jumps above and below the threshold z, or the rest of the tracking system's life. Following the argument introduced in Niño-Mora and Villar (2009), we consider the iterates $a_t(s, z)$ and $\phi_t(s, z)$, which are the action and STEV processes a_t and s_t generated under the z-threshold policy starting at s. They can be recursively computed as follows. Letting

$$\phi(s, z) \triangleq 1_{s>z}\phi^{(1)}(s) + 1_{s\leq z}(s)\phi^{(0)}(s),$$

where $1_{s>z}$ is the indicator of set $B(z)$, $\phi_0(s, z) \triangleq s$ and $\phi_t(s, z) \triangleq \phi\big(\phi_{t-1}(s, z), z\big)$ for $t \geq 1$. Further, $a_0(s, z) \triangleq 1_{s>z}$, and $a_t(s, z) \triangleq 1_{s>z}\big(\phi_t(s, z), z\big)$ for $t \geq 1$.
Note that the processes: $\phi_t(s, z)$ and $a_t(s, z)$, can be respectively analyzed as forward *orbits* through the initial state s of the underlying discrete deterministic dynamical systems: $(\mathbb{N}_0, \mathsf{S}, \phi)$ and $(\mathbb{N}_0, \{0, 1\}, a)$. Such orbits describe the evolution of the total cost and work measure and, depending on the value of the threshold z, they converge to some (asymptotically) *periodic* orbit. In case I, resulting orbits are closed since the processes converge to some *constant* orbit (or fixed point). Hence, asymptotic or closed-form formulae for the work and cost evaluation measures can be derived by studying the limiting behavior of the corresponding orbits. For an introductory review of discrete nonlinear dynamical systems and chaos theory see e.g. Wiggins (2003).

In the following we list our main results, drawing on the technical analysis of the marginal reward measure presented in Appendix C. Those results are based on properties of the underlying discrete dynamical systems, we shall perform an indexability analysis, as in this case SIC cannot be verified by algebraic means. This section outlines how to do so, and further shows how to use such properties to evaluate the index $\lambda^*(s)$. The total evaluation measures admit the following expressions

$$g(x, z) = \sum_{t=0}^{\infty} \beta^t \, a_t(s, z) \tag{6.20}$$

$$f(x,z) = d\,r \sum_{t=0}^{\infty} \beta^t (\phi a)_t(s,z) \tag{6.21}$$

Expression (6.20) admits closed-form expressions depending on the specific z-threshold value consired. As it is disccussed in Appendix C it is simply an alternated sum of β^t, depending on the resulting (asymptotically) *periodic* orbit of the discrete tracking system. However, as in the previuos model, (6.21) cannot be calculated in a closed form, and thus, truncation is neccesary for evaluating it numerically.

Next, we are able to describe a *recurrent cyclical* pattern in the resulting information state S_t process under a z-threshold policy, which allows us to describe the possible trajectories of the information state to be considered. Specifically, using properties of the Möbius Transformations we are able to establish regularities , in terms of the sequence of active and passive slots until a target is hunted, that allow us to derive the corresponding lower bounds on $w(s,z)$ for this case, which is the most complex of the three threshold cases. Those regularities are summarized in Lemma C.3, Lemma C.4 and Lemma C.5 of the Appendix C. In the following we list the main results for this case.

Using Lemma C.3, Lemma C.4 and Lemma C.5 we derive the lower bound on $w(s,z)$ for all $s \in \mathbb{S}$, as given by Lemma 4.3.

Lemma 6.2. *For all $z \in [\phi_\infty^{(1)}, \infty)$, $s \in \mathbb{S}$:* $\quad w(x,z) \geq (1-\beta)$.

The following proposition, based on the above lemmas, provides the conditions under which positivity of the marginal work measures is ensured.A derivation of this bound is outlined in Appendix C.

Proposition 6.3. *The marginal work measure $w(s,z)$ in problem (6.8) with $s \in \mathbb{S}$ and $z \in [\phi_\infty^{(1)}, \infty)$ is positive for $\beta \in [0,1)$.*

Next we compute the index (3.18) in this case, using the fact that $t_0^*(s,s) = 1$ and, given that $\phi^{(1)}(s) < s < \phi^{(0)}(s)$, as follows:

$$\lambda^{MP}(s) = \frac{d\,r \sum_{t=0}^{\infty} \beta^t \left[(\phi a)_t(\phi^{(0)}(s), s) - (\phi a)_t(\phi^{(1)}(s), s) \right]}{\sum_{t=0}^{\infty} \beta^t \left[a_t(\phi^{(1)}(s), s) - a_t(\phi^{(0)}(s), s) \right]}, \quad s \in [\phi_\infty^{(1)}, \infty) \tag{6.22}$$

Such an index can be expressed as an infinite sum of functions defined by a composition of the two the Möbius transformations describing the active and passive dynamics, depending on the concrete cycle that a given threshold s generates. Showing continuity of the index in this case calls for further research, but the experimental evidence suggests it holds.

Lemma 6.3. *For all $s \geq \phi_{\infty}^{(1)}$, it holds that:*

$$\sum_{t=1}^{\infty} \beta^t \left[\frac{\partial\, a_t(\phi^{(1)}(s), s)}{\partial s} - \beta \frac{\partial a_t(\phi^{(0)}(s), s)}{\partial s} \right] \leq 0 \tag{6.23}$$

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial (\phi a)_t(\phi^{(0)}(s), s)}{\partial s} - \frac{\partial (\phi a)_t(\phi^{(1)}(s), s)}{\partial s} \right] > 0 \tag{6.24}$$

The following proposition, based on the above lemma, provides the conditions under which MP index is monotone. Proof of Lemma 6.3 is outlined in Appendix C, and it follows from Proposition A.5 of the Appendix A.

Proposition 6.4. *The index $\lambda^{MP}(s) = \frac{c(s,s)}{w(s,s)}$ as defined in (6.22) for problem (6.8) is monotone increasing and continuous in the information state s for $s \in [\phi_{\infty}^{(1)}, \infty)$.*

Notice that for the limit case $z = \infty$, under the z-threshold policy S_t will never be above threshold starting from any possible initial level of the STEV s, hence the active set is the null set, i.e. $B(z) = \emptyset$. Thus, in this case every possible initial state is a passive initial state which reduces the computation of work and reward total measures significantly, as

$$S_t = \phi_t(s, z) = \phi_t^{(0)}(s, z) \text{ and } a_t = 0 \quad t \geq 0\, \forall s : s \in \mathbb{S}$$

Hence, in this case a_t converges to a constant orbit whose fixed point is 0 while s_t grows linearly in time up to infinite. Elementary arguments give that for any s in S the total work and cost measure have the following limit:

$$\begin{aligned} g(s, \infty) &= 0 \\ f(s, \infty) &= d\, r \left(\frac{s}{(1-\beta)} + \frac{\theta}{(1-\beta)^2} \right) \end{aligned}$$

Hence, for any s in $\mathbb{S}$ the corresponding marginal measures are:

$$w(s, \infty) = 1 \tag{6.25}$$

$$c(s, \infty) = d\, r \left(\frac{(s+\theta)^2}{(1-\beta)(1+s+\theta)} \right) \tag{6.26}$$

From this, it is readily obtained that the MP measure $\lambda(s,\infty) = c(s,\infty)$ can be expressed as follows:

$$\lambda^{MP}(s,\infty) = d\,r\left(\frac{(s+\theta)^2}{(1-\beta)(1+s+\theta)}\right)$$

Therefore the index (4.12) $\lambda^{MP}(s)$ has the limit:

$$\lim_{s\to\infty} \lambda^{MP}(s) = d\,r\left(\lim_{s\to\infty} \frac{(s+\theta)^2}{(1-\beta)(1+s+\theta)}\right) = \infty \tag{6.27}$$

6.3.1 Verification of PCL-indexability

Based on propositions 6.1-6.4, we conclude:

Theorem 6.1. *The single-target tracking ETD problem (6.8) is PCL-indexable for $\beta \in [0,1)$. Therefore, it is indexable for $\beta \in [0,1)$., and the MP index $\lambda^{MP}(s)$ calculated above is its Whittle index $\lambda^*(s)$.*

Short Proof: Proposition 6.1 and Proposition 6.3 ensure the positivity of the marginal work measure for $\beta \in [0,1)$. Proposition 6.2 and Proposition 6.4 ensure monotonicity and continuity.

We can also extend the result of Theorem 6.1 to the average criterion. Thus, denoting by $\lambda^*_\beta(s)$ the MP index for discount factor β, it holds that $\lambda^*_\beta(s)$ increases monotonically to a finite limiting index $\lambda^*_1(s)$ as $\beta \nearrow 1$.

6.3.2 Index Computation

Whittle's MP index has evaluation given by (6.18) and (6.22). As already mentioned during the indexability analysis, the index $\lambda^*(s)$ for every STEV state in practice must be computed by by truncating the infinite series defining them to a finite number of terms.

Performance Bound Computation

Once the indexability of subproblem (6.8) is ensured by Theorem 6.1 and having proposed a tractable procedure to compute its optimal value given any λ (i.e. the optimal active set $B^*(z)$ contains those information states s such that $\lambda^*(s) - \lambda \geq 0$), we can solve the Lagrangian dual problem (3.7) stated as

$$V_D^d(\boldsymbol{s_0}) = \max_{\lambda\geq 0} \sum_{n=1}^{N} \left[\max_{\pi_n\in\Pi_n} f(s_{n,0}\pi) + \lambda g(s_{n,0},\pi)\right] - \lambda\frac{M}{(1-\beta)} \tag{6.28}$$

Hence, we may use $V_D^d(\boldsymbol{s_0})$ as a lower bound on the best attainable performance for problem (2.5). In the next section we will compute such a bound for the simulated scenarios considered and use it to evaluate the suboptimality gap of our proposed policy and other possible heuristics.

Learn from yesterday,
live for today, hope for tomorrow.
Albert Einstein

Chapter 7

Computational Experiments

In this chapter we illustrate and extend the ideas on the Multitarget tracking Kalman Filter MARBP presented in Chapter 6. We discuss, through a series of computational experiments, index tractability, and relative and absolute performance of the Whittle index policy. We also discuss the convergence rate of the index evaluation method by truncation of the infinite series defining the Whittle index. Throughout the analysis, we will seek to draw insightful interpretations of the results in terms of the tracking problem of concern.

7.1 Index Evaluation

As an example of the use of our index computation method, we have implemented a MATLAB script for index evaluation based on the expressions (6.18) and (6.22). The Whittle index was then computed for a target instance with parameters $d = 1$, $r = 1$, and $q = 5$, so $\theta = 5$, $\phi_{\infty}^{(1)} = 0.8541$. The series in (6.18) and (6.22) were approximately evaluated by truncating them to $T = 10^2$ terms for $\beta = 0.1, 0.2, \ldots, 0.9$, and to $T = 10^5$ terms for $\beta = 0.9999$. For each β, the index $\lambda^{MP}(s)$ was evaluated on a grid of s values of width 10^{-3}. Note that for the case $\beta = 1$ evaluation of the marginal work measure by truncating the series to any number of time slots results in a 0 value, given the infinite asymptotic periodical cycles that govern the evolution of the total work measures under a threshold policy. Yet, the index converges to a limiting value for $\beta = 1$ as a consequence of these asymptotic periodical cycles which can be used to simplify the index computation in this case.

Figure 7.1 shows the results. As required by the PCL-indexability conditions, in each case the index $\lambda^{MP}(s)$ was monotone nondecreasing (in fact, strictly increasing) in s. Note that the index $\lambda^{MP}(s)$ is continuous in s, being also piecewise differentiable. Further, for fixed s the index $\lambda^{MP}(s)$ is increasing in β, converging as $\beta \nearrow 1$ to a limiting index that can be used for the average-criterion problem (6.7), which we have approximated by taking $\beta = 0.9999$. For each s, the time to compute $\lambda^{MP}(s)$ was negligible.

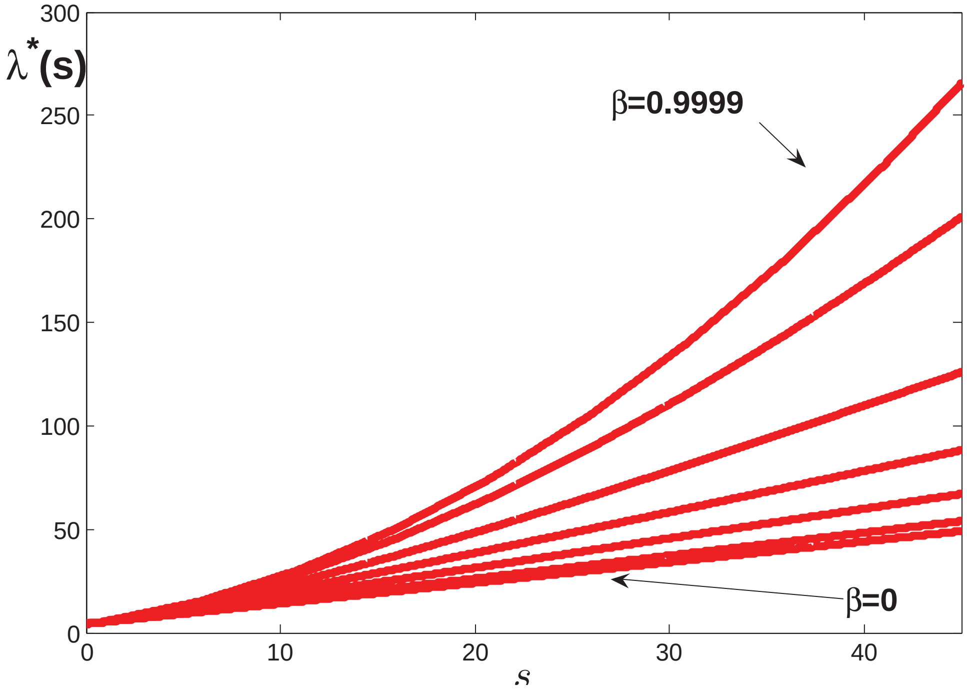


Figure 7.1: Whittle MP index for different discount factors β

From Figure 7.1, we draw the following conclusions regarding the intuition of the optimal tracking policy for one target in isolation.
As $\beta \nearrow 1$, the marginal profit of measuring the target grows non-linearly with its STEV s. That is, as the operating horizon of the tracking system grows, the marginal return of updating a target's position whose STEV increases ($s \nearrow \infty$) becomes significantly large .This reflects the fact that, for a finite measurement error variance $r < \infty$, if the STEV state goes to infinity, then the next TEV p_{t+1} will be $p_{t+1} = \infty$ if we do not measure the target, or equal to the measurement error variance $p_{t+1} = r$ if we do measure it. Naturally, if $r \searrow 0$, then the TEV of target's position can be practically reduced to 0 when measuring a target (even if its previous position was very uncertain). Thus, the marginal return of measuring the target as its STEV grows to infinity, interpreted as the marginal decrease in the total discounted precision cost, also grows to infinity.

7.1.1 Numerical Convergence of the Whittle Index Evaluation

The convergence rate of the above implemented Whittle index approximate evaluation provides meaningful information for the purpose of practical implementation of the resulting target update scheduling policy. Particularly, determining the number of discrete time slots necessary to achieve numerical convergence at some finite computational precision becomes relevant for achieving computational efficiency.

Hence, we have implemented a preliminary computational study in order to assess the convergence behavior of the infinite series defining the proposed Whittle index. Staring from a target instance with parameters as those in Section 7.1, we implemented a script that computes the Whittle' index $\lambda^*(s)$ at the STEV level $s = 1$ truncating the infinite series to time slot T with $T = 1, 2, \ldots, T_{max}$ at each iteration for $\beta = 0.1, 0.2, \ldots, 0.9$ respectively.

For $\beta \leq 0.9$, the numerical convergence of such a series is achieved in practice at some $T_{max} \leq 10^2$. Thus, since Whittle index verifies that $\lambda^*(s) = \lim_{T \to \infty} \lambda^*_T(s)$, we approximate it using the resulting $\lambda^*(s)$ computed truncating the infinite series up to time slot T_{max}, and we thus compute the approximation error $e(T)$ when considering T terms of the series as $\lambda^*_T(s) - \lambda^*_L(s)$. Next, we study the limiting behavior of the following error rate $\frac{e(T+1)}{e(T)}$.

Figure 7.2 shows the results. The Whittle index approximate evaluation appears to converge linearly. Further, the convergence rate appears to be equal to the discount factor β. In fact, we further propose an example, where we analytically derive such a result for a concrete case of the marginal work measure $w(s, s)$, for which a closed form expression is available. Extending the proof for the approximate Whittle index evaluation calls for further investigation.

Notice that, under such conditions, the limiting index for average-criterion problem tends to converge sublinearly, as shown in Figure 7.3. Therefore, precise enough approximations for the case when $\beta \nearrow 1$ will be more expensive computationally as β approaches 1. Further work is required to derive accurate index approximations which require a substantially lower computational effort for a given precision. Such approximations could exploit the structural properties of the STEV asymptotical cycles of the marginal measures under a threshold policy, as reviewed in the indexability analysis of section Section 6.3 and its Appendix in Appendix C, to approximate more efficiently their limiting values.

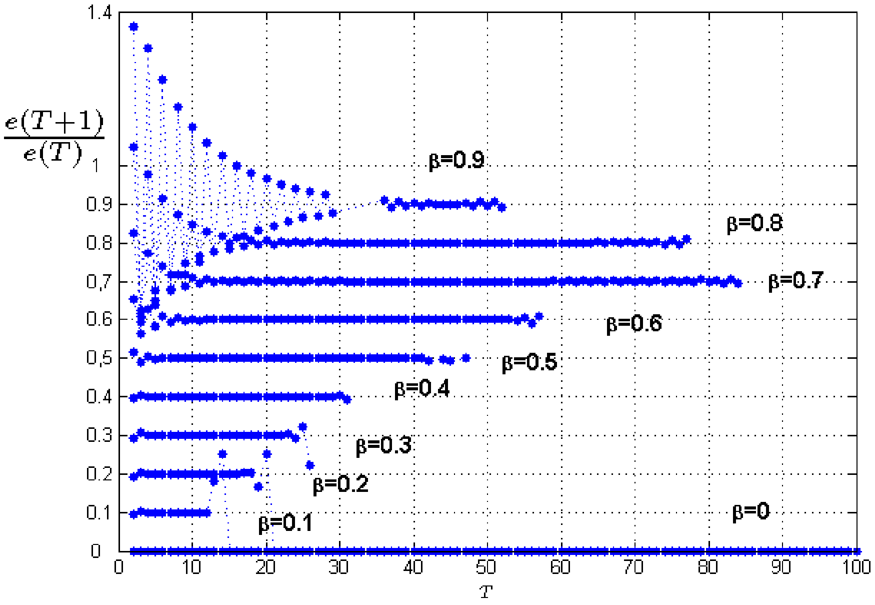


Figure 7.2: The convergence rate of the Whittle index for different discount factors β.

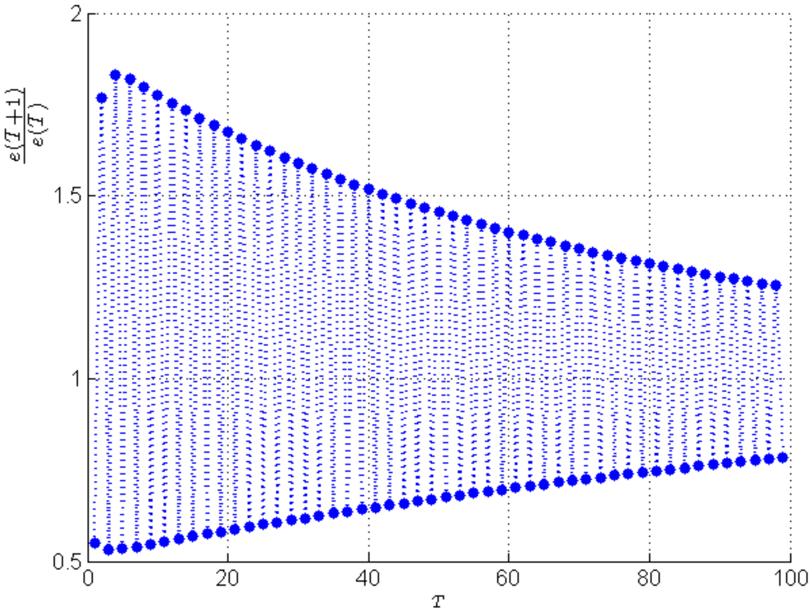


Figure 7.3: The convergence rate of the Whittle index for discount factor $\beta = 0.99$.

Example 7.1. From results in Section 4.3 and in Appendix C, fixing $s = 1$ in a target instance such as the one described in Section 7.1, it can be shown that:

$$w(s,s) = \begin{cases} 1 & s \in [0, \phi_\infty^{(1)}) \\ \\ (1-\beta) & s = \phi_\infty^{(1)} \\ \\ \frac{1-\beta}{1-\beta^2} = \frac{1}{1+\beta}, & s \in (\phi_\infty^{(1)}, \infty) \\ \\ 1, & s = \infty, \end{cases} \tag{7.1}$$

where for STEV states $x \in [\phi_\infty^1, \infty)$ in the absorbing set of states of case II, the asymptotic cycle of the work process a_t under a threshold policy is to alternate a_t into a passive period followed by an active period. Recall that, from definition 6.12 that:

$$w(s,s) = 1 + \beta\big(g(\phi_1^{(1)}(s), s) - g(\phi_1^{(0)}(s), s)\big) \tag{7.2}$$

It follows from result Lemma C.4 in Appendix C the fact that for $s, z \in [\phi^{(1)}(s), \phi^{(0)}(s)]$: $g(s,z) = \frac{1}{1-\beta^2}$ for $s > z$ and otherwise $g(s,z) = \frac{\beta}{1-\beta^2}$. Hence:

$$g(\phi_1^{(1)}(1), 1) = \sum_{t=0}^{\infty} \beta^{2t+1} = \frac{\beta}{(1-\beta^2)} \quad g(\phi_1^{(1)}(1), 1) = \sum_{t=0}^{\infty} \beta^{2t} = \frac{1}{(1-\beta^2)},$$

since $\phi^{(1)}(s) < s < \phi^{(0)}(s)$. Denote the marginal work in STEV s computed by truncating the infinite series up to time slot T as $\hat{w}_T(s,s)$, and notice that:

$$\hat{w}_T(s,s) = 1 + \beta\Big[\sum_{t=0}^{T} \beta^t\big(a_t(\phi_1^{(1)}(s), s) - a_t(\phi_1^{(0)}(s), s)\big)\Big] \tag{7.3}$$

Proposition 7.1. *For $s = 1$, $\theta = 0.5$, $d = r = 1$, the following holds:*

$$\lim_{T\to\infty} \left| \frac{\hat{w}(s,s)_{T+1} - w(s,s)}{\hat{w}(s,s)_T - w(s,s)} \right| = \beta \tag{7.4}$$

Proof. From (7.3) it holds that

$$\hat{w}(s,s)_T = 1 - \beta(1-\beta)\left[\sum_{t=0}^{T} \beta^{2t}\right] \quad \text{for } 0 \le \beta < 1$$

while,

$$\hat{w}(s,s)_T = 1 - \Big[(2T-2) - (2T-1)\Big] = 0 \quad \text{for } \beta = 1$$

Therefore,

$$\lim_{T\to\infty} \left| \frac{\hat{w}(s,s)_{T+1} - w(s,s)}{\hat{w}(s,s)_T - w(s,s)} \right| = \frac{\beta^{2T+2}}{\beta^{2T+1}} = \beta \quad \text{for } 0 \le \beta < 1$$

$$\lim_{T\to\infty} \left| \frac{\hat{w}(s,s)_{T+1} - w(s,s)}{\hat{w}(s,s)_T - w(s,s)} \right| = 1 \quad \text{for } \beta = 1$$

□

• • •

7.2 The Whittle Index and Other Index Policies

In this section we revise the definitions of the possible alternative heuristics for the Multi-armed RB Kalman Filter multitarget tracking problem (6.6) as stated in subsection 6.2.1. In the following section we report on simulation studies that compare the performance of Whittle index policy against these simpler alternatives.

7.2.1 The Myopic Index

The simplest case to consider is the *myopic case*, which corresponds to $\beta = 0$, under which $g(s,z) = a_0(s,z)$, $f(s,z) = d\,r\phi_1(s,z)$, $w(s,z) = 1$, $c(s,z) = d\,r\big[\phi^{(0)}(s) - \phi^{(1)}(s)\big]$, and hence $\lambda(s,z) = c(s,z)$ and $\lambda^*(s) = d\,r\big[\phi^{(0)}(s) - \phi^{(1)}(s)\big] = d\,r(\theta+s)^2/(1+\theta+s)$. Since $(d/ds)\lambda^*(s) = d\,r(\theta+s)(2+\theta+s)/((1+\theta+s)^2 > 0$, the *myopic index* $\lambda^*(s)$ is increasing for all $s \in \mathbb{S}$ and some $\theta > 0$. Therefore, it is straightforward that both conditions in Definition 3.2 hold and thus, by Theorem 3.1, the target's optimal tracking problem for $\beta = 0$ is indexable and $\lambda^*(s) = \lambda^{\text{Myopic}}(s)$ is its Whittle index.

The optimality of such a myopic index policy in the multi-target model for $\beta = 0$, can be also analyzed using the corresponding dynamic programming equations, as it minimizes the total cost function, i.e. the sum of the targets' tracking errors and energy expended for the next time slot. Notice that, for $\beta = 0$, the optimal policy is such that for target n we choose $a_{n,0}$ such that:

$$V_D^*(s_0) \triangleq d_n\, r_n \min_{a_{n,0}\in\{0,1\}} \left\{\phi_1^{a_{n,0}=0}(s_{n,0})\ ;\ \phi_1^{a_{n,0}=1}(s_{n,0})\right\}$$

Where we have assumed, without loss of generality, that $h = 0$. The above problem is equivalent to the following one:

$$V_D^*(s_0) \triangleq d_n \, r_n \min_{a_{n,0} \in \{0,1\}} \left\{0 \, ; \; \phi_1^{a_{n,0}=1}(s_{n,0}) - \phi_1^{a_{n,0}=0}(s_{n,0})\right\}$$

and since $\phi^{(1)}(s) < \phi^{(0)}(s) \forall s \in \mathbb{S}$, we further have that:

$$V_D^*(s_0) \triangleq d_n \, r_n \max_{a_{n,0} \in \{0,1\}} \left(\phi_1^{a_{n,0}=0}(s_{n,0}) - \phi_1^{a_{n,0}=1}(s_{n,0})\right)$$

By the above reasoning it is equivalent to choosing $a_{n,0}$ such that:

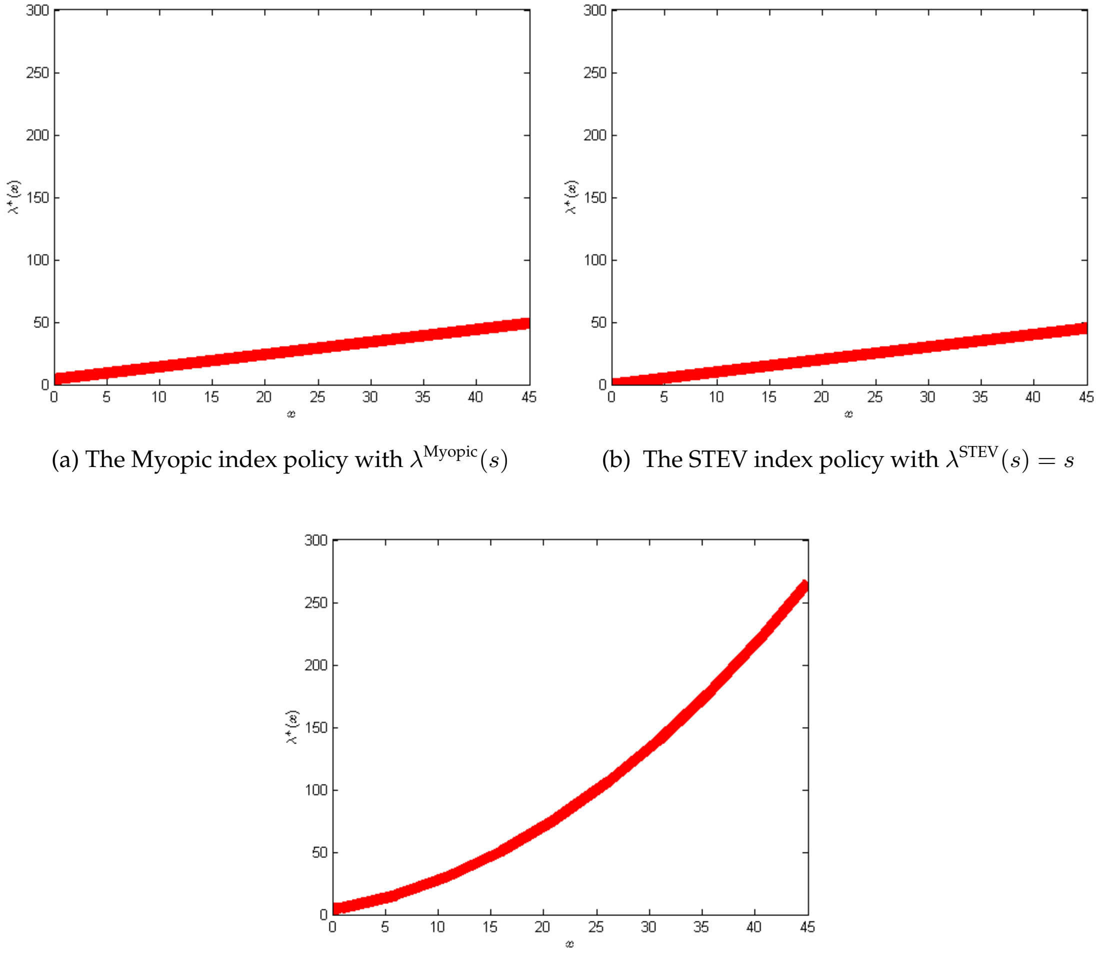
$$d_n \, r_n \min_{a_{n,0} \in \{0,1\}} \left\{\phi_1^{a_{n,0}=0}(s_{n,0}) \, ; \; \phi_1^{a_{n,0}=1}(s_{n,0})\right\} \Longleftrightarrow \max \left\{\lambda^{\text{Myopic}}(s_{n,0})\right\},$$

7.2.2 The STEV index policy

The *STEV index policy*, based on the index $\lambda^{\text{STEV}}(s) = d \, r \, s$. This heuristic is based on the TEV index policy proposed in Howard et al. (2004); La Scala and Moran (2006) and is called the *greedy policy* there. Both in La Scala and Moran (2006) and Howard et al. (2004) the authors claim that this policy optimizes the sum of the targets' track error variances over a finite horizon for $\beta = 1$, by deploying a scheduling *TEV* index policy for the case of two symmetric targets.

Further, in the completely symmetric case in which all N targets have the same state space model, measuring the M targets of highest $\lambda^{\text{Myopic}}(s_n)$, or measuring the M targets with the highest initial STEV $\lambda^{\text{STEV}}(s_n)$, results in an equivalent choice of targets to measure, and therefore in an identical system performance for the next PRI. Such a result holds because, under the identical targets assumption, for all targets and every possible STEV, the *myopic* index is a monotone transformation of the *STEV* index. Thus, in a completely symmetric scenario the Whittle index policy, the *STEV* index policy and the myopic index policy yield an identical tracking performance which is also optimal for $\beta = 0$. Yet, notice that for the general case of asymmetric targets such heuristics are not optimal nor does the above mentioned index policy equivalence hold.

To make a visual comparison of the three index policies, in Figure 7.4, we display the three index plot for $\beta = 1$. Notice that while the Myopic index and the STEV index grow linearly on the STEV, the Whittle index grows more than linearly on it, leading us to conclude that for targets with a large STEV state, priorities assigned by the simpler policies will differ significantly from what the Whittle index policy does. Further, we can thus expect that the performance improvements obtained by the Whittle index policy to be larger when the ratio of sensors to targets is small ($M/N \to 0$), since, in this



(a) The Myopic index policy with $\lambda^{\text{Myopic}}(s)$

(b) The STEV index policy with $\lambda^{\text{STEV}}(s) = s$

(c) The Whittle MP index policy with $\lambda^{\text{MP}}(s) = \lambda^*(s)$ with $\beta = 0.9999$

Figure 7.4: The Myopic, STEV and Whittle index policy

case, all targets will tend to have large TEVs during the whole tracking horizon.

7.2.3 The Whittle Index and the Gittins index: case $\theta = 0$

The motivation to study the case with $\theta = 0$ is that the model becomes classic. When the signal to noise ratio θ equals 0, under which active and passive dynamics reduce to: $\phi_t^{(1)}(s) = \frac{s}{s+1}$ while $\phi_t^{(0)}(s) = s$. Hence, the model is no longer *restless*. Following the previous section argument, it can easily be seen that the work measure:

$$g(s,z) = \begin{cases} \frac{1-\beta^{t_1^*(s,z)}}{1-\beta} = \sum_{t=0}^{t_1^*(s,z)-1} \beta^t, & s > z \\ 0, & s \leq z, \end{cases} \tag{7.5}$$

with $t_1^*(s,z) = \lceil \frac{s-z}{s\,z} \rceil$, whereas the cost measure $f(s,z)$ is characterized by

$$f(s,z) = \begin{cases} d\,r\big[\sum_{t=0}^{t_1^*(s,z)-1} \phi_t^{(1)}(s)\beta^t + \frac{\beta^{t_1^*(s,z)}}{(1-\beta)}\phi^{(1)}_{t_1^*(s,z)}(s)\big], & s > z \\ \\ d\,r\big[\frac{s}{(1-\beta)}\big], & s \leq z. \end{cases} \tag{7.6}$$

Thus, it can be computed that, for $s > z$, it holds that $w(s,z) = 1 - \beta^{t_1^*(s,z)}$, whereas $w(s,z) = 1$ when $s \leq z$ which further implies that $w(s,s) = 1$. In turn, the marginal cost function cane be computed to be $c(s,s) = \frac{d\,r}{(1-\beta)}\big[\phi^0(s) - \phi^1(s)\big]$. Hence $\lambda(s,z) = c(s,z)$ and $\lambda^*(s) = d\,r\frac{s^2}{(1+s)(1-\beta)}$.
Notice that $(d/ds)\lambda^*(s) = \frac{d\,r}{(1-\beta)}s(2+s)/((1+s)^2 > 0$, then the index $\lambda^*(s)$ is non decreasing for $s \in \mathbb{S}$ (and strictly increasing for $s \in \mathbb{S} \setminus 0$). Therefore, both conditions in Definition 3.2 hold and, by Theorem 3.1, the target's optimal tracking problem when $\theta = 0$ is indexable and $\lambda^*(s)$ is its Whittle index. Moreover in this case, $\lambda^*(s)$ is also its Gittins index, since the model formulation under $\theta = 0$ becomes *classic*. Notice also that the Gittins index in this case can be conveniently expressed as: $\frac{\lambda^{\text{myopic}}(s)}{(1-\beta)}$. Thus, the Gittins index representing the maximum rate of discounted reward per unit of discounted time that can be achieved under stopping rules for each initial target state is just the total discounted value over an infinite horizon of the Myopic index at s.

Notice that, since $\theta = q/r$, the case $\theta = 0$ occurs either in the case that the target's movement process is deterministic, i.e., $q = 0$ (in which the only source of error comes from the measurement process), or when its measurement process is absolutely uncertain, i.e., $r = \infty$. For the former case, notice that it is not necessary to assume that the target is frozen at a site to result in a classic model, what we are assuming by letting $q = 0$ is that target movement is completely known. In such a case, the TEV when not measuring the target is simply its previous value p_{t-1}, and continued measurement of the target eventually makes the TEV go to 0, because the measurement error vanishes. However, in the latter case $r = \infty$, measuring a target gives no relevant information. Thus, its TEV is infinite regardless of the selected sensing action. In such a case its index $\lambda^*(s) = \infty$ for all $s \in \mathsf{S}$, denoting the intuition that targets whose measurements processes are extremely unreliable would have absolute priority to access sensing resources, if the total error in prediction is to be minimized.

Further, given that for $\theta = 0$ the model is classic, in then the case of any N objective targets and $M = 1$ sensor, such an index policy is optimal.

7.2.4 The Whittle Index: case $\theta \to \infty$

The motivation to study this case it represents the case in which the evolution of the TEV over its state space is the largest. As a complementary analysis to the one presented in the previous subsection, in the case $\theta \to \infty$, under which active and passive dynamics are reduced to: $\phi_t^{(1)}(s) = \phi_\infty^{(1)} = 1$ while $\phi_t^{(0)}(s) \to \infty$ for all $s, z \in \mathbb{S}$. Hence, starting from any initial STEV in $\mathbb{S}$, the process s_t under a threshold policy infinitely alternates between 1, the minimum STEV level $\phi_\infty^{(1)} = 1$, and ∞, the maximum STEV level, for all $t = 1, 2, ...$ Following the previous section's argument, it can easily be seen that, for all $s, z \in [1, \infty)$,

$$g(s,z) = \begin{cases} \frac{1}{1-\beta^2}, & s > z \\ \frac{\beta}{1-\beta^2}, & s \leq z. \end{cases} \tag{7.7}$$

whereas the cost measure $f(s, z)$ tends to infinity irrespective of the initial state and threshold value. Thus, it can be shown that in this case for any $s, z \in \mathbb{S}$ it holds that $w(s, z) = w(s, s) = \frac{1}{1+\beta}$. Further, $c(s, s) \to \infty$ for any $s, z \in$ S .From where it follows $\lambda^*(s)$ is nondecreasing for $s \in \mathbb{S}$, therefore the target's optimal tracking problem is indexable and $\lambda^*(s)$ is its Whittle index.

Notice that the case $\theta \to \infty$ occurs either when the target's movement process is completely uncertain, i.e., $q = \infty$, or when its measurement process is exact, i.e. $r = 0$. In the former case, it is natural that regardless of how much precision we may have in the measurement at t, the TEV p_t grows to ∞ if the target remains unmeasured. Thus, the marginal return of measuring the target at any state goes to infinity.

The case in which $r = 0$ is more interesting, since in this case it follows from (6.3) that, when the target is measured its TEV goes to 0 immediately after, and further, when not measured its next TEV only increases in the measurement error q. Thus, such a target is measured once every two slots of the time, its TEV alternates between 0 and q, yielding a total discounted variance of $\frac{q}{(1-\beta^2)}$.

7.3 Bechmarking the Whittle Index Tracking Policy

7.3.1 Instances with Asymmetric Targets

We have performed some small-scale preliminary computational studies to assess the relative performance of the Whittle index proposed in Chapter 6 against the alternative reviewed policies: the STEV index policy, and the Myopic index policy.

First, we consider a base instance with a single radar and $N = 4$ symmetric targets with $q_n \equiv 0.5$, $r_n \equiv 1$, $d_n \equiv 1$, and zero measurement costs $h_n \equiv 0$. This base instance with identical targets of low *position to measurement noise variance ratio* θ was modified by varying $q^{(1)}$, the position noise's variance for target 1, while keeping constant r_n, the measurement noise's variance for all n targets. That is, for a given radar measuring precision and while the other target's movement processes remain invariant, the movement process for target 1 becomes more volatile. In particular, each instance, $q^{(1)}$ assumes values over the range $q^{(1)} \in \{0.5, 1, 2, \ldots, 10\}$. The discount factor is $\beta = 0.99$.

The Whittle index was computed on-line for each target, truncating the corresponding infinite series to 10^3 terms using expressions (6.18) and (6.22). For each instance and policy, the system was left to evolve over a horizon of $T = 10^4$ time slots. The initial state for each target n was taken to be $s_0^{(n)} = 0$, which corresponds to exact knowledge of the targets' initial positions.

Table 7.1 reports the resulting TEV performance objective value achieved under each policy for each value of parameter $q^{(1)}$ along with the lower bound obtained from the relaxation. The results show that the Whittle index policy outperforms both the myopic and the TEV index policy. As for the Whittle index policy's suboptimality gap, we can bound it using the relaxation's lower bound. Moreover, we observe that the Whittle index suboptimality gap is between 2 % and 5 %. The Whittle index policy's performance improvement over the myopic policy increases as $q^{(1)}$ gets larger. Note that such a performance gain is 5.42 % for the case in which $q^{(1)} = 3/2$, which is a quite significant amount. For the maximum value of the position noise's variance for target 1 considered, $q^{(1)} = 10$, such a gain is of 61.3 %.

Despite the fact that Whittle index policy also outperforms the TEV index policy, in this case the performance gain is not as significant as with respect to the Myopic index policy. In fact, the TEV policy is almost as good as the Whittle index policy for all cases. We note that, with the system starting from such a base instance, the TEV index policy will tend to give greater priority to target 1 as its movement becomes more uncertain, just as the Whittle index policy does. However, the Whittle index policy and the TEV index policy may prioritize targets differently if the base instance is such that identical targets share a high position variability, and thus a high *position to measurement*

Table 7.1: Benchmarking results (1): $q_n \equiv 0.5$ for all $n \neq 1$

$q^{(1)}$	TEV	Myopic	MP	LB
1/2	5.837	5.829	5.829	5.715
1	6.601	6.750	6.595	6.434
3/2	7.195	7.530	7.143	6.985
2	7.814	7.866	7.618	7.455
5/2	8.091	8.177	8.030	7.845
3	8.361	8.997	8.358	8.144
4	8.889	10.548	8.881	8.675
5	9.409	11.880	9.411	9.187
6	9.923	13.337	9.881	9.699
7	10.435	14.800	10.392	10.205
8	10.944	16.249	10.872	10.710
9	11.452	17.691	11.351	11.192
10	11.959	19.117	11.852	11.670

noise variance ratio, and we vary that instance by allowing a given target to become less volatile in its movement.

To illustrate such a fact, consider a base instance with a single radar and $N = 4$ symmetric targets with $q_n \equiv 10$, $r_n \equiv 1$, $d_n \equiv 1$, and zero measurement costs $h_n \equiv 0$. We next modify this base instance with identical targets of high *position to measurement noise variance ratio* by varying $q^{(1)}$, the position noise's variance for target 1, while keeping constant r_n, the measurement noise's variance for all n targets. That is, for a given radar measuring precision and while the other targets' movement processes remain invariant, the movement process for target 1 becomes less volatile. In particular, at each new instance, $q^{(1)}$ assumes values over the range $q^{(1)} \in \{0.5, 1, 2, \dots, 10\}$. The discount factor is again $\beta = 0.99$.

Table 7.2 reports the resulting TEV performance objective achieved under each policy for each value of parameter $q^{(1)}$. The results show that also in this case the Whittle index policy outperforms both the myopic and the STEV index policies, yet in this case the performance improvement now decreases as $q^{(1)}$ gets larger. For the minimum value of the position noise's variance for target 1 considered, $q^{(1)} = 0.5$, the performance gain of the Whittle index policy over the STEV index policy is 8.58 %, which is a significant amount. Among the STEV and myopic policies, the former performs better for smaller values of $q^{(1)}$, while the latter performs better for larger $q^{(1)}$. In fact, the myopic policy is as good as the Whittle index policy in the symmetric-target case $q^{(1)} = 10$ (and also in the cases $q^{(1)} = 8$ and $q^{(1)} = 9$). As for the Whittle index policy's suboptimality gap,

Table 7.2: Benchmarking results (2): $q_n \equiv 10$ for all $n \neq 1$

$q^{(1)}$	myopic	TEV	MP	LB
1/2	47.584	44.492	40.676	39.839
1	49.193	46.452	43.707	42.856
3/2	50.267	47.902	45.959	45.097
2	51.302	49.475	47.815	46.943
5/2	52.246	56.777	49.409	48.531
3	53.094	51.584	50.855	49.838
4	54.804	54.181	53.367	52.325
5	55.394	56.906	56.140	54.351
6	56.908	56.142	55.396	54.491
7	59.403	59.210	58.924	56.478
8	60.431	60.740	60.431	58.013
9	61.936	62.270	61.936	59.504
10	63.441	63.799	63.441	62.529

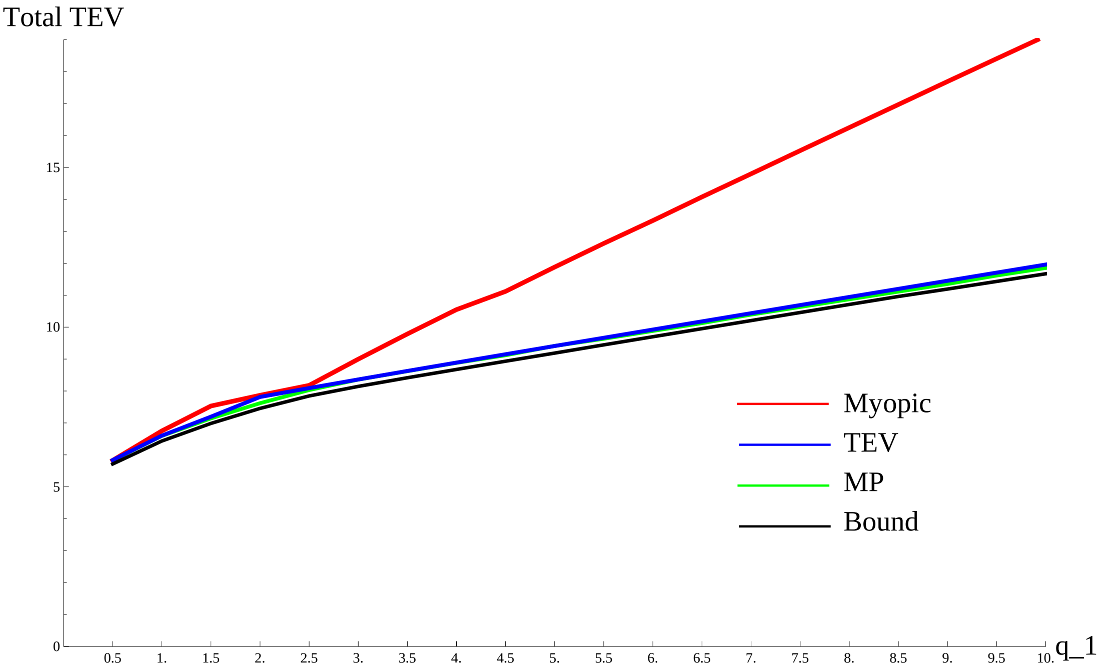


Figure 7.5: The Whittle MP index Benchmarking results (1): $q_n \equiv 1/2$ for all $n \neq 1$.

bounding it above by means of the relaxation's lower bound, we note that the Whittle MP index suboptimality gap is between 2.31 % and 11.68 %.

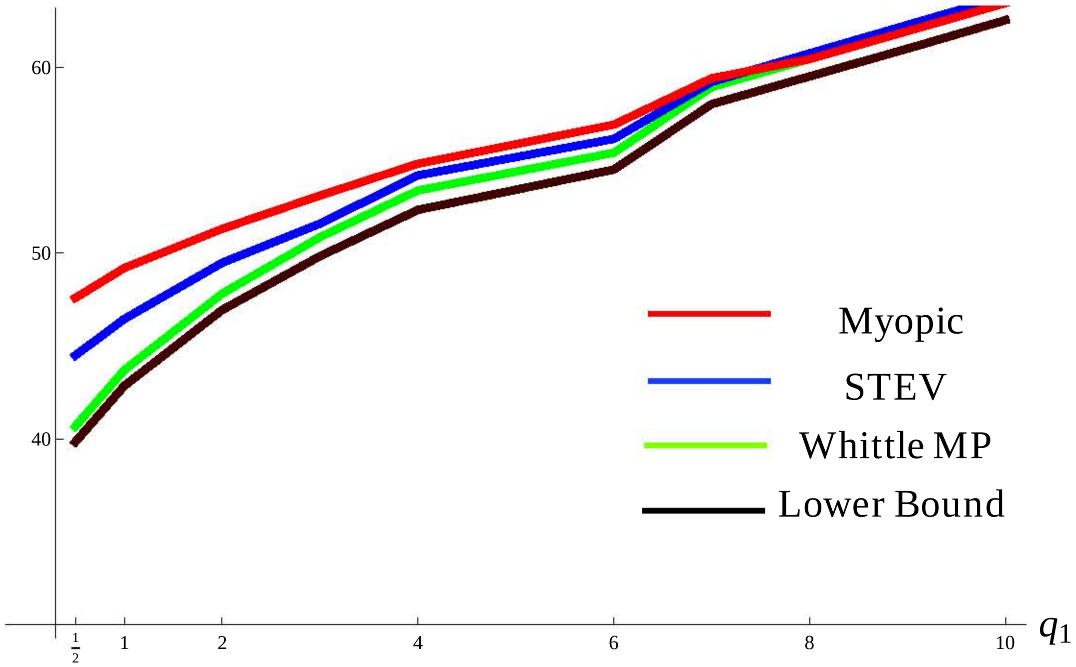


Figure 7.6: The Whittle MP index Benchmarking results (2): $q_n \equiv 10$ for all $n \neq 1$.

7.3.2 Asymptotic Optimality

Together with the RB indexability property introduced in Whittle (1988), Whittle conjectured that for a population with N projects, the policy of being active in the M projects of greatest Whittle index is asymptotically optimal as M and N tend to ∞ in constant ratio R with $R = M/N$.

Such a conjecture can be formulated in terms of the problem under study as follows. Denote as p_j the proportion of targets of type j in the total number of targets, which is characterized by the parameter specification r_j, d_j, q_j, h_j.

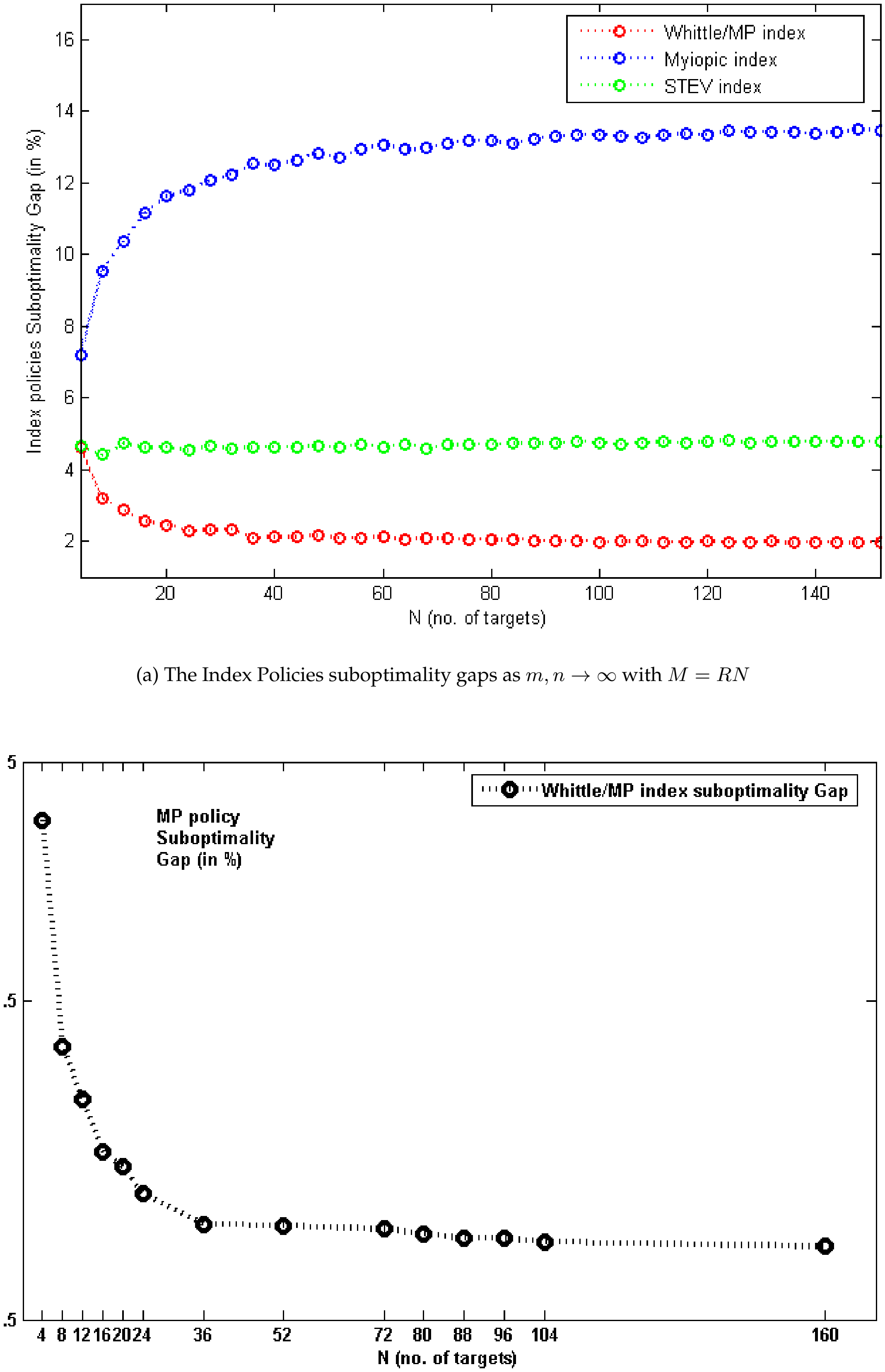
Conjecture 7.1. For population of fixed composition in the sense that $p_j \to p$ as $N \to \infty$, with all N targets being indexable, Whittle conjectured that

$$V_D^*(\mathbf{s};\lambda) \to V^{\mathrm{L}}(\mathbf{s};\lambda) \text{ as } M, N \to \infty \text{ and } R = M/N$$

In Weber and Weiss (1990) the authors provided some counterexamples which elucidated that in general asymptotic optimality of such index policy need not be the case. Further, they established a sufficient condition for such conjecture to hold. Unfortunately, evaluating such a condition for the model at hand is not an easy task, calling for further research.

We have performed a small-scale preliminary computational study to assess the conditions under which we can expect such a conjecture to hold for the present model. We consider a base instance with one beam per 4 objective targets (i.e. $R = 1/4$) for tracking a population of $N = 4$ different targets (i.e. $p = 1/N$), with $q_n \equiv n$, $r_n \equiv 1$, $d_n \equiv 1$, a discount factor of $\beta = 0.99$ and zero measurement costs $h_n \equiv 0$. This base instance was modified by letting the total population of targets N vary over the range $N \in 4 * \{1, 2, \ldots, 40\}$. For each instance the Whittle index policy was computed on-line for each target, truncating the corresponding infinite series to 10^3 terms and the system was left to evolve over a horizon of $T = 10^4$ time slots. The initial state for each target n was taken to be $s_{n,0} = 0$.

Based on the resulting TEV performance objective achieved under the Whittle index policy, and on the lower bound provided by the Lagrangian relaxation approach discussed above, an upper bound for the Whittle index policy the suboptimality gap is computed for each population size N. The results, illustrated in Figure 7.6, show that the upper bound of the Whittle index policy suboptimality gap initially decreases fast as N gets larger, tending to stabilize around 2 % for the largest values of N considered. Such a result seems to suggest that we can expect the proposed Whittle policy to be nearly optimal for cases in which, given a constant radar per target ratio M/N, target heterogeneity grows as the total number of objective targets N grows. Regarding the other policies, we observe that the STEV index policy suboptimality gap is approximately around 4.5 % for all N whereas the myopic index policy's suboptimality gap initially increases fast as N gets larger, tending to stabilize around 13.5 % for the largest values of N considered.



(a) The Index Policies suboptimality gaps as $m, n \to \infty$ with $M = RN$

(b) The Whittle MP index suboptimality gap as $m, n \to \infty$ with $M = RN$

Figure 7.7: The Index Policies suboptimality gap and the asymptotical optimality of the Whittle index hypothesis

Part IV

Conclusions

Aim above morality.
Be not simply good; be good for something.
Henry David Thoreau.

Chapter 8

Summary of Contributions

This thesis has addressed two concrete applications of the MARBP with a real-state variable. The applied problems: (i) hunting multiple elusive hiding targets, and (ii) tracking multiple moving targets, are relevant to the performance optimization of modern sensor systems.

The goal in each of them is to obtain a near-optimal resource allocation policy, which performs well (both in relative and absolute terms). The approach deployed to achieve such a goal, is to design the index policy based on the lagrangian relaxation and decomposition indexation approach introduced by Whittle (1988) and in recent developments by Niño-Mora. This approach allows to design an index rule based on a structural property of the optimal solution to the decomposed parametric-optimization subproblems.

The resulting index policy assigns a value to each possible resource use as a function of its state, and that value *prioritizes* alternative uses in the following sense: those resource allocation options that have the largest index value (as long as they exceed the investment cost) are selected to be engaged, given the resource endowment constraint.

Given a stochastic resource allocation problem formulated as a MARBP, the application of such an indexation approach poses two severe challenges:

(i) The existence of a priority index policy, based on a structural property of the individual arm's optimal policy, is not ensured in the restless case.

(ii) Even if existence of the index is established, computing it in a tractable fashion, may require significant effort.

These two challenges have been addressed by work revised in Niño-Mora (2007a) for the discrete state case. Niño-Mora provided the first tractable Sufficient Indexability Conditions (SIC) for discrete-state MARBP, along with an index algorithm. The resulting approach has proven to be fruitful both in theoretical and algorithmic aspects, as well as in terms of the wide scope of successfully addressed applications.

The application of the indexation approach to real-state variant of the MARBP shares the above mentioned challenges, which have been addressed by the applications researched in this thesis by means of the recent extension of the sufficient indexability conditions introduced by work done in Niño-Mora (2008). In the model analyzed in Niño-Mora (2008), direct verification of the SIC and obtaining a closed-form index formula are possible.

Yet, the real-state MARBPs analyzed in this dissertation posit extra challenges given the technical difficulties introduced by its uncountable state space. Further, the evolution over that uncountable state space is determined by non-linear dynamics, of the type known as Möbiuos transformations or LFT. Specifically, these two facts cause that the state variable and action processes for these models follow infinite possible sequences of values, even if operated under special families of policies (as the threshold policies). All these difficulties result in the lack of a closed form expression for performance measures defining the Whittle index, which complicates significantly the required indexability analysis.

Despite all the above obstacles, we accomplish the main goal of deriving a index-based scheduling policy by exploiting properties of the non-linear dynamics as Möbius transformations, which allow us to reduce and describe the possible structure of state and action trajectories in terms of periodical cycles. This particular achievement is a core contribution of this dissertation shared by the applications studied in Part II and Part III.

Besides the practical contribution of proposing tractable index rules for these two intractable problems, which as shown by Chapter 5 and Chapter 7 exhibit a nearly optimal performance, a main contribution for both applications is to analyze the conditions of the PCL-indexability of the two MARBPs of concern.

Regarding the search problem studied in Chapter 4, of hunting a set of elusive targets by a sensor network of at most as many sensors as targets, the application of the indexation approach yields the following results. Among the main theoretical contributions, we highlight the following: we provide a tractable condition on the the discount factor that ensures the PCL-indexability of the total expected discounted problem, and we propose a tractable index rule for an intractable Partially Observed Markov Decision Process (POMDP). We further provide computational evidence that suggests the optimality of the resulting policy in several scenarios, in which we simultaneously observe that other heuristic perform significantly badly.

In the second problem presented in Chapter 6, of tracking a set of moving targets by a sensor network of at most as many sensors as targets, the application of the indexation approach establishes the PCL-indexability of the ETD subproblem. We further provide computational evidence that suggests that resulting index policy is in general scenarios nearly optimal, outperforming simpler heuristics in the case where targets differ in their motion or measurement parameters. Further, we find results that suggest that its suboptimality gap is bounded and small when the system's size goes to infinity keeping constant the ratio of sensors to targets.

Finally, Chapter 5 and Chapter 7 of thesis contribute to the growing body of computational evidence indicating that Marginal Productivity (MP) Whittle index policies typically achieve a near-optimal performance and in some cases substantially outperform benchmark policies derived from conventional approaches. Traditional myopic heuristics, have been generally found in the literature to be optimal in symmetric scenarios (i.e., where all targets are identical), yet in this dissertation we provide significant evidence which suggests that in asymmetric scenarios they can exhibit a poor performance, both in comparison with the Whittle policy and in absolute terms.

Further, the application of this approach produces insightful interpretations and intuitions regarding the concrete applications. For instance, as pointed out in Chapter 5, when searching an elusive target that reacts to actions by hiding, to hunt it as fast as possible it is of key importance to allow for idling periods so that the target exposes itself. In the tracking model, as pointed out in Chapter 7, as target motion becomes more volatile than its measurment process (i.e., as $\theta \nearrow \infty$), the resulting MP Whittle index significantly differs from the other simpler heuristics, by growing exponentially in the target's STEV rather than linearly.

To conclude, the tractable MP Whittle index sensing policies can be used to solve nearly optimally the underlying (multiple) sequential estimation problems. Both in Chapter 4 and in Chapter 6, the problems can be interpreted as minimum variance estimation problems, in which an unobservable vector of variables (i.e., presence of exposed targets, or target's positions) must be estimated through a noisy measurement vector. Yet, instead of simply using the corresponding filter equations after the realization of the observational data, the observer may control, at each period, the composition of the measurement vector so as to achieve the best overall precision for all the predictive horizon.

References

Anderson, J. (2005). *Hyperbolic geometry*. Springer Verlag.

Bellman, R. (1956). A problem in the sequential design of experiments. *Sankhyā*, 16(3–4):221–229.

Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Bertsekas, D. (2007). *Dynamic programming and optimal control, vol. II*. Athena Scientific.

Bertsimas, D. and Niño-Mora, J. (1996). Conservation laws, extended polymatroids and multiarmed bandit problems; a polyhedral approach to indexable systems. *Math. Oper. Res.*, 21(2):257–306.

Bradt, R. N., Johnson, S. M., and Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.*, 27(4):1060–1074.

Castanon, D. (1997). Approximate dynamic programming for sensor management. In *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, volume 2, pages 1202–1207. IEEE.

Dusonchet, F. and Hongler, M. (2003). Continuous-time restless bandit and dynamic scheduling for make-to-stock production. *Robotics and Automation, IEEE Transactions on*, 19(6):977–990.

Ehsan, N. and Liu, M. (2004). On the optimality of an index policy for bandwidth allocation with delayed state observation and differentiated services. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1974–1983. IEEE.

Gittins, J. C. (1989). *Multi-armed Bandit Allocation Indices*. Wiley, Chichester, UK.

Gittins, J. C. and Jones, D. M. (1974). A dynamic allocation index for the sequential design of experiments. In Gani, J., Sarkadi, K., and Vincze, I., editors, *Progress in Statis-*

tics (*European Meeting of Statisticians, Budapest, 1972*), pages 241–266. North-Holland, Amsterdam, The Netherlands.

Heilmann, W. (1978). Solving stochastic dynamic programming problems by linear programmingan annotated bibliography. *Mathematical Methods of Operations Research*, 22(1):43–53.

Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, New York, NY.

Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York, NY.

Hero, A., Kastella, K., Castanon, D., and Cochran, D. (2006). *Foundations and applications of sensor management*. Springer.

Hong, S. and Jung, Y. (1998). Optimal scheduling of track updates in phased array radars. *IEEE Transactions on Aerospace and Electronic Systems*, 34(3):1016–1022.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA.

Howard, S., Suvorova, S., and Moran, B. (2004). Optimal policy for scheduling of Gauss–Markov systems. *7th International Conference on Information Fusion (Stockholm, Sweden)*.

Kaelbling, L., Littman, M., and Cassandra, A. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134.

Koopman, B. (1946). Search and screening (oeg report no. 56, the summary reports group of the columbia university division of war research). alexandria, virginia: available from the center for naval analyses. Technical report.

Kreucher, C., Blatt, D., Hero, A., and Kastella, K. (2006). Adaptive multi-modality sensor scheduling for detection and tracking of smart targets. *Digital Signal Processing*, 16(5):546–567.

Krishnamurthy, V. and Evans, R. J. (2001). Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking. *IEEE Trans. Signal Process.*, 49(12):2893–2908.

La Scala, B. and Moran, B. (2006). Optimal target tracking with restless bandits. *Digital Signal Process.*, vol. 16(5):pp.479–487.

Liu, B., JI, C., Zhang, Y., and Hao, C. (2009). Blending sensor scheduling strategy with particle filter to track a smart target. *Wireless Sensor Network*, 1:300–305.

Liu, K. and Zhao, Q. (2008). A restless bandit formulation of opportunistic access: Indexablity and index policy. In *Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2008. SECON Workshops' 08. 5th IEEE Annual Communications Society Conference on*, pages 1–5. IEEE.

Moran, W., Suvorova, S., and Howard, S. (2008). Application of sensor scheduling concepts to radar. *Foundations and Applications of Sensor Management*, pages 221–256.

Ng, G. and Ng, K. (2000). Sensor management–what, why and how. *Information Fusion*, 1(2):67–75.

Niño-Mora, J. (2001). Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability*, 33(1):76–98.

Niño-Mora, J. (2002). Dynamic allocation indices for restless projects and queueing admission control: a polyhedral approach. *Mathematical programming*, 93(3):361–413.

Niño-Mora, J. (2006a). Marginal productivity index policies for scheduling a multiclass delay-/loss-sensitive queue. *Queueing Systems*, 54(4):281–312.

Niño-Mora, J. (2006b). Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock m/g/1 queues. *Mathematics of Operations Research*, pages 50–84.

Niño-Mora, J. (2007a). Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):161–198.

Niño-Mora, J. (2007b). Marginal productivity index policies for admission control and routing to parallel multi-server loss queues with reneging. volume 4465 of *Lecture Notes in Computer Science*, pages 138–149. Springer.

Niño-Mora, J. (2008). An index policy for dynamic fading-channel allocation to heterogeneous mobile users with partial observations. In *Next Generation Internet Networks, 2008. NGI 2008*, pages 231–238. IEEE.

Niño-Mora, J. (2009). A restless bandit marginal productivity index for opportunistic spectrum access with sensing errors. *Network Control and Optimization. Lecture Notes in Computer Science. Springer*, Volume 5894,:60–74.

Niño-Mora, J. (2011a). Conservation laws and related applications. *In Wiley Encyclopedia of Operations Research and Management Science.*

Niño-Mora, J. (2011b). Multi-armed restless bandits, index policies, and dynamic priority allocation. *Boletín de la Sociedad de Estadística e Investigación Operativa*, 26(2):124–133.

Niño-Mora, J. and Villar, S. (2009). Multitarget tracking via restless bandit marginal productivity indices and Kalman filter in discrete time. In *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009*, pages 2905–2910. IEEE.

Niño-Mora, J. and Villar, S., S. (2011). Sensor scheduling for hunting elusive hiding targets via whittle's restless bandit index policy. In *NetGCOOP 2011 : International conference on NETwork Games, COntrol and OPtimization*. IEEE.

Papadimitriou and Tsitsiklis (1994). The complexity of optimal queueing network control. *Structure in Complexity Theory Conference, 1994., Proceedings of the Ninth Annual*, pages 318–322, 1–8.

Puterman, M. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc.

Raissi-Dehkordi, M. and Baras, J. (2002). Broadcast scheduling in information delivery systems. In *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, volume 3, pages 2935–2939. IEEE.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.

Rucker, J. (2006). Using agent-based modeling to search for elusive hiding targets. Technical report, DTIC Document.

Savage, C. and La Scala, B. (2009). Sensor management for tracking smart targets. *Digital Signal Processing*, 19(6):968–977.

Smith, W. (1956). Various optimizers for single-stage production. *Naval Research Logistics Quarterly*, 3(1-2):59–66.

Song, N. and Teneketzis, D. (2004). Discrete search with multiple sensors. *Mathematical Methods of Operations Research*, 60(1):1–13.

Stone, L. (1975). *Theory of optimal search*, volume 118. Elsevier Science.

Stromberg, D. (1996). Scheduling of track updates in phased array radars. In *Radar Conference, 1996., Proceedings of the 1996 IEEE National*, pages 214–219.

Thompson, W. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

van Keuk, G. and Blackman, S. (1993). On phased-array radar tracking and parameter control. *IEEE Transactions on aerospace and electronic systems*, 29(1):186–194.

Varaiya, P., Walrand, J., and Buyukkoc, C. (1985). Extensions of the multiarmed bandit problem: the discounted case. *Automatic Control, IEEE Transactions on*, 30(5):426–439.

Veatch, M. and Wein, L. (1996). Scheduling a make-to-stock queue: Index policies and hedging points. *Operations Research*, pages 634–647.

Washburn, R. (2008). Application of multi-armed bandits to sensor management. *Foundations and Applications of Sensor Management*, pages 153–175.

Washburn, R., Schneider, M., and Fox, J. (2002). Stochastic dynamic programming based approaches to sensor resource management. In *Information Fusion, 2002. Proceedings of the Fifth International Conference on*, volume 1, pages 608–615. IEEE.

Weber, R. (1992). On the gittins index for multiarmed bandits. *The Annals of Applied Probability*, pages 1024–1033.

Weber, R. and Weiss, G. (1990). On an index policy for restless bandits. *Journal of Applied Probability*, 27(3):637–648.

Whittle, P. (1980). Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 143–149.

Whittle, P. (1988). Restless bandits: activity allocation in a changing world. *Journal of Applied Probability*, 25:287–298.

Wiggins, S. (2003). *Introduction to applied nonlinear dynamical systems and chaos*, volume 2. Springer Verlag.

Williams, J. (2007). *Information theoretic sensor management*. PhD thesis, Massachusetts Institute of Technology.

Xiong, N. and Svensson, P. (2002). Multi-sensor management for information fusion: issues and approaches. *Information fusion*, 3(2):163–186.

Appendix A

Appendix: A Review of Möbius Transformations

Here we review some auxiliary material useful for the indexability analysis done in of Chapter 4 and Chapter 6.

In both chapters we have considered two distinct iterated mappings of the form $x \mapsto \phi^a(x)$ where x denotes the belief state and $a = 0, 1$ stands for passive and active actions respectively. Letting $\phi_0^a(x) \triangleq x$ and $\phi_t^a(x) \triangleq \phi^a(\phi_{t-1}^a(x))$ for $t \geq 1$, where the functions $\phi^a(x)$ are respectively defined for $a = 0, 1$ by (4.7).

For the sake of establishing PCL indexability, we must study the behavior of the t-th iterate of both mappings in order to derive some required properties of $\phi_t^a(x)$. To prove all these properties, it is convenient to visualize these mappings, especially $\phi_t^1(x)$ which is the most complex of the two recursions, as Möbius Transformations.

Definition A.1. As in Anderson (2005).
A Möbius transformation is a function m: $\mathbb{C} \rightarrow \mathbb{C}$ of the form

$$m(x) = \frac{ax + b}{cx + d}$$

where a,b,c,$d \in \mathbb{C}$ and $ad - cb \neq 0$

Proposition A.1. *Given two Möbius Transformations $m(x)$ and $n(x)$, defined by: $m(x) = \frac{ax+b}{cx+d}$ and $n(x) = \frac{\alpha x+\beta}{\gamma x+\delta}$, the composition of $m(x)$ and $n(x)$ is a Möbius Transformation with:*

$$n\, o\, m(x) = \frac{(\alpha\, a + \beta\, c)x + (\alpha\, b + \beta\, d)}{(\gamma\, a + \delta\, c)x + (\gamma\, b + \delta\, d)}$$

Corollary A.1. *If we define a 2×2 matrix with its entries as the pairs of coefficients of the a* *given Möbius transformation $m(x)$ and $n(x)$, we get:*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad N = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Then, the entries of the matrix representation associated to the composition $n\,o\,m(x)$ correspond to the entries of the product matrix $N \times M$.

It follows from Proposition A.1 and Corollary A.1 that the closed-form representation of the t^{th} composition of any Möbius transformation $m(x)$:

$$\underbrace{m\,o\,m \ldots\, o\,m}_{t}\,(x),$$

denoted as $m_t(x)$, has an associated matrix representation of M^t.

Hence, the associated matrix M^t is obtained as the t^{th} power of the matrix M. We denote λ_1, λ_2 to the eigenvalues of the M matrix, then it holds that:

$$M^t = C \quad \begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix} \quad C^{-1} \tag{A.1}$$

Where C is a matrix in which the i^{th} column is an eigenvector corresponding to the i^{th} eigenvalue of M, and where the eigenvalues of the matrix M and can be computed to be equal to:

$$\lambda_{1,2} \quad = \quad \frac{1}{2}\left(a + d \mp \sqrt{(a-d)^2 + 4\,b\,c}\right) \tag{A.2}$$

Notice that for the case $c \neq 0$ (i.e. in the case we have a non linear function of x), the matrix C is computed to be:

$$C = \begin{pmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{pmatrix},$$

with

$$\gamma_{1,2} \quad = \quad \frac{(a-d) \mp \sqrt{(a-d)^2 + 4\,b\,c}}{2\,c}$$

notice that in this case it holds that $\gamma_i = (\lambda_i - d)/c$.

Thus, from (A.1) it follows that the t^{th} iterate of a Möbius transformation has the following associated matrix representation:

$$M^t(\gamma_1,\gamma_2,k) = \begin{pmatrix} \gamma 1 - k^t\gamma_2 & (k^t-1)\gamma_1\gamma_2 \\ 1-k^t & k^t\gamma_1 - \gamma_2 \end{pmatrix}$$

where $k \triangleq \frac{\lambda_2}{\lambda_1}$.

Finally, the above expression for M^t allows us to conclude the following results:

Proposition A.2.

$$m_t(x) \quad = \quad \frac{\gamma_1(x-\gamma_2) + k^t\,\gamma_2\,(\gamma_1 - x)}{x - \gamma_2 + k^t(\gamma_1 - x)},$$

Proposition A.3.

$$\lim_{t\to\infty} m_t(x) = \begin{cases} \frac{\gamma_1(x-\gamma_2)}{x-\gamma_2} = \gamma_1 & \textit{if } |k| < 1 \\ \\ \frac{\gamma_2(\gamma_1 - x)}{\gamma_1 - x} = \gamma_2 & \textit{if } |k| > 1 \end{cases}$$

Proposition A.4.

$$\frac{\partial\, m_t(x)}{\partial t} = \frac{k^t(x-\gamma_1)(x-\gamma_2)(\gamma_1-\gamma_2)\log(k)}{[(\gamma_2 - x) + k^t\,(x-\gamma_1)]^2}$$

Proposition A.5.

$$\frac{\partial\, m_t(x)}{\partial x} = \frac{k^t(\gamma_1-\gamma_2)^2}{[(\gamma_2 - x) + k^t(x-\gamma_1)]^2}$$

Whereas for the case $c = 0$ and $a \neq d$ (i.e. in the case we have a linear function of x), the matrix C in (A.1) is computed to be:

$$C = \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix},$$

with $\gamma = -\frac{b}{a-d}$. Thus, from (A.1) it follows that the t^{th} iterate of a such a Möbius transformation has the following associated matrix representation:

$$M^t(\gamma,k) = \begin{pmatrix} 1 & -\gamma\,(1-k^t) \\ 0 & k^t \end{pmatrix}$$

Finally, the above expression for M^t allows us to conclude the following results:

Proposition A.6.

$$m_t(x) = \gamma - k^{-t}\,(\gamma - x)$$

Proposition A.7.

$$\lim_{t\to\infty} m_t(x) = \begin{cases} \infty & \textit{if } |k| < 1 \\ \\ \gamma & \textit{if } |k| > 1 \end{cases}$$

Proposition A.8.

$$\frac{\partial\, m_t(x)}{\partial t} = -k^t\,(\gamma - x)\,\log(k)$$

Proposition A.9.

$$\frac{\partial\, m_t(x)}{\partial x} = k^{-t}$$

Straightforward application of the previously revised results allows us to conclude that: (i) the $\phi^a(x)$ functions in (4.7) define two Möbius transformations (where $0/1$ stands for the passive/active dynamics respectively) with associated matrix representations given by:

$$\Phi^0 = \begin{pmatrix} \rho^0 & p^0 \\ 0 & 1 \end{pmatrix} \qquad \Phi^1 = \begin{pmatrix} \rho^1\,\alpha - (1-\alpha)\,p1 & p^1 \\ -(1-\alpha) & 1 \end{pmatrix}$$

(ii) Thus, results A.2-A.5 and A.6-A.9 apply to the functions in (4.7). We then compute the eigenvalues of both matrices to respectively be:

$$\lambda^0_1 = \rho^0, \lambda^0_2 = 1 \qquad \lambda^1_{1,2} = \frac{A \mp \sqrt{A^2 - 4\,(1-\alpha)\,p^1}}{2} + 1 \tag{A.3}$$

with $A \triangleq (1-\rho^1) + (1-\alpha)(\rho^1 + p^1)$. Further, we have that:

$$k^0 = 1/\rho^0 \qquad k^1 = \frac{A + \sqrt{A^2 - 4\,(1-\alpha)\,p^1} + 2}{A - \sqrt{A^2 - 4\,(1-\alpha)\,p^1} + 2} \tag{A.4}$$

$$\gamma = \frac{p^o}{1-\rho^0} \qquad \gamma^1_{1,2} = \frac{A \pm \sqrt{A^2 - 4\,(1-\alpha)\,p^1}}{2\,(1-\alpha)} \tag{A.5}$$

We have defined $\phi^a_\infty \triangleq \lim_{t\to\infty} \phi^a_t(x)$. Notice that, from results A.3 and A.7 it follows that

$$\phi^0_\infty = \gamma = \frac{p^o}{1-\rho^0} \quad , \quad \phi^1_\infty = \gamma_2 = \frac{A - \sqrt{A^2 - 4\,(1-\alpha)\,p^1}}{2\,(1-\alpha)}.$$

By application of A.2 and A.6 we solve for the passive and active recursions in closed-form. Thus, the expression for the t^{th} passive/active iteration on some belief state x are respectively computed as follows:

$$\phi^0_t(x) = \phi^0_\infty - \left(\rho^0\right)^t \left(\phi^0_\infty - x\right) \tag{A.6}$$

$$\phi^1_t(x) = \frac{\gamma_1(x - \phi^1_\infty) + (k^1)^t \phi^1_\infty(\gamma_1 - x)}{x - \phi^1_\infty + (k^1)^t(\gamma_1 - x)}, \tag{A.7}$$

Notice that from (A.6) and (A.7) it follows that for $a = 0, 1$:

$$x < \phi^a_\infty \quad \longrightarrow \quad \phi^a_t(x) < \phi^a_\infty \quad \forall t \geq 0 \tag{A.8}$$

To show (A.8) we apply results A.2 and A.3 and the results A.6 and A.7 to conclude that:

$$\text{For } x < \phi^1_\infty, \quad \phi^1_t(x) < \phi^1_\infty \quad \text{ since } \gamma_1 > \phi^1_\infty \tag{A.9}$$

$$\text{If } x < \phi^0_\infty, \quad \phi^0_t(x) < \phi^0_\infty \tag{A.10}$$

Next, by results A.4 and A.8 it follows that:

$$\frac{\partial\, \phi^0_t(x)}{\partial t} = -\left(\rho^0\right)^t \left(\phi^0_\infty - x\right)\, \log(\rho^0) \tag{A.11}$$

$$\frac{\partial\, \phi^1_t(x)}{\partial t} = = \frac{(k^1)^t(x-\gamma_1)(x-\phi^1_\infty)(\gamma_1-\phi^1_\infty)\log(k^1)}{\left[(\phi^1_\infty - x) + k^t\,(x-\gamma_1)\right]^2} \tag{A.12}$$

Hence,

$$\text{sgn}\frac{\partial\, \phi^0_t(x)}{\partial t} = \text{sgn}(\phi^0_\infty - x) \tag{A.13}$$

$$\text{sgn}\frac{\partial\, \phi^1_t(x)}{\partial t} = \text{sgn}(\phi^1_\infty - x) \tag{A.14}$$

Finally, by results A.5 and A.9 we have that:

$$\frac{\partial\, \phi^0_t(x)}{\partial x} = \rho^t \tag{A.15}$$

$$\frac{\partial\, \phi^1_t(x)}{\partial x} = \frac{(k^1)^t(\gamma_1-\phi^1_\infty)^2}{\left[(\phi^1_\infty - x) + (k^1)^t(x-\gamma_1)\right]^2} \tag{A.16}$$

Therefore,

$$\text{sgn}\frac{\partial\, \phi^0_t(x)}{\partial x} = \text{sgn}\left(\rho^t\right) > 0 \tag{A.17}$$

$$\text{sgn}\frac{\partial\, \phi^1_t(x)}{\partial x} = \text{sgn}\left(k^1\right)^t > 0 \tag{A.18}$$

Finally, also notice that:

$$\frac{\partial^2\, \phi_t^0(x)}{\partial x^2} = 0 \tag{A.19}$$

$$\frac{\partial^2\, \phi_t^1(x)}{\partial x^2} = -2\, \frac{(k^1)^t(\gamma_1 - \phi_\infty^1)^2}{\left[(\phi_\infty^1 - x) + (k^1)^t(x - \gamma_1)\right]^3} \left(1 - (k^1)^t\right) \tag{A.20}$$

Appendix B

Appendix to Chapter 4

B.1 Work-Reward Measures Analysis

In order to prove Proposition Proposition 4.1 in Chapter 4, we invoked Lemmas providing lower bounds on the marginal work measures $w(x, z)$. In this Appendix we shall derive those bounds in detail.

In the elusive target hunt model model presented in Chapter 4 we have considered two iterated mappings of the form $x \mapsto \phi^{(a)}(x)$ where x denotes the initial information state and $a = 0, 1$ stands for passive and active actions respectively. Letting $\phi_0^{(a)}(x) \triangleq x$ and $\phi_t^{(a)}(x) \triangleq \phi^{(a)}(\phi_{t-1}^{(a)}(x))$ for $t \geq 1$, and defining:

$$\phi^{(0)}(x) = p^{(0)} + \rho^{(0)}x \tag{B.1}$$

$$\phi^{(1)}(x) = p^{(1)} + \rho^{(1)}\frac{\alpha x}{1-(1-\alpha)x} \tag{B.2}$$

For the sake of establishing PCL-indexability, we are interested in studying the behavior of the t-th iterate of both mappings. In order to do we visualize both dynamics as Möbius Transformations or Linear Fractional Transformations (LFTs), with associated matrix representations given by:

$$\Phi^0 = \begin{pmatrix} \rho^{(0)} & p^{(0)} \\ 0 & 1 \end{pmatrix} \qquad \Phi^1 = \begin{pmatrix} p^{(1)} - \alpha(1-q^{(1)}) & p^{(1)} \\ -(1-\alpha) & 1 \end{pmatrix}$$

Note that for equation (B.1), the corresponding LFT is a combination of a translation and a rotation (since in this case $c = 0$ and $a = d$) and thus, one of its fixed points is at infinity. The attractive fixed points for these recursions are:

$$\phi_\infty^{(0)} \triangleq \frac{p^{(0)}}{1-\rho^{(0)}} \qquad \phi_\infty^{(1)} \triangleq \frac{\gamma - \sqrt{\gamma^2 - 4p^{(1)}(1-\alpha)}}{2(1-\alpha)}$$

with $\gamma \triangleq 1 - \rho^{(1)} + (p^{(1)} + \rho^{(1)})(1-\alpha)$.

For proving Lemma 4.1 we deploy results and properties of the Möbius transformations, which are listed as lemmas after the the proof of Lemma 4.1.

Proof of Lemma 4.1
For all $z < \phi_{\infty}^{(1)}$, consider first the case with $0 < x \leq z$, which is the first part of Lemma 4.1.

$$(i) \quad w(x,z) > \min\{1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta) + \beta\,(1-\alpha)\,z}, 1-\beta\} \geq 0 \ \text{ for any } x \in (0,z],\ 0 \leq \beta \leq 1.$$

To compute a lower bound on $w(x,z)$ we must first compute $g(\phi^1(x),z)$ and $g(\phi^0(x),z)$ according to (4.14), for which we need to establish whether $\phi^{(1)}(x)$ and $\phi^{(0)}(x)$ are initially active or passive states. From Lemma B.1 it follows that both $\phi^{(1)}(x) > x$ and $\phi^{(0)}(x) > x$, yet the total work measure is computed differently depending on the relation of $\phi^{(1)}(x)$, $\phi^{(0)}(x)$ and z. Hence,
1) If $z < \phi^{(1)}(x)$:
When $\phi^{(1)}(x) > z$ it also holds that $\phi^{(0)}(x) > z$. Thus, it holds that $z < \phi^{(1)}(x) < \phi^{(0)}(x)$.

$$\begin{aligned}
w(x,z) &= 1 + \beta\ [1-(1-\alpha)\,x\,]\ g(\phi^{(1)}(x),z) - \beta\, g(\phi^{(0)}(x),z) \\
&> 1 + \beta\ [1-(1-\alpha)\,x\,]\ g(\phi^{(1)}(x),z) - \beta\, g(\phi^{(1)}(x),z) \\
&\qquad \text{(By Lemma B.1 and Lemma B.3)} \\
&> 1 - \beta\,(1-\alpha)\,x\,g(\phi^{(1)}(x),z) \quad (\phi^{(1)}(x) > z) \\
&> 1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z} \geq 0 \quad \text{(By Lemma B.4 and } (x \leq z)\text{)} \qquad \blacksquare \text{(B.3)}
\end{aligned}$$

2) If $\phi^{(1)}(x) \leq z$, we have two possible cases:
a) $\phi^{(0)}(x) \leq z$ and $t_0^*(\phi^{(1)}(x),z) = t_0^*(\phi^{(0)}(x),z)$,
We write by $y_1^1 \triangleq \phi^{(0)}_{t_0^*(\phi^{(1)}(x),z)}\left(\phi^{(1)}(x),z)\right)$ and $y_0^1 \triangleq \phi(0)_{t_0^*(\phi^{(0)}(x),z)}\left(\phi^{(0)}(x),z)\right)$. Notice that y_1^1, y_0^1 stand for the first belief state to reach the active set by having been active/passive in its initial state and then following a z-threshold policy. Further, by Lemma B.5 it may be the case that $t_0^*(\phi^{(1)}(x),z) = t_0^*(\phi^{(0)}(x),z)$.
Thus, if $\phi^{(1)}(x) \leq z$, $\phi^{(0)}(x) \leq z$, and $t_0^*(\phi^{(1)}(x),z) - t_0^*(\phi^{(0)}(x),z) = 0$ we have that $y_1^1 < y_0^1$, by Lemma B.1 and Lemma B.7. Thus, by the same reasoning deployed in i) we

can invoke Lemma B.3 to conclude that the lower bound on $w(x,z)$ for this case is:

$$\begin{aligned}
w(x,z) &= 1+\beta^{t_0^*(\phi^{(1)}(x),z)+1}\,[1-(1-\alpha)\,x\,]\;g(y_1^1,z)-\beta^{t_0^*(\phi^{(0)}(x),z)+1}\,g(y_0^1,z)\\
&> 1+\beta^{t_0^*(\phi^{(1)}(x),z)+1}\,[1-(1-\alpha)\,x\,]\;g(y_1^1,z)-\beta^{t_0^*(\phi^{(1)}(x),z)+1}\,g(y_1^1,z)\\
&\qquad \text{(By Lemma B.1, Lemma B.7 and Lemma B.3)}\\
&> 1-\beta^{t_0^*(\phi^{(1)}(x),z)+1}\,(1-\alpha)\,x\,g(y_1^1,z)\quad (y_1^1>z)\\
&> 1-\beta^{t_0^*(\phi^{(1)}(x),z)}\,\frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z}\geq 0 \qquad \text{(By Lemma B.4 and } (x\leq z))\\
&> 1-\ \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z},\\
&> 1-\beta^{t_0^*(\phi^{(1)}(x),z)}\geq 1-\beta\quad (t_0^*(\phi^{(1)}(x),z)\geq 1)\\
&> \min\{1-\frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z},1-\beta,\}\geq 0 \qquad \blacksquare
\end{aligned} \tag{B.4}$$

b) $\phi^{(0)}(x)\leq z$ and $t_0^*(\phi^{(1)}(x),z)-t_0^*(\phi^{(0)}(x),z)=1$,

For any $x>z$, we will further define $t_0^*(x,z)=0$, so the case $t_0^*(\phi^{(1)}(x),z)-t_0^*(\phi^{(0)}(x),z)=1$ includes both that $\phi^{(1)}(x)\leq z$, $\phi^{(0)}(x)>z$ or $\phi^{(1)}(x)\leq z$, $\phi^{(0)}(x)\leq z$. Hence, we write $g(x,z)$ for both of these cases as follows:

$$g(\phi^{(1)}(x),z)=\beta^{t_0^*(\phi^{(1)}(x),z)}g(y_1^1,z),\ \text{and}\ g(\phi^{(0)}(x),z)=\beta^{t_0^*(\phi^{(1)}(x),z)-1}g(y_0^1,z).$$

Next, by Lemma Lemma B.7 it holds that $y_1^1>y_0^1$, which therefore by Lemma B.3 implies that $g(y_0^1,z)>g(y_1^1,z)$. Thus,

$$\begin{aligned}
w(x,z) &= 1+\beta^{t_0^*(\phi^{(1)}(x),z)+1}\,[1-(1-\alpha)\,x\,]\;g(y_1^1,z)-\beta^{t_0^*(\phi^{(1)}(x),z)}\,g(y_0^1,z)\\
&= 1-\beta^{t_0^*(\phi^{(1)}(x),z)}\left[g(y_0^1,z)-\beta\,[1-(1-\alpha)\,x\,]\;g(y_1^1,z)\right]\\
&> 1-\beta^{t_0^*(\phi^{(1)}(x),z)}\geq 1-\beta\geq 0\quad \left(\text{By Lemma B.9 and } (t_0^*(\phi^{(1)}(x),z)\geq 1)\,\right)\\
&> 1-\beta \qquad \blacksquare
\end{aligned} \tag{B.5}$$

Thus, we conclude that, for case ii), $\phi^{(1)}(x)\leq z$ b) $t_0^*(\phi^{(1)}(x),z)-t_0^*(\phi^{(0)}(x),z)=1$:

$$w(x,z)\quad >\quad 1-\beta\quad \blacksquare \tag{B.6}$$

Next, we address the final case in Lemma 4.1, part (ii) with $z<x\leq\phi_\infty^{(0)}$).

$$(ii)w(x,z)>\frac{(1-\beta)}{(1-\beta)+\beta\,(1-\alpha)\,z}\geq 0\quad \text{for any } x\in(z,\phi_\infty^0],\quad 0\leq\beta\leq 1.$$

$$\begin{aligned} w(x,z) &= g(x,z) - \beta\, g(\phi^{(0)}(x),z) \quad (x > z) && \text{(B.7)} \\ w(x,z) &\geq g(x,z)\,(1-\beta) \qquad \text{(By Lemma B.8 and Lemma B.3)} \\ w(x,z) &> \frac{(1-\beta)}{(1-\beta)+\beta\,(1-\alpha)\,x} > (1-\beta) > 0 \quad \text{(By Lemma B.4)} \quad \blacksquare && \text{(B.8)} \end{aligned}$$

Notice that (B.7) follows from the marginal work definition given in (4.10) for any $x > z$, since in this case it holds that $g(x,z) = 1 + \beta\,[1-(1-\alpha)x]g(\phi^{(1)}(x),z)$. Next, by Lemma B.8 we have that, for any $x \in (z, \phi^{(0)}_\infty]$, it holds that: $x \leq \phi^{(0)}(x)$ and then, by Lemma B.3 we have that $g(x,z) \geq g(\phi^{(0)}(x),z)$.

Next, we address the final case in Lemma 4.1, part (iii) with $\phi^{(0)}_\infty) < x \leq 1$.
The lower bound on $w(x,z)$ in this case coincides with (B.3). By Lemma B.1, for any $x \in [\phi^{(0)}_\infty, 1]$ it holds that: $z < \phi^{(1)}(x) < \phi^{(0)}(x)$. Then, by Lemma B.3 and Lemma B.4 we conclude that:

$$w(x,z) > 1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,z} \quad (x > z)$$

Actually, it follows from Lemma B.14 that a tighter bound can be applied in this case.

$$\begin{aligned} w(x,z) &> 1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta)+\beta\,(1-\alpha)\,\phi^{(1)}_\infty} \quad (x > z \geq \phi^{(0)}_\infty), \\ &> \frac{(1-\beta)+\beta\,(1-\alpha)\,(\phi^{(1)}_\infty - x)}{(1-\beta)+\beta\,(1-\alpha)\,\phi^{(1)}_\infty} \quad (x > z \geq \phi^{(0)}_\infty) \end{aligned}$$

Notice that, for $\beta = 1$, such a lower bound is negative for all $x > \phi^{(0)}_\infty$, since by Lemma B.1 it holds that $\phi^{(1)}_\infty < \phi^{(0)}_\infty$. Thus, the strategy of establishing the positivity of the lower bound on $w(x,z)$ is not useful in this case.
Furthermore, not only the lower bound on $w(x,z)$ but also **the marginal work measure** $w(x,z)$ for $\beta = 1$ can be shown to be negative in this case.

$$\begin{aligned} w(x,z) &= 1 + \beta\,[1-(1-\alpha)\,x\,]\; g(\phi^{(1)}(x),z) - \beta\, g(\phi^{(0)}(x),z) \\ &= 1 + \beta\,[1-(1-\alpha)\,x\,]\left[\sum_{t=0}^{\infty} \beta^t\, \theta(\phi^{(1)}(x),z,t)\right] - \beta\left[\sum_{t=0}^{\infty} \beta^t\, \theta(\phi^{(0)}(x),z,t)\right] \\ &= (1-\beta) + \left[\sum_{t=1}^{\infty} \beta^t\, \theta(x,z,t) - \beta \sum_{t=1}^{\infty} \beta^t\, \theta(\phi^{(0)}(x),z,t)\right] && \text{(B.9)} \end{aligned}$$

Notice that, by Lemma B.2 for $x > \phi^{(0)}_\infty$ it holds that $x > \phi^{(0)}(x)$ and then by Lemma Lemma B.7 it holds that $\phi^{(1)}_t(x) > \phi^{(1)}_t\left(\phi^{(0)}(x)\right)$ for any $t \geq 0$. Thus, it also holds that:

$$\theta(x,z,t) < \theta\left(\phi^{(0)}(x),z,t\right) \quad t \geq 0 \qquad \text{(B.10)}$$

Thus, if $\sum_{t=1}^{\infty}\theta(x,z,t) - \beta\sum_{t=1}^{\infty}\beta^t\theta(\phi^{(0)}(x),z,t) < 0$, then from (B.9) we conclude that $w(x,z) > 0$ if

$$1-\beta > \sum_{t=1}^{\infty}\theta(x,z,t) - \beta\sum_{t=1}^{\infty}\beta^t\theta(\phi^{(0)}(x),z,t) \tag{B.11}$$

From (B.10), it is straightforward to see that for $\beta = 1$, the above condition (B.11) never holds, thus $w(x,z) < 0$ for any $x \in [\phi_{\infty}^{(0)}, 1]$ and $z \in [0, \phi_{\infty}^{1}]$ if $\beta = 1$.
Also, note that, if $x - \beta\phi^{(0)}(x) < 0$ a sufficient condition for $w(x,z)$ to be negative for $\beta < 1$ is

$$(1-\beta) < \beta\,(1-\alpha)\,\left(x - \beta\phi^{(0)}(x)\right)$$

Hence, there exists a β^* for all possible $p^0, p^1, q^0, q^1, \alpha$ such that $w(x,z) > 0$ for all $0 \leq \beta < \beta^*$. Such a β^* is the lowest discount factor β for which (B.11) does not hold. Since the infinite sum in (B.11) does not admit a closed form expression, β^* cannot be computed exactly, yet a lower bound on β^* can be obtained in closed form by, for instance, imposing that the lowest bound on $w(x,z)$ is strictly positive. Hence,

$$\begin{aligned} w(x,z) &> 1-\beta\,(1-\alpha)\,x\,g(\phi^{(1)}(x),z) \\ &> 1-\frac{\beta\,(1-\alpha)}{1-\beta(1-(1-\alpha)\phi_{\infty}^{(1)})} \\ &> \frac{1-\beta\,\left[1-(1-\alpha)(\phi_{\infty}^{(1)}-1)\right]}{1-\beta(1-(1-\alpha)\phi_{\infty}^{(1)})} \end{aligned} \tag{B.12}$$

Thus, by (B.12) we conclude that:

$$\beta^* > \frac{1}{1+\left[1-(1-\alpha)(1-\phi_{\infty}^{(1)})\right]} \triangleq \beta^{(1)} \tag{B.13}$$

Notice that $\beta^{(1)}$ is the lowest value of the maximum discount factor derived following this procedure of imposing that the minimum $w(x,z)$ is strictly positive. Tighter lowest bounds on β^* can be obtained by sequentially approximating $g(\phi^1(1),z)$ with more terms (instead of using its upper bound). For instance, instead of using $g(\phi^1(1),z) < \frac{1}{1-\beta(1-(1-\alpha)\phi_{\infty}^{(1)})}$ we could consider:

$$g(\phi^1(1),z) < 1+\beta\left[1-(1-\alpha)(p^1+\rho^1)\right]\left(\frac{1}{1-\beta(1-(1-\alpha)\phi_{\infty}^{(1)})}\right)$$

Thus, we would obtain a discount factor $\beta^{(2)}$ for which (B.14) holds.

$$\begin{aligned} w(x,z) &> 1-\beta\,(1-\alpha)\,x\,g(\phi^{(1)}(x),z) \\ &> 1-\beta\,(1-\alpha)\left[1+\left(\frac{\beta\left[1-(1-\alpha)(p^{(1)}+\rho^{(1)})\right]}{1-\beta(1-(1-\alpha)\phi_\infty^{(1)})}\right)\right] \end{aligned} \tag{B.14}$$

Where $\beta^* > \beta^{(2)} > \beta^{(1)}$. By approximating $g(\phi^{(1)}(x),z)$ using n terms of the infinite sum assuming the largest value of the $n+1$ remaining terms we can obtain further lower bounds on β^* which we denote by β^n, such that $\beta^* > \cdots > \beta^{(n)} > \cdots > \beta^{(2)} > \beta^{(1)}$.

Thus, we conclude that $w(x,z) > 0$ when $x \in [\phi_\infty^{(0)}, 1]$ and $z \in [0, \phi_\infty^{(1)}]$ only for $\beta < \beta^*$ where β^* is the lowest discount factor for which (B.9) is strictly positive. ■

For proving Lemma 4.1 we have deployed the following results and properties of the Möbius transformations, listed as lemmas. Further, they shall be invoked for proving the rest of results shown in this Appendix.

Lemma B.1. *For any $p^{(0)} > p^{(1)}$, $q^{(0)} < q^{(1)}$ and $\rho^{(1)} > 0$:*

$$\phi_t^{(1)}(x) < \phi_t^{(0)}(x) \qquad \textit{for all } t \geq 1,$$

From which it follows that: a) $\phi^{(1)}(x) < \phi^{(0)}(x)$ and b) $\phi_\infty^{(1)} < \phi_\infty^{(0)}$

Proof. We shall prove it by induction.
For $t = 1$ we have that:

$$\phi^{(1)}(x) = (p^{(1)} + \rho^{(1)}x) - \rho^{(1)}x\frac{(1-\alpha)}{1-(1-\alpha)x}. \tag{B.15}$$

Thus, for any $\rho^{(1)} \geq 0$, $\phi^{(1)}(x) \leq (p^{(1)} + \rho^{(1)}x)$ while $\phi^{(0)}(x) = (p^{(0)} + \rho^{(1)}x)$. Then, $\phi^{(0)}(x) - \phi^{(1)}(x) \geq (p^{(0)} - p^{(1)})\,(1-x) + (q^{(1)} - q^{(0)})\,x\ > 0$,
for any $p^{(0)} > p^{(1)}$ and $q^{(0)} < q^{(1)}$, as it occurs in this model. Then, for $t = 1$ it holds that $\phi^{(1)}(x) < \phi^{(0)}(x)$.

Next, we assume it true for t, i.e. $\phi_t^{(1)}(x) < \phi_t^{(0)}(x)$ and consider the case for $t+1$. Notice that: $\phi_{t+1}^{(a)} = \phi^{(a)}(\phi_t^{(a)}(x))$ for $a = 0, 1$, then by (B.15):

$$\phi_{t+1}^{(1)}(x) \leq p^{(1)} + \rho^{(1)}\phi_t^{(1)}(x).$$

Finally, since $\phi_t^{(1)}(x) < \phi_t^{(0)}(x)$ we conclude that:

$$\phi_{t+1}^{(0)}(x) - \phi_{t+1}^{(1)}(x) \geq (p^{(0)} - p^{(1)})\,(1 - \phi_t^{(0)}(x)) + (q^{(1)} - q^{(0)})\,\phi_t^{(0)}(x) \ > 0.$$

□

Lemma B.2. *For $a = 0, 1$ and any $x < \phi_\infty^{(a)}$, $0 \leq t < \infty$, $\phi_t^{(a)}(x) < \phi_\infty^{(a)}$, whereas, for any $x > \phi_\infty^{(a)}$, $\phi_t^{(a)}(x) > \phi_\infty^{(a)}$.*

Proof. It follows straightforwardly from the result (A.8) in the appendix on Möbius transformations implying that $\phi_t^{(a)}(x)$ is a deceasing function in t to the right of its fixed point $\phi_\infty^{(a)}$ and a increasing function in t to the left of its fixed point. □

Lemma B.3. *For any x, y such that $z < x < y \leq 1$ with $z \in [0, \phi_\infty^{(1)}]$,*

$$g(x, z) > g(y, z)$$

Proof. By results (A.18) and (A.14) it holds that:

$$\phi_t^{(1)}(y) > \phi_t^{(1)}(x) > \phi_\infty^{(1)} \qquad \text{if } y > x > z \geq \phi_\infty^{(1)}, \text{ for } 0 \leq t < \infty.$$

Therefore, for any $x, y \in (z, 1] : x < y$ with $z \in [0, \phi_\infty^{(1)}]$ it holds that

$$1 - (1-\alpha)\,\phi_t^{(1)}(y) < 1 - (1-\alpha)\,\phi_t^{(1)}(x) < 1 - (1-\alpha)\,\phi_\infty^{(1)}, \qquad \text{for } 0 \leq t < \infty.$$

Hence, $\theta(y, z, t) < \theta(x, z, t)$ for all $t < \infty$ while $\theta(y, z, \infty) = \theta(x, z, \infty)$. Thus, from (4.14) it follows that: $g(x, z) > g(y, z)$. □

Lemma B.4. *For any $x \in (z, 1]$ with $z \in [0, \phi_\infty^{(1)}]$,*

$$\frac{1}{1 - \beta\,(1 - (1-\alpha)\,x)} < g(x, z) < \frac{1}{1 - \beta\,(1 - (1-\alpha)\,z)}, \qquad z \leq \phi_\infty^{(1)} < x$$

$$\frac{1}{1 - \beta\,(1 - (1-\alpha)\,\phi_\infty^{(1)})} < g(x, z) < \frac{1}{1 - \beta\,(1 - (1-\alpha)\,z)}, \qquad z < x \leq \phi_\infty^{(1)}$$

Proof. By result (A.18) it holds that:

$$(a)\ x > \phi_t^{(1)}(x) > \phi_\infty^{(1)} \geq z \qquad \text{if } x > \phi_\infty^{(1)} \geq z, \text{ for } 1 \leq t < \infty, \qquad \text{and}$$

$$(b)\ \phi_\infty^{(1)} > \phi_t^{(1)}(x) > x > z \qquad \text{if } \phi_\infty^{(1)} \geq x > z, \text{ for } 1 \leq t < \infty.$$

Therefore, for (a) and (b) respectively it holds that

$$1-(1-\alpha)\,x < 1-(1-\alpha)\,\phi_t^{(1)}(x) < 1-(1-\alpha)\,z, \quad \text{for } 1 \leq t < \infty.$$

$$1-(1-\alpha)\,\phi_\infty^{(1)} < 1-(1-\alpha)\,\phi_t^{(1)}(x) < 1-(1-\alpha)\,z, \quad \text{for } 1 \leq t < \infty.$$

Hence, $[1-(1-\alpha)\,x]^t < \theta(x,z,t) < [1-(1-\alpha)\,z]^t$ for all $0 < t < \infty$ and $\left[1-(1-\alpha)\,\phi_\infty^{(1)}\right]^t < \theta(x,z,t) < [1-(1-\alpha)\,z]^t$ for all $0 < t < \infty$. Thus, from (4.14) it follows that:

$$\sum_{t=0}^{\infty}\beta^t\,\left[1-(1-\alpha)\,x\right]^t < g(x,z) < \sum_{t=0}^{\infty}\beta^t\,\left[1-(1-\alpha)\,z\right]^t$$

$$\sum_{t=0}^{\infty}\beta^t\,\left[1-(1-\alpha)\,\phi_\infty^{(1)}\right]^t < g(x,z) < \sum_{t=0}^{\infty}\beta^t\,\left[1-(1-\alpha)\,z\right]^t$$

□

Lemma B.5. *For any $x \in (0,z]$,*

$$t_0^*(\phi^{(1)}(x),z) - t_0^*(\phi^{(0)}(x),z) \in \{0,1\}$$

Proof. From the definition of $t_0^*(x,z)$ provided in Chapter 4 it follows that such a function will be a *floor* function since, for $x \leq z$, $t_0^*(x,z)$ satisfies:

$$\phi^{(0)}_{t_0^*(x,z)-1}(x) \leq z < \phi^{(0)}_{t_0^*(x,z)}(x).$$

Thus, using the property below:

$$\lfloor x \rfloor - \lfloor y \rfloor \leq \lfloor x-y \rfloor,$$

it can be shown that the maximum value for $t_0^*(\phi^{(1)}(x),z) - t_0^*(\phi^{(0)}(x),z)$ is 1 since the function $t_0^*(x,z)$ is decreasing in x for a given z and $\phi^{(1)}(x) < \phi^{(0)}(x)$. □

Lemma B.6. *For any $x \in (z,1]$,*

$$t_1^*(\phi^{(0)}(x),z) - t_1^*(\phi^{(1)}(x),z) \in \{0,1\}$$

Proof. From the definition of $t_1^*(x,z)$ provided in Chapter 4 it follows that such function will be a *ceiling* function since for $x > z$, $t_1^*(x,z)$ satisfies:

$$\phi^{(1)}_{t_1^*(x,z)}(x) < z \leq \phi^{(1)}_{t_1^*(x,z)-1}(x).$$

Thus, using the property below:

$$\lceil x \rceil - \lceil y \rceil \leq \lceil x - y \rceil + 1,$$

it can be shown that the maximum value for $t_1^*(\phi^{(0)}(x), z) - t_1^*(\phi^{(1)}(x), z)$ is 1 since the function $t_1^*(x, z)$ is increasing in x for a given z and $\phi^{(1)}(x) < \phi^{(0)}(x)$. □

Lemma B.7. *For $a = 0, 1$ and any $x < y < \phi_\infty^{(a)}$, $0 \leq t < \infty$, it holds that:* $\phi_t^{(a)}(x) < \phi_t^{(a)}(y) < \phi_\infty^{(a)}$ *whereas for any $x > y > \phi_\infty^{(a)}$, $\phi_t^{(a)}(x) > \phi_t^{(a)}(y) > \phi_\infty^{(a)}$.*

Proof. We begin the proof by invoking results (A.17) and (A.18) from the appendix on Möbius transformations, from which it follows that:

$$\frac{\partial \phi_t^{(a)}(x)}{\partial x} > 0 \qquad \text{for } a = 0, 1 \text{ for all } t < \infty \tag{B.16}$$

□

Lemma B.8. *For $a = 0, 1$ and any $x < \phi_\infty^{(a)}$, $0 \leq t < \infty$, it holds that:* $\phi_t^{(a)}(x) < \phi_{t+1}^{(a)}(x) < \phi_\infty^{(a)}$ *whereas, for any $x > \phi_\infty^{(a)}$, $\phi_t^{(a)}(x) > \phi_{t+1}^{(a)}(x) > \phi_\infty^{(a)}$.*

Proof. This proof is completed by invoking results (A.13) and (A.14) from the appendix on Möbius transformations, from which it follows that:

$$\frac{\partial \phi_t^{(a)}(x)}{\partial t} = \text{sgn}\Big(\phi_\infty^{(a)} - x\Big) \qquad \text{for } a = 0, 1.$$

Thus, for $x < \phi_\infty^{(a)}$, $\frac{\partial \phi_t^{(a)}(x)}{\partial t} > 0$, and hence from result (A.8) it holds that $\phi_t^{(a)}(x) < \phi_\infty^{(a)}$ for all $t < \infty$, whereas, for $x > \phi_\infty^{(a)}$, $\frac{\partial \phi_t^{(a)}(x)}{\partial t} < 0$, and hence it holds that $\phi_t^{(a)}(x) > \phi_\infty^{(a)}$ for all $t < \infty$. □

Lemma B.9.

$$g(y_0^1, z) - \beta\ [1 - (1 - \alpha)\, x\,]\ g(y_1^1, z < 1.$$

Sketch of Proof:
The total work measure $g(y_0^1, z) < g(y_1^1, z)$, yet $|g(y_0^1, z) - \beta\ [1 - (1 - \alpha)\, x\,]\, g(y_1^1, z)| \leq 1$. To see this, first bound the maximum difference in the to work processes by the first difference by $|d| < 1$, then it follows that $|g(y_0^1, z) - [1 - (1 - \alpha)\, x\,]\, g(y_1^1, z)| < d\frac{1-\beta\,[1-(1-\alpha)\,x\,]}{1-\beta\,[1-(1-\alpha)\,z\,]} \leq 1$ ■

Next, we continue with the proof of Lemma 4.3 invoked in Chapter 4 to show Proposition 4.3.

Proof of Lemma 4.3

For any threshold $z \in [\phi^{(1)}_\infty, \phi^{(0)}_\infty)$, we summarize the main results regarding information state cycles under a z-threshold policy in results Lemma B.10, Lemma B.11 and Lemma B.12.

Lemma B.10. *For $z \in [\phi^{(1)}_\infty, \phi^{(0)}_\infty)$ and $x \in (0, 1]$, as long as $x_t \neq 0$, (i.e, as long as the target remains* unhunted*), the* hitting time *for the belief state process to the interval $(\phi^{(1)}(z), \phi^{(0)}(z)]$ is* finite, *and once the belief state reaches this set of states, the probability of abandoning it is zero. The subset $(\phi^{(1)}(z), \phi^{(0)}(z)]$, as long as $x_t \neq 0$, is "*absorbing*" (i.e., it is never abandoned until the target is hunted).*

Proof. If $x \leq z$ after $t_0^*(x, z)$ slots under deterministic dynamics, we reach the set $(z, \phi^{(0)}(z)]$. While if $x > z$ and if after $t_1^*(x, z)$ active slots the target has not been found, then after $t_1^*(x, z) + t_0^*(\phi^{(1)}_{t_1^*(x,z)}(x), z)$ periods the set $(z, \phi^{(0)}(z)]$ is reached. Notice that, the maximum value of x_t to reach the active set $B(z)$ coming from the passive set $B(z)^c$ is $\phi^{(0)}(z)$; then the minimum value of the information state to reach the passive set $B(z)^c$ coming from $(z, \phi^{(0)}(z)]$ is $\lim_{x \to z^+} \phi^{(1)}(x) = \phi^{(1)}(z)$. Thus, once x_t is in $(z, \phi^{(0)}(z)]$ we know that the interval $(\phi^{(1)}(z), \phi^{(0)}(z)]$ is never abandoned, alternating infinitely within it between the interval $(\phi^{(1)}(z), z] \subset B(z)^c$ (passive slots), and the interval $(z, \phi^{(0)}(z)] \subset B(z)$ (active slots), until the target is found.

□

Furthermore, as stated in result Lemma B.11, within that "*absorbing*" set of states the *possible composition of cycles (in terms of the concrete sequence of active/passive time slots) is reduced to three cases: case 1: 1 passive slot & A active slots, with $A \geq 2$; case 2: 1 passive slot & 1 active slot; case 3: P passive slots &* 1 *active slot1, with $P \geq 2$.*

Lemma B.11. *(a) For $z \in (\phi^{(1)}_\infty, \phi^0(\phi^{(1)}_\infty))$: if $x \in (\phi^{(1)}(z), z]$, then $t_0^*(x, z) = 1$; If $x \in (z, \phi^{(0)}(z)]$, $t_1^*(x, z) > 1$.*

(b) For $z \in (\phi^1(\phi^{(0)}_\infty), \phi^0(\phi^{(1)}_\infty))$: if $x \in (\phi^{(1)}(z), z]$ then $t_0^*(x, z) = 1$. If $x \in (z, \phi^{(0)}(z)]$, then $t_1^*(x, z) = 1$.

(c) For $z \in (\phi^0(\phi^{(1)}_\infty), \phi^{(0)}_\infty)$,: if $x \in (\phi^{(1)}(z), z]$, then $t_0^*(x, z) > 1$. If $x \in (z, \phi^{(0)}(z)]$, then $t_1^*(x, z) = 1$.

Proof. For case (i) and (ii) notice that, if $x, z \in (\phi^{(1)}_{\infty}, \phi^{(0)}(\phi^{(1)}_{\infty}))$, it is easy to see that $t^*_0(x, z) = 1$; and by the same reasoning, if $x, z \in (\phi^{(1)}(\phi^{(0)}_{\infty}), \phi^{(0)}_{\infty})$ then $t^*_1(x, z) = 1$.
Thus, for $z \in (\phi^{(1)}_{\infty}, \phi^{(1)}(\phi^{(1)}_{\infty}))$ the process x_t will jump above $(\phi^{(1)}(z), z]$ only after 1 passive time slot, while for $z \in (\phi^{(1)}(\phi^{(0)}_{\infty}), \phi^{(0)}_{\infty})$ the process x_t will leave the interval $(z, \phi^{(0)}_{\infty})$ just after 1 active time slot.
Finally, since $\phi^{(1)}(\phi^{(0)}_{\infty}) < \phi^{(0)}(\phi^{(1)}_{\infty})$ for $z \in (\phi^{(1)}(\phi^{(0)}_{\infty}), \phi^{(0)}(\phi^{(1)}_{\infty}))$ the process x_t will evolve in such a way that there will be 1 active and 1 passive period when searching the sites under a threshold policy. Since, if $x, z \in (\phi^{(1)}(\phi^{(0)}_{\infty}, \phi^{(0)}(\phi^{(1)}_{\infty}))$ then $t^*_0(x, z) = 1$ and $t^*_1(x, z) = 1$. An analogous reasoning follows for case (iii).

□

Notice from Lemma B.11 the existence of belief state cycles under a z-threshold policy different from the ones described above is ruled out.

Lemma B.12. *Furthermore, for cases (i) and (iii) it may also occur that the active slots or the passive slots respectively composing the cycle generate* regular *cycles or* irregular *cycles as described by the lemma below.*

(a) For $z \in (\phi^{(1)}_{\infty}, \phi^{0}(\phi^{(1)}_{\infty}))$:, $\quad a.1)\ \forall x \in (z, \phi^{(0)}(z)]$ *then* $t^*_1(x, z) = A \geq 2$

$a.2)\ \forall x \in (z, x^*]$ *then* $t^*_1(x, z) = A \geq 2$ *, and*

$a.3) x \in (x^*\phi^{(0)}(z)]$ *then* $t^*_1(x, z) = A + 1$

Further in cases a.2) and a.3) it holds that $\quad \phi^{(0)}(\phi^{(1)}_{A}(x)) \in (x^*\phi^{(0)}(z)]$ *and* $\phi^{(0)}(\phi^{(1)}_{A+1}(x)) \in (z, x^*]$

(b) For $z \in (\phi^{(0)}(\phi^{(1)}_{\infty}), \phi^{(0)}_{\infty})$:, $\quad b.1)\ \forall x \in (\phi^{(1)}(z), z],$ *then* $t^*_0(x, z) = P \geq 2$

$b.2)\ \forall x \in (x^*, z], t^*_0(x, z) = P \geq 2$ *, and*

$b.3) \forall x \in (\phi^{(1)}(z), x^*] :$ *then* $t^*_0(x, z) = P + 1$

Further in cases b.2) and b.3) it holds that $\quad \phi^{(1)}(\phi^{(0)}_{P}(x)) \in (\phi^{(1)}(z), x^*]$ *and* $\phi^{(1)}(\phi^{(0)}_{P+1}(x)) \in (x^*, z]$

Sketch of Proof:

The above result follows from properties of the floor and ceiling function defining $t^*_0(x, z)$ and $t^*_1(x, z)$. For some threshold within each case, the absorbing set of states will have $t^*_0(x, z)$ or $t^*_1(x, z)$ yielding two possible results (whose difference is 1). The properties of the active and passive dynamics will next guarantee that these two possible values (either for the active or the passive slots) occur alternating.
Next, with the above Lemmas we shall prove Lemma 4.3.

For $x \in (0, z]$, with z in $[\phi_\infty^{(1)}, \phi_\infty^{(0)})$ such that:

(ii) $z \in (\phi^{(1)}(\phi_\infty^{(0)}), \phi^{(0)}(\phi_\infty^{(1)}))$ as in Lemma B.11 holds; or z as in (i) a) of Lemma B.12; or z as in (iii) a) of Lemma B.12 with $t_0^*(\phi^{(1)}(x), z) = t_0^*(\phi^{(0)}(x), z)$;

it holds that:

$$\begin{aligned}
w(x,z) &= 1 + \beta\ [1 - (1-\alpha)\, x\,]\ g(\phi^{(1)}(x), z) - \beta\, g(\phi^{(0)}(x), z) \\
&> 1 + \beta\ [1 - (1-\alpha)\, x\,]\ g(\phi^{(1)}(x), z) - \beta\, g(\phi^{(1)}(x), z) \quad \text{(By Lemma B.13)} \\
&> 1 - \beta^{t_0^* \phi^{(1)}(x), z)}\ \beta(1-\alpha)\, x\, g(y_1^1, z) \quad (y_1^1 > z) \\
&> 1 - \beta^{t_0^* \phi^{(1)}(x), z)}\ \frac{\beta\,(1-\alpha)\,x}{(1-\beta) + \beta\,(1-\alpha)\,\phi_\infty^{(1)}} \geq 0 \quad \Big(\text{By Lemma B.14 and } (x \leq \phi_\infty^{(1)})\Big) \\
&> 1 - \frac{\beta\,(1-\alpha)\,x}{(1-\beta) + \beta\,(1-\alpha)\,\phi_\infty^{(1)}} \quad \blacksquare \qquad \text{(B.17)}
\end{aligned}$$

For the remaining threshold values z in $(\phi_\infty^{(1)}, \phi_\infty^{(0)})$:
z as in (i) b) of Lemma B.12; or z is as in (iii) a) of Lemma B.12 with $t_0^*(\phi^{(0)}(x), z) + 1 = t_0^*(\phi^{(1)}(x), z)$ or z as in (iii) b) of Lemma B.12 It holds that:

$$\begin{aligned}
w(x,z) &= 1 + \beta^{t_0^*(\phi^{(1)}(x), z)+1}\ [1 - (1-\alpha)\, x\,]\ g(y_1^1, z) - \beta^{t_0^*(\phi^{(1)}(x), z)}\ g(y_0^1, z) \\
&= 1 - \beta^{t_0^*(\phi^{(1)}(x), z)}\ \left[g(y_0^1, z) - \beta\ [1 - (1-\alpha)\, x\,]\ g(y_1^1, z)\right] \\
&> 1 - \beta^{t_0^*(\phi^{(1)}(x), z)} \geq 1 - \beta \geq 0 \quad \big(\text{By Lemma B.15 and } (t_0^*(\phi^{(1)}(x), z) \geq 1)\ \big) \\
&> 1 - \beta \qquad \blacksquare \qquad \text{(B.18)}
\end{aligned}$$

Lemma B.13. *For any $x, y \in (z, \phi_\infty^{(0)}]$: $x < y$, and (ii) $z \in (\phi^{(1)}(\phi_\infty^{(0)}), \phi^{(0)}\phi_\infty^{(1)}))$; or z as in (i) a) of Result Lemma B.12; or z as in (iii) a) of Result Lemma B.12*

$$g(x, z) > g(y, z)$$

Proof. For (ii) $z \in (\phi^{(1)}(\phi_\infty^{(0)}), \phi^{(0)}(\phi_\infty^{(1)}))$, the process a_t (starting at either x or y and as long as the target remains unhunted) will be:

$$a_t = \begin{cases} 1, & \text{for } t = 2\,n, \qquad n = 0, 1, 2, \ldots, \\ 0, & \text{for } t = 2\,n + 1, \quad n = 0, 1, 2, \ldots. \end{cases} \qquad \text{(B.19)}$$

Given that $x < y$, from proposition Lemma B.7 in $t = 1$ it follows that $\phi^{(1)}(x) < \phi^1(y)$. This fact ensures that the belief state starting in y will always be above the process x,

and this in turn implies that $g(x,z) > g(y,z)$.
Let us show it by induction, if we know that for $t = 0$ it holds that: $x < y$ thus $\phi^{(1)}(x) < \phi^{1}(y)$ in $t = 1$, hence in $t = 2$: $\phi^{(0)}(\phi^{(1)}(x)) < \phi^{(0)}(\phi^{1}(y))$. Let us assume it true for some t, so that for $t = 2\,n$ the belief state process is such that $x_t < y_t$, hence it holds that $\phi^{(1)}(x_t) < \phi^{(1)}(y_t)$ in $t+1 = 2\,n+1$ and $\phi^{(0)}(\phi^{(1)}(x_t)) < \phi^{(0)}(\phi^{(1)}(y_t))$ in $t+2 = 2\,n+2$. Then, if true in t it is also true in $t+1$ since it holds that $x_{t+1} < y_{t+1}$, and this in turn implies that $x_{t+2} < y_{t+2}$ by proposition Lemma B.7 (both for t odd and even).

For (i) a) as in Lemma B.12, it will hold that $t_1^*(x,z) = t_1^*(y,z) = A$ (with $A \in \{1,2,\dots\}$) and the process a_t (starting at either x or y and as long as the target remains unhunted) will be:

$$a_t = \begin{cases} 1, & \text{for } t = n\,(A+1)\ \dots\ t = n\,(A+1)+(A-1), \quad n = 0,1,2,\dots, \\ 0, & \text{for } t = n\,(A+1)+A, \quad n = 0,1,2,\dots. \end{cases} \tag{B.20}$$

Given that $x < y$, from proposition Lemma B.7 in $t = 1$ it follows that $\phi_A^{(1)}(x) < \phi_A^{(1)}(y)$ and also $\phi^{(0)}(\phi_A^{(1)}(x) < \phi^{(0)}(\phi_A^{(1)}(y)$. These facts ensure that the belief state starting in y will always be above the process x, and this in turn implies that $g(x,z) > g(y,z)$.

For (iii) a) as in Lemma B.12, it will hold that $t_0^*(x,z) = t_0^*(y,z) = P$, thus the process a_t (starting at either x or y) will be:

$$a_t = \begin{cases} 1, & \text{for } t = n\,(P+1), \quad n = 0,1,2,\dots, \\ 0, & \text{for } t = n\,(P+1)+1\dots t = n\,(P+1)+P, \quad n = 0,1,2,\dots. \end{cases} \tag{B.21}$$

Given that $x < y$, from proposition Lemma B.7 in $t = 1$ it follows that $\phi^{(1)}(x) < \phi^{(1)}(y)$ and $\phi_P^{(0)}(\phi^{(1)}(x)) < \phi_P^{(0)}(\phi^{1}(y))$. These facts ensure that the belief state starting in y will always be above the process x, and this in turn implies that $g(x,z) > g(y,z)$.

□

Lemma B.14. *For any $x \in (z,1]$, with $z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$,*

$$1 < g(x,z) < \frac{1}{1-\beta(1-(1-\alpha)\phi_\infty^{(1)})}$$

Sketch of Proof:
This follows from the fact that as $z \searrow \phi_\infty^{(1)}$, the cycle tends to be composed of a larger number of active periods $A \to \infty$ (as in Case I of the possible thresholds). As when

$z \nearrow \phi_\infty^{(0)}$ the cycle tends to be composed of a larger number of passive periods $P \to \infty$ (as in Case III of the possible thresholds). These arguments explain the above bounds on the work measure for the case II of the possible thresholds.

Lemma B.15. *For any $x \in (0, 1]$, let y_a^1 be the first value of the belief state process to reach the set $(z, \phi_\infty^{(0)}]$having selected the action $a = 0, 1$ at the initial state x and following a z-threshold policy thereafter, if z is as in (i) b) of Lemma B.12; or as in (i) a) of Result B.12 with $t_1^*(\phi^{(0)}(x), z) + 1 = t_1^*(\phi^{(1)}(x), z)$ for $x > z$; or z as in (iii) b) of Result B.12; or z is as in (iii) a) of Result B.12 with $t_0^*(\phi^{(0)}(x), z) + 1 = t_0^*(\phi^{(1)}(x), z)$ for $x \leq z$.*

$$g(y_0^1, z) - g(y_1^1, z) < 1 \Longrightarrow g(y_0^1, z) - \beta(1 - (1 - \alpha)x)g(y_1^1, z) < 1$$

Sketch of Proof:
In all these cases it holds that $y_0^1 < y_1^1$, and this fact ensures that either the belief state starting in x and active will always be above the process starting in x and passive, or that both processes will be intertwined (as in (i) b) or (iii) b) where cycles are *alternating* in their composition). Yet, in all cases it holds that:

Next, we continue to prove part (ii) of Lemma 4.3.
If $x \in (z, \phi_\infty^{(0)}]$, with z in $[\phi_\infty^{(1)}, \phi_\infty^{(0)})$ such that (ii) $z \in (\phi^{(1)}(\phi_\infty^{(0)}), \phi^{(0)}(\phi_\infty^{(1)}))$ as in Lemma B.11; or z as in (i) a) of Lemma B.12 with $t_1^*(x, z) = t_1^*(\phi^{(0)}(x), z)$; or z as in (iii) a) of Lemma B.12; It holds that $x < \phi^{(0)}(x)$ and hence:

$$\begin{aligned} w(x, z) &= g(x, z) - \beta\, g(\phi^{(0)}(x), z) \quad (x > z) && \text{(B.22)} \\ &\geq g(x, z)\,(1 - \beta) \qquad \text{(By Lemma B.8 and B.13)} \\ &> (1 - \beta) > 0 \qquad \text{(By Lemma B.14)} \quad \blacksquare && \text{(B.23)} \end{aligned}$$

For the remaining threshold values z in $(\phi_\infty^{(1)}, \phi_\infty^{(0)})$:
z as in (i) a) of Lemma B.12 with $t_1^*(x, z)+1 = t_1^*(\phi^{(0)}(x), z)$; z or as in (i) b) of Lemma B.12; or z is as in (iii) b) of Lemma B.12, it holds that:

$$\begin{aligned} w(x, z) &= 1 + \beta[1 - (1 - \alpha)\,x]\, g(\phi^{(1)}(x), z) - \beta\, g(\phi^{(0)}(x), z) \\ &= 1 + \beta[1 - (1 - \alpha)\,x]\, g(y_1^1, z) - \beta\, g(y_0^1, z) \\ &> 1 - \beta\left(g(y_0^1, z) - [1 - (1 - \alpha)\,x]\, g(y_1^1, z)\right) \\ &> (1 - \beta) \quad \text{(By Lemma B.15)} \quad \blacksquare \end{aligned}$$

Next, we prove part (iii) of Lemma 4.3.

Let $t_1^*(x, \phi_\infty^{(0)})$ be the number of active slots until we reach the "*absorbing*" set of states

$(\phi^{(1)}(z), \phi^{(0)}(z))$ for any $z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$ starting from some $x > \phi_\infty^{(0)}$. Let the value of the first belief state to reach the "*recurrent*" set after being active/passive in the first slot be denoted as $y_1^0 \triangleq \left(\phi^1_{t_1^*(\phi^{(1)}(x), \phi_\infty^{(0)})}(\phi^{(1)}(x))\right)$ and $y_0^0 \triangleq \left(\phi^1_{t_1^*(\phi^{(0)}(x), \phi_\infty^{(0)})}(\phi^{(0)}(x))\right)$. Then, it holds that:

$$\begin{aligned}
w(x,z) \;&=\; 1 + \beta[1-(1-\alpha)\,x]\left[\sum_{t=0}^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})-1} \beta^t\,\theta(\phi^{(1)}(x), \phi_\infty^{(0)}, t)\right] + \\
&\quad \beta^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)}))}\,\theta\left(\phi^{(1)}(x), \phi_\infty^{(0)}, \left[t_1^*(\phi^{(1)}(x), \phi_\infty^{(0)}) - 1\right]\right)\; g(y_0^1, z)\ldots \\
&\quad -\beta\left[\sum_{t=0}^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)})-1} \beta^t\,\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, t)\right] - \\
&\quad \beta^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)}))}\,\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, \left[t_1^*(\phi^{(0)}(x), \phi_\infty^{(0)}) - 1\right])g(y_0^0, z) \\
&=\; (1-\beta) + \\
&\quad \left(\left[\sum_{t=1}^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})} \beta^t\,\theta(x, \phi_\infty^{(0)}, t)\right] - \beta\left[\sum_{t=1}^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)})-1} \beta^t\,\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, t) - 1\right]\right) + \\
&\quad \beta^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})}\,\theta\left(x, \phi_\infty^{(0)}, \left[t_1^*(\phi^{(1)}(x), \phi_\infty^{(0)})\right]\right)\; g(y_0^1, z) \\
&\quad -\beta^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)} z))}\,\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, \left[t_1^*(\phi^{(0)}(x), \phi_\infty^{(0)}) - 1\right])g(y_0^0, z)
\end{aligned} \tag{B.24}$$

Define further,

$$w^{(3)}(x,z) \;=\; (1-\beta) + \left(\left[\sum_{t=1}^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})} \beta^t\,\theta(x, \phi_\infty^{(0)}, t)\right] - \beta\left[\sum_{t=1}^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)})-1} \beta^t\,\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, t)\right]\right)$$

By proposition Lemma B.6, we have that:

$$w^{(3)}(x,z) \;>\; (1-\beta) + \left(\sum_{t=1}^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})} \beta^t\,\left(\theta(x, \phi_\infty^{(0)}, t) - \beta\,\left[\theta(\phi^{(0)}(x), \phi_\infty^{(0)}, t)\right]\right)\right) \triangleq w_{\min}^{(3)}(x, \phi_\infty^{(0)})$$

Then, it holds that

$$\begin{aligned} w(x,z) \;&>\; w^{(3)}(x,z) + \beta^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})}\, \theta\left(x, \phi_\infty^{(0)}, \left[t_1^*(\phi^{(1)}(x), \phi_\infty^{(0)}) - 1\right]\right) \times \\ &\left(\beta^d\, d\, \left[1 - (1-\alpha)\, \left(\phi^1_{t_1^*(\phi^{(1)}(x),z)-1}(\phi^{(1)}(x))\right)\right]\, g(y_0^1, z) - g(y_0^0, z)\right) \end{aligned}$$

Where $d \in \{0,1\}$ according to Lemma B.6. For $d = 0$, $y_0^1 < y_0^0$ and from Lemma B.13 we conclude that $g(\ y_0^1, z) - g(\ y_0^0, z) > 0$. For $d = 1$, $y_0^1 > y_0^0$ and from Lemma B.15 we conclude that $g(\ y_0^0, z) - g(\ y_0^1, z) < 1$.

It follows that $w(x,z)$ will be positive as long as $w^{(3)}(x,z) > 0$. But, $w^{(3)}(x,z)$ may be negative. From Lemma B.2 and for any $x > \phi_\infty^{(0)}$ it follows that $x > \phi^{(0)}(x)$, and then by Lemma B.7 it holds that $\phi_t^1(x) > \phi_t^1(\phi^{(0)}(x))$ for any $0 \le t < \infty$. Note that for $\beta = 1$, and $t_1^*(\phi^{(1)}(x), \phi_\infty^{(0)}) \ge 1$, it holds that $w^{(3)}(x,z) < 0$ since $\phi_t^1(x) > \phi_t^1(\phi^{(0)}(x))$ implies that $\theta(x,z,t) < \theta(\phi^{(0)}(x), z, t)$. Notice that $w^{(3)}(x,z) > 0$ when $w^{(3)}(x,z) - (1-\beta) < 0$ iff

$$(1-\beta) > \left(\left[\sum_{t=1}^{t_1^*(\phi^{(1)}(x),\phi_\infty^{(0)})} \beta^t\, \theta(x, \phi_\infty^{(0)}, t)\right] - \beta\, \left[\sum_{t=0}^{t_1^*(\phi^{(0)}(x),\phi_\infty^{(0)})-1} \beta^t\, \theta(\phi^{(0)}(x), \phi_\infty^{(0)}, t)\right]\right) \tag{B.25}$$

Denote by β_2^* the lowest discount factor β for which (B.25) does not hold. Then, it holds that for $x \in (\phi_\infty^{(0)}, 1]$: $w(x,z) > 0$ for $\beta < \beta_2^*$.
It is straightforward to see that $\beta_2^* > \beta^*$, given that for deriving β^* we must consider an infinite sum of terms, instead of a finite one, as in this case.
Thus, we conclude that, for $x \in (\phi_\infty^{(0)}, 1]$ when $z \in [\phi_\infty^{(1)}, \phi_\infty^{(0)})$, $w(x,z) > 0$ for $\beta < \beta^*$. ■

Next, we prove Lemma 4.5. We start by part (i) by which for any threshold $z \ge \phi_\infty^{(0)}$, $w(x,z) = 1$.

In this threshold case the marginal work measure can be computed in closed form. Notice that by proposition Lemma B.2 any $x \le z$ is such that $\phi^{(0)}(x) \le z$ and $\phi^{(1)}(x) \le z$, which hence implies from (4.26) that $g(\phi^{(0)}(x), z) = g(\phi^{(1)}(x), z) = 0$. Thus, from (4.10) we conclude that $w(x,z) = 1$.

Now, we prove part (ii) in Lemma 4.5 by which for any threshold $z \ge \phi_\infty^{(0)}$, $w(x,z) > 0$ for any $x \in (z, 1]$ and $\beta < \beta^*$.

It follows from (4.26) and (4.10) that:

$$\begin{aligned} w(x,z) &= 1+\beta[1-(1-\alpha)\,x]\left[\sum_{t=0}^{t_1^*(\phi^{(1)}(x),z)-1} \beta^t\,\theta(\phi^{(1)}(x),z,t)\right] \\ &\quad -\beta\left[\sum_{t=0}^{t_1^*(\phi^{(0)}(x),z)-1} \beta^t\,\theta(\phi^{(0)}(x),z,t)\right] \end{aligned} \tag{B.26}$$

Define $t_1^*(x,z)=0$ for any $x \le z$. Then by Proposition Lemma B.6 we can bound (B.26) as follows:

$$\begin{aligned} w(x,z) &\geq 1+\beta[1-(1-\alpha)\,x]\left[\sum_{t=0}^{t_1^*(\phi^{(1)}(x),z)-1} \beta^t\,\theta(\phi^{(1)}(x),z,t)\right] \\ &\quad -\beta\left[\sum_{t=0}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\,\theta(\phi^{(0)}(x),z,t)\right] \\ &\geq 1+\left[\sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\,\theta(x,z,t)\right] \\ &\quad -\beta\left[1+\sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\,\theta(\phi^{(0)}(x),z,t)\right] \\ &\geq (1-\beta)+\left[\sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\,\theta(x,z,t)-\beta\sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\,\theta(\phi^{(0)}(x),z,t)\right] \end{aligned} \tag{B.27}$$

Once more it follows from the definition of β^* detailed in the proof of Proposition 4.1 part c) that (B.27) will be strictly positive for all $\beta < \beta^*$. To see this it is enough to notice that by Proposition Lemma B.2, for any $x > \phi_\infty^{(0)}$ it holds that $x > \phi^{(0)}(x)$, and then by Proposition Lemma B.7 it holds that $\phi_t^1(x) > \phi_t^1(\phi^{(0)}(x))$ for any $t \geq 0$, which implies that $\theta(x,z,t) < \theta(\phi^{(0)}(x),z,t)$. Thus, for $w(x,z)$ to be strictly positive it should be the case that

$$(1-\beta) > \beta \sum_{t=1}^{t_1^*(\phi^{(0)}(x),z)} \beta^t\,\theta(\phi^{(0)}(x),z,t) - \sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\theta(x,z,t); \tag{B.28}$$

whenever $\beta\sum_{t=1}^{t_1^*(\phi^{(0)}(x),z)} \beta^t\,\theta(\phi^{(0)}(x),z,t) - \sum_{t=1}^{t_1^*(\phi^{(1)}(x),z)} \beta^t\theta(x,z,t) < 0$. Denote by $\beta^*_{(3)}$

the lowest discount factor β for which (B.28) does not hold. Then, it holds that for $x \in (z, 1]$, $w(x, z) > 0$ for $\beta < \beta_3^*$.
It is straightforward to see that $\beta_3^* > \beta^*$, given that for deriving β^* we must consider an infinite sum of terms, instead of a finite one, as in this case.
Thus, we conclude that, for $x \in (z, 1]$ when $z \in [\phi_\infty^{(0)}, 1]$, $w(x, z) > 0$ for $\beta < \beta^*$. ■

These are the results obtained which prove that, provided the MP index is continuous, a) $\frac{\partial w(x,x)}{\partial x} \leq 0$ and b) $\frac{\partial r(x,x)}{\partial x} \geq 0$.
These results are required to show Proposition 4.2, Proposition 4.4, and Proposition 4.4.
Next, we prove Lemma 6.1.
For $z < \phi_\infty^{(1)}$, to show $\frac{\partial w(x,x)}{\partial x} \leq 0$ and b) $\frac{\partial r(x,x)}{\partial x} \geq 0$, we start by writing $w(x, x)$ and $r(x, x)$ in closed form as follows:

$$\begin{aligned}
w(x,x) &= 1 + \beta\ [1-(1-\alpha)\,x\,]\ g(\phi^{(1)}(x), x) - \beta g(\phi^{(0)}(x), x) \\
&= 1 + \beta\ [1-(1-\alpha)\,x\,] \sum_{t=0}^{\infty} \beta^t\, \theta(\phi^{(1)}(x), x, t) - \beta \sum_{t=0}^{\infty} \beta^t\, \theta(\phi^{(0)}(x), x, t) \\
&= (1-\beta)\ + \sum_{t=1}^{\infty} \beta^t \left[\theta(x, x^-, t) - \beta\, \theta(\phi^{(0)}(x), x, t)\right] \\
&= \sum_{t=0}^{\infty} \beta^t \left[\theta(x, x^-, t) - \beta\, \theta(\phi^{(0)}(x), x, t)\right], \qquad \text{(B.29)}
\end{aligned}$$

where to compute $g(\phi^{(1)}(x), z)$ and $g(\phi^{(0)}(x), z)$ we have used the fact that $t_0^*(\phi^{(1)}(x), x) = t_0^*(\phi^{(0)}(x), x) = 1$ and with where x^- stands for the sensing policy with active set equal to $B(x^-) = [x, 1]$.

$$\begin{aligned}
r(x,x) &= R\,(1-\alpha)\,x + \beta\ [1-(1-\alpha)\,x\,]\ f(\phi^{(1)}(x), x) - \beta f(\phi^{(0)}(x), x) \\
&= R\,(1-\alpha) \Bigg[x + \beta\ [1-(1-\alpha)\,x\,] \sum_{t=0}^{\infty} \beta^t\, \phi_t^1(x)\, \theta(\phi^{(1)}(x), x, t) - \\
&\quad \beta \sum_{t=0}^{\infty} \beta^t\, \phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t) \Bigg] \\
&= R\,(1-\alpha) \left[\left(x - \beta\phi^{(0)}(x)\right) + \sum_{t=1}^{\infty} \beta^t \left[\phi_t^1(x)\, \theta(x, x^-, t) - \beta\phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t)\right] \right], \\
&= R\,(1-\alpha) \left[\sum_{t=0}^{\infty} \beta^t \left[\phi_t^1(x)\, \theta(x, x^-, t) - \beta\phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t)\right] \right], \qquad \text{(B.30)}
\end{aligned}$$

where we have again used the fact that $t_0^*(\phi^{(1)}(x), x) = t_0^*(\phi^{(0)}(x), x) = 1$.
Next, we take partial derivative with respect to x and we obtain:

$$\frac{\partial w(x,x)}{\partial x} = \sum_{t=1}^{\infty} \beta^t \left[\frac{\partial \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \theta(\phi^{(0)}(x), x, t)}{\partial x} \right]. \tag{B.31}$$

$$\frac{\partial r(x,x)}{\partial x} = R\,(1-\alpha) \sum_{t=0}^{\infty} \beta^t \left[\frac{\partial \phi_t^1(x)\, \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t)}{\partial x} \right]. \tag{B.32}$$

To simplify notation for $t \geq 0$ we let

$$u_t \triangleq \left[\frac{\partial \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \theta(\phi^{(0)}(x), x, t)}{\partial x} \right], \tag{B.33}$$

$$v_t \triangleq \left[\frac{\partial \phi_t^1(x)\, \theta(x, x^-, t)}{\partial x} - \beta \, \frac{\partial \phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t)}{\partial x} \right]. \tag{B.34}$$

Notice that $u_0 = 0$ and $v_0 = (1 - \beta\rho^{(0)})$.
For some $x > z$ with $z < \phi_\infty^{(1)}$ it holds that for all $t \geq 0$

$$\begin{aligned} \frac{\partial \theta(x,z,t)}{\partial x} &= \frac{\partial \left[(1-(1-\alpha)\, \phi_{t-1}^1(x))\, \theta(x, z, (t-1)) \right]}{\partial x} \\ &= -(1-\alpha) \frac{\partial \phi_{t-1}^1(x)}{\partial x}\, \theta(x, z, (t-1)) + \\ &\quad \left[1 - (1-\alpha)\, \phi_{t-1}^1(x) \right] \frac{\partial \left[\theta(x, z, (t-1)) \right]}{\partial x} \end{aligned} \tag{B.35}$$

$$\frac{\partial \phi_t^1(x) \theta(x,z,t)}{\partial x} = \frac{\partial \phi_t^1(x)}{\partial x} \theta(x,z,t) - \phi_t^1(x) \frac{\partial \theta(x,z,t)}{\partial x}. \tag{B.36}$$

Then, using (B.35) we compute u_t as follows:

$$\begin{aligned} u_t = &-(1-\alpha) \left[\frac{\partial \phi_{t-1}^1(x)}{\partial x}\, \theta(x, x^-, (t-1)) - \beta \, \frac{\partial \phi_{t-1}^1(\phi^{(0)}(x))}{\partial x}\, \theta(\phi^{(0)}(x), x, (t-1)) \right] \\ &+ \Bigg(\left[1 - (1-\alpha)\, \phi_{t-1}^1(x) \right] \frac{\partial \left[\theta(x, x^-, (t-1)) \right]}{\partial x} \\ &- \beta \left[1 - (1-\alpha)\, \phi_{t-1}^1(\phi^{(0)}(x)) \right] \frac{\partial \left[\theta(\phi^{(0)}(x) x, x^-, (t-1)) \right]}{\partial x} \Bigg) \end{aligned} \tag{B.37}$$

Notice that rearranging terms of the above expression and using (B.36) we conclude that:

$$u_t = u_{t-1} - (1-\alpha)v_{t-1} \tag{B.38}$$

Further, using the result below

$$\begin{aligned}
\phi_t^1(x)\,\theta(x,z,t) &= \phi^1(\phi_{t-1}^1(x))\left[1-(1-\alpha)\,\phi_{t-1}^1(x)\right]\,\theta(x,z,(t-1)) \\
\phi_t^1(x)\,\theta(x,z,t) &= \left[p^{(1)}(1-\phi_{t-1}^1(x)) + \alpha\,\phi_{t-1}^1(x)(1-q^{(1)})\right]\,\theta\left(x,z,(t-1)\right),
\end{aligned}$$

we conclude that:

$$\begin{aligned}
\frac{\partial\phi_t^1(x)\,\theta(x,z,t)}{\partial x} = &\left[(1-q^{(1)})\alpha - p^{(1)}\right]\left[\frac{\partial\phi_{t-1(x)}^1}{\partial x}\theta(x,z,(t-1)) + \frac{\partial\theta(x,z,(t-1))}{\partial x}\phi_{t-1}^1(x)\right] \\
&+ p^{(1)}\frac{\partial\theta(x,z,(t-1))}{\partial x},
\end{aligned}$$

From the above result and (B.36), we derive the following relation

$$v_t = \left[(1-q^{(1)})\alpha - p^{(1)}\right]v_{t-1} + p^{(1)}u_{t-1} \tag{B.39}$$

Now, we shall prove that $\frac{\partial w(x,x)}{\partial x} \leq 0$ and $\frac{\partial r(x,x)}{\partial x} \geq 0$ hold, by showing that

Lemma B.16.

$$(a)\ \sum_{t=1}^{\infty}\beta^t u_t \leq 0 \tag{B.40}$$

$$(b)\ \sum_{t=0}^{\infty}\beta^t v_t \geq 0, \tag{B.41}$$

Proof. First, multiply both sides of equation (B.38) by β^t and sum (from 1 to infinite) to obtain:

$$\begin{aligned}
\sum_{t=1}^{\infty}\beta^t u_t &= \sum_{t=1}^{\infty}\beta^t\left[u_{t-1} - (1-\alpha)v_{t-1}\right] \\
\sum_{t=1}^{\infty}\beta^t u_t &= \left(\beta u_0 + \beta\sum_{t=1}^{\infty}\beta^t u_t\right) - (1-\alpha)\left(\beta v_0 + \beta\sum_{t=1}^{\infty}\beta^t v_t\right) \\
\sum_{t=1}^{\infty}\beta^t u_t &= -\beta\frac{(1-\alpha)}{(1-\beta)}\sum_{t=0}^{\infty}\beta^t v_t
\end{aligned} \tag{B.42}$$

We can now proceed analogously with (B.39) to conclude that:

$$\begin{aligned}
\sum_{t=1}^{\infty} \beta^t v_t &= \sum_{t=1}^{\infty} \beta^t \left[\left[(1-q^{(1)})\alpha - p^{(1)} \right] v_{t-1} + p^{(1)} u_{t-1} \right] \\
\sum_{t=1}^{\infty} \beta^t v_t &= \left[(1-q^{(1)})\alpha - p^{(1)} \right] \left(\beta v_0 + \beta \sum_{t=1}^{\infty} \beta^t v_t \right) + p^{(1)} \left(\beta u_0 + \beta \sum_{t=1}^{\infty} \beta^t u_t \right) \\
\sum_{t=1}^{\infty} \beta^t v_t + v_0 - v_0 &= \left[(1-q^{(1)})\alpha - p^{(1)} \right] \beta \left(\sum_{t=0}^{\infty} \beta^t v_t \right) + p^{(1)} \beta \left(\sum_{t=1}^{\infty} \beta^t u_t \right) \\
\sum_{t=0}^{\infty} \beta^t v_t &= \frac{v_0 + p^{(1)} \beta \left(\sum_{t=1}^{\infty} \beta^t u_t \right)}{\left(1 - \beta \left[(1-q^{(1)})\alpha - p^{(1)} \right] \right)} \qquad \text{(B.43)}
\end{aligned}$$

Finally, we plug in result (B.42) in (B.43)

$$\left[\left(1 - \beta \left[(1-q^{(1)})\alpha - p^{(1)} \right] \right) + p^{(1)} \beta^2 \frac{(1-\alpha)}{(1-\beta)} \right] \sum_{t=0}^{\infty} \beta^t v_t = v_0 \qquad \text{(B.44)}$$

From (B.44) it follows that $\sum_{t=0}^{\infty} \beta^t v_t \geq 0$ which ensures that $\sum_{t=0}^{\infty} \beta^t u_t \leq 0$. □

Therefore, the MP index $\lambda^{MP}(x)$ for the set of states $x \in (0, \phi_{\infty}^{(1)})$ is:

$$\lambda^{MP}(x) = \frac{R\,(1-\alpha) \left[\sum_{t=0}^{\infty} \beta^t \left[\phi_t^1(x)\, \theta(x, x^-, t) - \beta \phi_t^1(\phi^{(0)}(x))\, \theta(\phi^{(0)}(x), x, t) \right] \right]}{\sum_{t=0}^{\infty} \beta^t \left[\theta(x, x^-, t) - \beta\, \theta(\phi^{(0)}(x), x, t) \right]}, \quad x \in (0, \phi_{\infty}^{(1)}) \qquad \text{(B.45)}$$

Which is an infinite sum of continuous functions of the information state x. It is further nondecreasing in x since we have shown that $r(x, x)$ is nondecreasing in x and $w(x, x)$ is nonincreasing in x.

Next, we continue to prove Lemma 6.3.

$$\begin{aligned}
w(x,x) &= 1+\beta\ [1-(1-\alpha)\,x\,]\ g(\phi^{(1)}(x),x)-\beta g(\phi^{(0)}(x),x) \\
&= 1+\beta\ [1-(1-\alpha)\,x\,]\sum_{t=0}^{\infty}\beta^t\,\theta(\phi^{(1)}(x),x,t)\,a_t(\phi^{(1)}(x),z) \\
&= -\beta\sum_{t=0}^{\infty}\beta^t\,\theta(\phi^{(0)}(x),x,t)\,a_t(\phi^{(0)}(x),z) \\
&= (1-\beta a_t(\phi^{(0)}(x),x))\ +\ \sum_{t=1}^{\infty}\beta^t\left[\,\theta(x,x^-,t)a_t(x,x^-)-\beta\,\theta(\phi^{(0)}(x),x,t)a_t(\phi^{(0)}(x),z)\right] \\
&= \sum_{t=0}^{\infty}\beta^t\left[\,\theta(x,x^-,t)a_t(x,x^-)-\beta\,\theta(\phi^{(0)}(x),x,t)a_t(\phi^{(0)}(x),z)\right], \qquad \text{(B.46)}
\end{aligned}$$

where where x^- stands for the sensing policy with active set equal to $B(x^-)=[x,1]$.

$$\begin{aligned}
r(x,x) &= R\,(1-\alpha)\,x+\beta\ [1-(1-\alpha)\,x\,]\ f(\phi^{(1)}(x),x)-\beta f(\phi^{(0)}(x),x) \\
&= R\,(1-\alpha)\left[x+\beta\ [1-(1-\alpha)\,x\,]\sum_{t=0}^{\infty}\beta^t\,\phi_t^1(x,z)\,\theta(\phi^{(1)}(x),x,t)a_t(\phi^{(1)}(x),z)-\right. \\
&\qquad \left.\beta\sum_{t=0}^{\infty}\beta^t\,\phi_t^1(\phi^{(0)}(x),z)\,\theta(\phi^{(0)}(x),x,t)a_t(\phi^{(0)}(x),z)\right] \\
&= R\,(1-\alpha)\left[\left(x-\beta\phi^{(0)}(x)a_t(\phi^{(0)}(x),x)\right)\ +\ \sum_{t=1}^{\infty}\beta^t\left[\phi_t^1(x)\,\theta(x,x^-,t)a_t(x,x^-)\right.\right. \\
&\qquad \left.\left.-\beta\phi_t^1(\phi^{(0)}(x))\,\theta(\phi^{(0)}(x),x,t)a_t(\phi^{(0)}(x),z)\right]\right], \\
&= R\,(1-\alpha)\left[\sum_{t=0}^{\infty}\beta^t\left[\phi_t^1(x)\,\theta(x,x^-,t)a_t(x,x^-)-\beta\phi_t^1(\phi^{(0)}(x))\,\theta(\phi^{(0)}(x),x,t)a_t(\phi^{(0)}(x),z)\right]\right], \\
&\qquad \text{(B.47)}
\end{aligned}$$

In this case, to show $\frac{\partial w(x,x)}{\partial x}\leq 0$ and b) $\frac{\partial r(x,x)}{\partial x}\geq 0$, we start by writing $w(x,x)$ and $r(x,x)$ in closed form using the Lemma B.11 and Lemma B.12 to write $a_t(x,x^-)$ and $a_t(\phi^{(0)}(x),x)$ in closed form. Next, notice that for all $x\in[\phi_\infty^{(1)},\phi_\infty^{(0)})$, given that $\phi^{(1)}(x)<x<\phi^{(0)}(x)$, it therefore holds that:

$$a_0(x,x^-)=1\,,a_1(x,x^-)=0;\qquad a_0(x,x)=0\,,a_0(\phi^{(0)}(x),x)=1 \qquad \text{(B.48)}$$

Next, consider for instance, cases (ii) in Lemma B.11 and (iii) a) in Lemma B.12. Using results (B.19) and (B.20) we can write $a_t(x,x^-)$ and $a_t(\phi^{(0)}(x),x)$ in closed form. Notice that case (ii) in Lemma B.11 corresponds to $P=1$ in case (iii) a) of Lemma B.12. Further, define $\hat{\theta}(x,z,t) \triangleq \prod_{s=0}^{t-1} 1-(1-\alpha)\phi_t^{P,1}(x)$, with $\phi_0^{P,1}(x) \triangleq x$, $\phi_t^{P,1}(x) \triangleq \phi^{P,1}\left(\phi_{t-1}^{P,1}(x)\right)$ and $\phi^{P,1}(x) \triangleq \phi_P^0\left(\phi^{(1)}(x)\right)$. Thus, (B.46) and (B.47) reduce to the expressions below:

$$\begin{aligned} w(x,x) &= \sum_{t=0}^{\infty} \beta^{(P+1)t}\left[\hat{\theta}(x,x^-,t) - \beta\,\hat{\theta}(\phi^{(0)}(x),x,t)\right], \qquad\qquad (\text{B.49}) \\ r(x,x) &= R\,(1-\alpha)\left[\sum_{t=0}^{\infty} \beta^{(P+1)t}\left[\phi_t^{P,1}(x)\,\theta(x,x^-,t) - \beta\phi_t^{P,1}(x)(\phi^{(0)}(x))\,\theta(\phi^{(0)}(x),x,t)\right]\right], \end{aligned}$$

Next, we take partial derivative with respect to x and we obtain:

$$\frac{\partial w(x,x)}{\partial x} = \sum_{t=1}^{\infty} \beta^{(P+1)t}\left[\frac{\partial\hat{\theta}(x,x^-,t)}{\partial x} - \beta\,\frac{\partial\hat{\theta}(\phi^{(0)}(x),x,t)}{\partial x}\right] \tag{B.50}$$

$$\frac{\partial r(x,x)}{\partial x} = R\,(1-\alpha)\sum_{t=0}^{\infty} \beta^{(P+1)t}\left[\frac{\partial\phi_t^{P,1}(x)\,\hat{\theta}(x,x^-,t)}{\partial x} - \beta\,\frac{\partial\phi_t^{P,1}((\phi^{(0)}(x))\,\hat{\theta}(\phi^{(0)}(x),x,t)}{\partial x}\right] \tag{B.51}$$

To simplify notation for $t \geq 0$ we let

$$\hat{u}_t \triangleq \left[\frac{\partial\hat{\theta}(x,x^-,t)}{\partial x} - \beta\,\frac{\partial\hat{\theta}(\phi^{(0)}(x),x,t)}{\partial x}\right] \tag{B.52}$$

$$\hat{v}_t \triangleq \left[\frac{\partial\phi_t^{P,1}(x)\,\hat{\theta}(x,x^-,t)}{\partial x} - \beta\,\frac{\partial\phi_t^{P,1}((\phi^{(0)}(x))\,\hat{\theta}(\phi^{(0)}(x),x,t)}{\partial x}\right] \tag{B.53}$$

Now, as we did for showing monotonicity in the threshold values of case I, we shall prove that $\frac{\partial w(x,x)}{\partial x} \leq 0$ and $\frac{\partial r(x,x)}{\partial x} \geq 0$ hold, by showing that

Lemma B.17.

$$(a)\ \sum_{t=1}^{\infty} \beta^{(P+1)t}\hat{u}_t \leq 0 \tag{B.54}$$

$$(b)\ \sum_{t=0}^{\infty} \beta^{(P+1)t}\hat{v}_t \geq 0, \tag{B.55}$$

Proof. After algebraical manipulations analogous to the ones deployed in case I (i.e. (B.36)), and rearranging terms of the resulting expressions, it follows that:

$$\hat{u}_t = \hat{u}_{t-1} - (1-\alpha)\hat{v}_{t-1} \tag{B.56}$$

and

$$\hat{v}_t = A\hat{v}_{t-1} + B\hat{u}_{t-1} \tag{B.57}$$

where $A \triangleq \alpha(\rho^{(0)})^p\rho^{(1)} - (1-\alpha)B$ and $B \triangleq \left[\phi_\infty^{(0)}(1-(\rho^{(0)})^P) + (\rho^{(0)})^P p^{(1)}\right]$. Notice that for the number of passive slots equal to 1, i.e., $P = 1$, we recover case in which the cycle is to alternate one passive and one active slot.
Analogous results to (B.42) and (B.44) are derived in this case to be:

$$\sum_{t=1}^{\infty} \beta^{(P+1)t}\hat{u}_t = -\beta^{(P+1)}\frac{(1-\alpha)}{(1-\beta^{(P+1)t})}\sum_{t=0}^{\infty}\beta^{(P+1)t}\hat{v}_t, \tag{B.58}$$

and

$$\left[\left(1-\beta^{(P+1)}A\right) + B\beta^{2(P+1)}\frac{(1-\alpha)}{(1-\beta^{P+1})}\right]\sum_{t=0}^{\infty}\beta^{P+1}\hat{v}_t = v_0 \tag{B.59}$$

From (B.59), and given that $v_0 = (1-\beta\rho^{(0)})$ for all x in case II, it follows that $\sum_{t=0}^{\infty}\beta^{P+1}\hat{v}_t \geq 0$ which ensures that $\sum_{t=0}^{\infty}\beta^{P+1}\hat{u}_t \leq 0$. □

Therefore, the MP index $\lambda^{MP}(x)$ in the set of states $x \in (\phi_\infty^{(1)}, \phi_\infty^{(0)}]$ for cases (ii) in result B.11 and (iii) a) in B.12 is:

$$\lambda^{MP}(x) = \frac{R\,(1-\alpha)\left[\sum_{t=0}^{\infty}\beta^{(P+1)t}\left[\phi_t^{P,1}(x)(x)\,\theta(x,x^-,t) - \beta\phi_t^{P,1}(x)(\phi^{(0)}(x))\,\theta(\phi^{(0)}(x),x,t)\right]\right]}{\sum_{t=0}^{\infty}\beta^{(P+1)t}\left[\,\hat{\theta}(x,x^-,t) - \beta\,\hat{\theta}(\phi^{(0)}(x),x,t)\right]}, \tag{B.60}$$

Which is nondecreasing in x since we have shown that $r(x,x)$ is nondecreasing in x and $w(x,x)$ is nonincreasing in x.

Next, we consider the case (i) a) in Lemma B.12. Using result (B.20) we can write $a_t(x,x^-)$ and $a_t(\phi^{(0)}(x),x)$ in closed form. Notice again that case (ii) in result Lemma B.11 corresponds to $A = 1$ in case (iii) a) of Lemma B.12. Further, define

$$\tilde{\theta}(x,z,t) \triangleq \prod_{r=0}^{t-1}\left(\prod_{s=0}^{A-1} 1-(1-\alpha)\phi_t^{1,A}(x)\right) = \prod_{r=0}^{t-1}\theta(\phi_t^{1,A}(x),z,A),$$

with $\phi_0^{1,A}(x) \triangleq x$, $\phi_t^{1,A}(x) \triangleq \phi^{1,A}\left(\phi_{t-1}^{1,A}(x)\right)$ and $\phi^{1,A}(x) \triangleq \phi^0\left(\phi_A^1(x)\right)$. Thus, (B.46) and

(B.47) reduce to the expressions below:

$$\begin{aligned} w(x,x) &= \sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\Big[\theta(\phi_t^{1,A}(x),x^-,s)\,\tilde{\theta}(x,x^-,t) \\ &\quad -\beta\theta(\phi_t^{1,A}(\phi^{(0)}(x),x,s)\,\tilde{\theta}(\phi^{(0)}(x),x,t)\Big], \qquad \text{(B.61)} \end{aligned}$$

$$\begin{aligned} r(x,x) &= R\,(1-\alpha)\Bigg[\sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\phi_s^1(\phi_t^{1,A}(x))\,\theta(\phi_t^{1,A}(x),x^-,s)\,\tilde{\theta}(x,x^-,t) - \\ &\quad \beta\,\phi_s^1(\phi_t^{1,A}(\phi^{(0)}(x))\,\theta(\phi_t^{1,A}(\phi^{(0)}(x),x,s)\,\tilde{\theta}(\phi^{(0)}(x),x,t)\Bigg], \end{aligned}$$

Define further, $\check{\theta}(x,z,t,s) \triangleq \theta(\phi_t^{1,A}(x),z,s)\,\tilde{\theta}(x,z,t)$. Next, we take partial derivative with respect to x and we obtain:

$$\frac{\partial w(x,x)}{\partial x} = \sum_{t=1}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\left[\frac{\partial\check{\theta}(x,x^-,t,s)}{\partial x} - \beta\,\frac{\partial\check{\theta}(\phi^{(0)}(x),x,t,s)}{\partial x}\right] \qquad \text{(B.62)}$$

$$\frac{\partial r(x,x)}{\partial x} = R\,(1-\alpha)\sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\left[\frac{\partial\phi_s^1(\phi_t^{1,A}(x)\,\check{\theta}(x,x^-,t)}{\partial x} - \beta\,\frac{\partial\phi_s^1(\phi_t^{1,A}(\phi^{(0)}(x))\,\check{\theta}(\phi^{(0)}(x),x,t)}{\partial x}\right] \qquad \text{(B.63)}$$

To simplify notation for $t \geq 0$ and $s = 0,1,\ldots,A-1$ we let

$$\check{u}_{t,s} \triangleq \left[\frac{\partial\check{\theta}(x,x^-,t,s)}{\partial x} - \beta\,\frac{\partial\check{\theta}(\phi^{(0)}(x),x,t,s)}{\partial x}\right] \qquad \text{(B.64)}$$

$$\check{v}_{t,s} \triangleq \left[\frac{\partial\phi_s^1(\phi_t^{1,A}(x))\,\check{\theta}(x,x^-,t,s)}{\partial x} - \beta\,\frac{\partial\phi_s^1(\phi_t^{1,A}(\phi^{(0)}(x)))\,\check{\theta}(\phi^{(0)}(x),x,t,s)}{\partial x}\right] \qquad \text{(B.65)}$$

Now, as we did for case I, we shall prove that $\frac{\partial w(x,x)}{\partial x} \leq 0$ and $\frac{\partial r(x,x)}{\partial x} \geq 0$ hold, by showing that

Lemma B.18.

$$(a)\ \sum_{t=1}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\check{u}_{t,s} \leq 0 \qquad \text{(B.66)}$$

$$(b)\ \sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\check{v}_{t,s} \geq 0, \qquad \text{(B.67)}$$

Proof. After algebraical manipulations analogous to the ones deployed in case I (i.e. (B.36)), and rearranging terms of the resulting expressions, it follows that for all $t \geq 0$ and $s \in \{0, \dots (A-1)\}$:

$$\check{u}_{t,s} = \check{u}_{t,s-1} - (1-\alpha)\check{v}_{t,s-1} \tag{B.68}$$

and

$$\check{v}_{t,s} = \left[(1-q^{(1)})\alpha - p^{(1)}\right] \check{v}_{t,s-1} + p^{(1)}\check{u}_{t,s-1} \tag{B.69}$$

Now it follows from (B.16) that for all $t \geq 0$ it holds that:

$$(a) \sum_{s=0}^{A-1} \beta^s \check{u}_{t,s} \leq 0 \tag{B.70}$$

$$(b) \sum_{s=0}^{A-1} \beta^s \check{v}_{t,s} \geq 0, \tag{B.71}$$

Finally, by (B.70) (a)-(b), we conclude that (B.70) holds. □

Notice that for $A = 1$ we recover case I, in which the cycle is to alternate one passive and one active slot.
Therefore, the MP index $\lambda^{MP}(x)$ in the set of states $x \in (\phi_\infty^{(1)}, \phi_\infty^{(0)}]$ for cases (ii) in result Lemma B.11 and (i) a) in Lemma B.12 is:

$$\lambda^*(x) = \frac{R\,(1-\alpha)\left[\sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\left[\phi_s^1(\phi_t^{1,A}(x))\,\check{\theta}(x,x^-,t) - \beta\phi_s^1(\phi_t^{1,A}(\phi^{(0)}(x)))\,\check{\theta}(\phi^{(0)}(x),x,t)\right]\right]}{\sum_{t=0}^{\infty}\sum_{s=0}^{A-1} \beta^{(A+1)t+s}\left[\check{\theta}(x,x^-,t) - \beta\,\check{\theta}(\phi^{(0)}(x),x,t)\right]}, \tag{B.72}$$

Which is nondecreasing in x since we have shown that $r(x,x)$ is nondecreasing in x and $w(x,x)$ is nonincreasing in x.

It remains to prove that cases in result B.11 and (i) b) and (iii) b) (corresponding to the *irregular* cycles) will also have a monotone index function in x. In both cases, the marginal work and reward measures can be decomposed into two infinite sums (depending on which part of the irregular cycle the period is) and the belief state process can be expressed in terms of Möbiuos transformations (as it was done for this two cases). Finally, with these expressions it follows by an analogous argument to the one applied for regular cycles that $\frac{\partial w(x,x)}{\partial x} \leq 0$ and $\frac{\partial r(x,x)}{\partial x} \geq 0$.

Appendix C

Appendix to Chapter 6

C.1 Work-Reward Measures Analysis

In order to prove Proposition Proposition 6.1 in Chapter 6, we invoked a Lemma providing a lower bound on the marginal work measures $w(x, z)$. In this Appendix we shall outline how to derive that bound in detail. Further, we outline the proof of the lemmas required to ensure the monotonicity of the resulting index.

In the multi-target tracking model presented in Chapter 6 we have considered two iterated mappings of the form $s \mapsto \phi^{(a)}(s)$ where s denotes the initial Scaled Tracking Error Variance (STEV) and $a = 0, 1$ stands for passive and active actions respectively. Letting $\phi_0^{(a)}(s) \triangleq s$ and $\phi_t^{(a)}(s) \triangleq \phi^{(a)}(\phi_{t-1}^{(a)}(s))$ for $t \geq 1$, and defining:

$$\phi^{(0)}(s) = s + \theta \tag{C.1}$$

$$\phi^{(1)}(s) = \frac{s + \theta}{s + \theta + 1} \tag{C.2}$$

where $\theta = \frac{q}{r}$ stands for the *position to measurement noise variance ratio*.

For the sake of establishing PCL-indexability, we are interested in studying the behavior of the t-th iterate of both mappings. In order to do we visualize both dynamics as Möbius Transformations or Linear Fractional Transformations (LFTs), with associated matrix representations given by:

$$\Phi^0 = \begin{pmatrix} 1 & \theta \\ 0 & 1 \end{pmatrix} \qquad \Phi^1 = \begin{pmatrix} 1 & \theta \\ 1 & (1+\theta) \end{pmatrix}$$

Note that for equation (C.1), the corresponding LFT is a pure translation (since in this case $c = 0$ and $a = d$) and thus, both fixed points are at infinity.

Lemma C.1. *For $a = 0, 1$ and any $x < \phi_\infty^{(a)}$, $0 \leq t < \infty$, $\phi_t^{(a)}(x) < \phi_\infty^{(a)}$ whereas for any*

$x > \phi_\infty^{(a)}$, $\phi_t^{(a)}(x) > \phi_\infty^{(a)}$.

Proof. It follows straightforward from the result (A.8) in the appendix on Möbius transformations. □

Lemma C.2. *For any $s \in (0, z]$,*

$$t_0^*(\phi^{(1)}(s), z) - t_0^*(\phi^{(0)}(s), z) \in \{0, 1\}$$

Proof. From the definition of $t_0^*(s, z)$ provided in Chapter 6 it follows that such function will be a *floor* function since for $s \leq z$, $t_0^*(s, z)$ satisfies:

$$\phi^{(0)}_{t_0^*(s,z)-1}(s) \leq z < \phi^{(0)}_{t_0^*(s,z)}(s).$$

Thus, using the property below:

$$\lfloor x \rfloor - \lfloor y \rfloor \leq \lfloor x - y \rfloor,$$

it can be shown that the maximum value for $t_0^*(\phi^{(1)}(s), z) - t_0^*(\phi^{(0)}(s), z)$ is 1. □

$\langle\langle$ *Notice that for any $s \leq z$ the critical iteration of the passive dynamics, denoted by $t_0^*(s, z)$, is an integer such that:*

$$\frac{z-s}{\theta} < t_0^*(s, z) \leq \left(\frac{z-s}{\theta}\right) + 1,$$

which leads us to conclude that:

$$t_0^*(s, z) = \left\lfloor \frac{z-s}{\theta} \right\rfloor + 1$$

Similarly, the critical iteration of the active recursion, denoted by $t_1^*(s, z)$, is an integer such that:

$$\frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} \leq t_1^*(s, z) < \frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} + 1,$$

which again leads us to conclude that:

$$t_1^*(s, z) = \left\lceil \frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} \right\rceil$$

$\rangle\rangle$

Below, we provide the proof of lemma Lemma 6.1.

Proof. For all $s < \phi_{\infty}^{(1)}$, it holds that:

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \right] > 0 \tag{C.3}$$

We start by realizing that: that:

$$\operatorname{sgn} \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \right],$$

which can be computed as follows

$$\begin{aligned} \operatorname{sgn} \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} \frac{\partial (\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \frac{\partial (\phi^{(1)}(s))}{\partial s} \right] &= \\ \operatorname{sgn} \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \frac{1}{(s+\theta+1)^2} \right] & \end{aligned} \tag{C.4}$$

It follows from Proposition A.5 of the Appendix A, that

$$\operatorname{sgn} \left[\frac{\partial \phi_t^{(1)}(\phi^{(0)}(s))}{\partial s} - \frac{\partial \phi_t^{(1)}(\phi^{(1)}(s))}{\partial s} \right] > 0 \quad \forall s \in \mathbb{S},$$

since for $k > 0$ the sign of the derivative of the LFT with respect to its argument s is non negative, thus it is increasing in its argument, and it holds that $\phi^{(0)} > \phi^{(1)}$and $\frac{1}{(s+\theta+1)^2} < 1$.

$\langle\langle$ *To compute the values of Proposition A.5, it is useful to consider that*

$$\gamma_{1,2} = \frac{1}{2} \left(-\theta \mp \sqrt{\theta(4+\theta)} \right) \tag{C.5}$$

and The eigenvalues of matrix Φ^1 *are the following:*

$$\lambda_{1,2} = \frac{(2+\theta) \mp \sqrt{\theta(4+\theta)}}{2} \tag{C.6}$$

$$\tag{C.7}$$

Thus, $k \geq 0$ *since it is defined as the ratio of the eigenvalues.* $\rangle\rangle$

□

For Case II in the single target tracking model, we shall deploy results which are analogous to the ones deployed in the single target hunt model. Below, we summarize them. These results describe the STEV state cycles under a z-threshold policy in results Lemma C.3, Lemma C.4 and Lemma C.5.

Lemma C.3. *For* $z \in [\phi^{(1)}_\infty, \phi^{(0)}_\infty)$ *and* $s \in \mathbb{S}$, *the* hitting time *of the STEV state process to the interval* $(\phi^{(1)}(z), \phi^{(0)}(z)]$ *is* finite *and once the belief state reaches this set of states, the probability of abandoning it is zero. The subset* $(\phi^{(1)}(z), \phi^{(0)}(z)]$ *is "absorbing".*

Proof. If $s \leq z$ after $t_0^*(s,z)$ slots under deterministic dynamics, we reach the set $(z, \phi^{(0)}(z)]$; while if $s > z$ and after $t_1^*(s,z) + t_0^*(\phi^{(1)}_{t_1^*(s,z)}(s), z)$ periods the set $(z, \phi^{(0)}(z)]$ is reached. Notice that the maximum value of s_t to reach the active set $B(z)$ coming from the passive set $B(z)^c$ is $\phi^{(0)}(z)$; then the minimum value of s_t to reach the passive set $B(z)^c$ coming from $(z, \phi^{(0)}(z)]$ is $\lim_{x \to z^+} \phi^{(1)}(s) = \phi^{(1)}(z)$. Thus, once s_t is in $(z, \phi^{(0)}(z)]$ we know that the interval $(\phi^{(1)}(z), \phi^{(0)}(z)]$ is never abandoned, alternating infinitely within it between the interval $(\phi^{(1)}(z), z] \subset B(z)^c$ (passive slots), and the interval $(z, \phi^{(0)}(z)] \subset B(z)$ (active slots), until the target is found.

□

Furthermore, as stated in Lemma C.4, within that "*absorbing*" set of states the *possible composition of cycles (in terms of the concrete sequence of active/passive time slots) is reduced to three cases: case 1: 1 passive slot & A active slots, with* $A \geq 21$; *case 2: 1 passive slot & 1 active slot; case 3:* P *passive slots &* 1 *active slot1, with* $P \geq 2$.

Lemma C.4. *(a) For* $z \in (\phi^{(1)}_\infty, \phi^0(\phi^{(1)}_\infty))$: *if* $s \in (\phi^{(1)}(z), z]$, *then* $t_0^*(s,z) = 1$; *If* $s \in (z, \phi^{(0)}(z)]$, $t_1^*(s,z) > 1$.

(b) For $z \in (\phi^1(\phi^{(0)}_\infty), \phi^0(\phi^{(1)}_\infty))$: *if* $s \in (\phi^{(1)}(z), z]$ *then* $t_0^*(s,z) = 1$. *If* $s \in (z, \phi^{(0)}(z)]$, *then* $t_1^*(s,z) = 1$.

(c) For $z \in (\phi^0(\phi^{(1)}_\infty), \phi^{(0)}_\infty)$,: *if* $s \in (\phi^{(1)}(z), z]$, *then* $t_0^*(s,z) > 1$. *If* $x \in (z, \phi^{(0)}(z)]$, *then* $t_1^*(s,z) = 1$.

Notice that the existence of belief state cycles under a z-threshold policy different from the ones described above is ruled out.

Lemma C.5. *Furthermore, for cases (i) and (iii) it may also occur that:*

(a) For $z \in (\phi^{(1)}_{\infty}, \phi^{0}(\phi^{(1)}_{\infty}))$:, $a.1)\ \forall s \in (z, \phi^{(0)}(z)]$ *then* $t^*_1(s,z) = A \geq 2$

$a.2)\ \forall s \in (z, s^*]$ *then* $t^*_1(s,z) = A \geq 2$ *, and*

$a.3) s \in (s^* \phi^{(0)}(z)]$ *then* $t^*_1(s,z) = A + 1$

Further in cases a.2) and a.3) it holds that $\phi^{(0)}(\phi^{(1)}_{A}(s)) \in (x^* \phi^{(0)}(z)]$ *and* $\phi^{(0)}(\phi^{(1)}_{A+1}(s)) \in (z, s^*]$

(b) For $z \in (\phi^{(0)}(\phi^{(1)}_{\infty}), \phi^{(0)}_{\infty})$:, $b.1)\ \forall s \in (\phi^{(1)}(z), z],$ *then* $t^*_0(s,z) = P \geq 2$

$b.2)\ \forall s \in (s^*, z], t^*_0(s,z) = P \geq 2$ *, and*

$b.3) \forall s \in (\phi^{(1)}(z), s^*]:$ *then* $t^*_0(s,z) = P + 1$

Further in cases b.2) and b.3) it holds that $\phi^{(1)}(\phi^{(0)}_{P}(s)) \in (\phi^{(1)}(z), s^*]$ *and* $\phi^{(1)}(\phi^{(0)}_{P+1}(s)) \in (s^*, z]$

Proof. Using these results, it can be shown that for $s \in \mathbb{S}$ such that $s > z$ and $z \in (\phi^{(1)}_{\infty}, \phi^{(0)}_{\infty})$:

$$\begin{aligned}
w(s,z) &= \big(g^c(s,z) - g^c(\phi^0(s),z)\big) + \Big(\beta^{T^*_1} g(s^1,z) - \beta^{T^*_0} g(s^0,z)\Big) \\
&\geq \left[\left(\frac{1-\beta^{t^*_1(s,z)}}{1-\beta}\right)(1-\beta) - \beta^{t^*_1(s,z)+1}\right] + \Big[\beta^{t^*_1(s,z)}(1-\beta^2) g^*(s^*,z)\Big] \\
&\geq 1 - \beta^{t^*_1(s,z)} \\
&\geq 1 - \beta \qquad \blacksquare
\end{aligned} \tag{C.8}$$

where $g^c(.,z)$ are cumulated work measure until the process s_t reaches the absorbing set of states, and $g^*(s^*,z)$ is the total work measure once the absorbing set of sates is reached. Further, for $z \in (\phi^{(1)}_{\infty}, \infty)$, it holds that $g^*(s^*,z) \geq \frac{1}{1-\beta^2}$. It holds that, for $s \leq z$ and $z \in (\phi^{(1)}_{\infty}, \phi^{(0)}_{\infty})$:

Using these results, it can also be shown that for $s \in \mathbb{S}$ such that $s \leq z$ and $z \in (\phi^{(1)}_{\infty}, \phi^{(0)}_{\infty})$:

$$\begin{aligned}
w(s,z) &= 1 + \beta g(\phi^{(1)}(s,z)) - g(s,z) \\
&= 1 + \beta(\beta^{t^*_0(\phi^{(1)}(s),z)} g(s^*,z) - \beta^{t^*_0(s,z)} g(s^*,z) \\
&\geq 1 - (1-\beta^2)\beta^{t^*_0(s,z)} g(s^*,z) \\
&\geq 1 - \beta^{t^*_0(s,z)} \\
&\geq 1 - \beta \qquad \blacksquare
\end{aligned} \tag{C.9}$$

□

Finally, to prove lemma Lemma 6.3 by which, for all $s \geq \phi_\infty^{(1)}$, it holds that:

$$\sum_{t=1}^{\infty} \beta^t \left[\frac{\partial\, a_t(\phi^{(1)}(s), s)}{\partial s} - \beta \frac{\partial a_t(\phi^{(0)}(s), s)}{\partial s} \right] \leq 0 \tag{C.10}$$

$$\sum_{t=0}^{\infty} \beta^t \left[\frac{\partial(\phi a)_t(\phi^{(0)}(s), s)}{\partial s} - \frac{\partial(\phi a)_t(\phi^{(1)}(s), s)}{\partial s} \right] > 0 \tag{C.11}$$

Proof. Regarding the (C.10), such a result follows from the fact that the marginal work measure $w(s, s)$ can be shown to be equal to a constant value representing the discounted fraction of time the tracking system will be active during the infinite horizon under the s-threshold value. Specifically, it can be shown that for any $s \in (\phi_\infty^{(1)}, \phi_\infty^{(0)})$

$$w(s,s) = \frac{1-\beta}{1-\beta^{C(s)}} = \left[\sum_{t=0}^{t=C(s)-1} \beta^t \right]^{-1},$$

where $C(S)$ is the length of the period cycle (or orbit) that threshold s generates. According to results C.3, C.4 and C.5, $C(s) = A + P$, A/P is the number of active/passive periods. Hence, for $A = P = 1$ the $w(s, s) = \frac{1-\beta}{1-\beta^2} = \frac{1}{1+\beta}$. In general,

$$w(s,s) = \frac{\sum_{t=0}^{t=A-1} \beta^t}{\sum_{t=0}^{t=A+P-1} \beta^t}.$$

Further, $C(s)$ is decreasing in s (as it varies between $1/C(s) = 1$, when $s \searrow \phi_\infty^{(1)}$ and $1/C(s) = 1/2$ as $s \nearrow \infty$). Thus,

$$\frac{\partial w(s,s)}{\partial s} \leq 0$$

Regarding the (C.11), it follows from an analogous reasoning to the one deployed in the proof of Lemma 6.1. Under all possible cycle compositions (i.e. in terms of active and passive slots), the resulting $(\phi a)_t(s)$ processes are again LFT whose derivative with respect to the argument is again positive. Using this fact, and knowing that $\phi^{(0)} > \phi^{(1)}$and $\frac{1}{(s+\theta+1)^2} < 1$, (C.11) can be shown for every possible cycle in C.3, C.4 and C.5.

□