UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 10-15
Statistics and Econometrics Series 06
April 2010

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

MULTITARGET TRACKING VIA RESTLESS BANDIT MARGINAL PRODUCTIVITY INDICES AND KALMAN FILTER IN DISCRETE TIME

José Niño-Mora and Sofía S. Villar

Abstract

This paper designs, evaluates, and tests a tractable priority-index policy for scheduling target updates in a discrete-time multitarget tracking model, which aims to be close to optimal relative to a discounted or average performance objective accounting for tracking-error variance and measurement costs. The policy is to be used by a sensor system composed of M phased-array radars coordinated to track the positions of N targets moving according to independent scalar Gauss-Markov linear dynamics, which therefore allows for the use of the Kalman Filter for track estimation. The paper exploits the natural problem formulation as a multiarmed restless bandit problem (MARBP) with real-state projects subject to deterministic dynamics by deploying Whittle's (1988) index policy for the MARBP. The challenging issues of indexability (existence of the index) and index evaluation are resolved by applying a method recently introduced by the first author for the analysis of real-state restless bandits. Computational results are reported demonstrating the tractability of index evaluation, the substantial performance gains that the Whittle's marginal productivity (MP) index policy achieves against myopic policies advocated in previous work and the resulting index policies suboptimality gaps. Further, a preliminary small scale computational study shows that the MP index policy exhibits a nearly optimal behavior as the number of distinct objective targets grows with the number of radars per target constant.

Keywords: multitarget tracking; sensor management; phased array radar; radar scheduling; scaled track-error variance (STEV); Kalman Filter; index policy, marginal productivity (MP) index, real-state multiarmed restless bandit problem (MARBP)

Sofía S. Villar and José Niño-Mora are with the Department of Statistics, Carlos III University of Madrid, 28911 Leganés (Madrid), Spain. Email: svillar@est-con.uc3m.es (Sofía S. Villar) and jnino@est-econ.uc3m.es (José Niño-Mora). **Acknowledgements:** Work funded by the Spanish Ministry of Science and Innovation project MTM2007-63140 and an I3 faculty endowment grant, and by the Autonomous Community of Madrid grant CCG08-UC3M1ESP-4162

Multitarget Tracking via Restless Bandit Marginal Productivity Indices and Kalman Filter in Discrete Time

José Niño-Mora and Sofía S. Villar*†

April 30, 2010

Abstract

This paper designs, evaluates, and tests a tractable priority-index policy for scheduling target updates in a discrete-time multitarget tracking model, which aims to be close to optimal relative to a discounted or average performance objective accounting for tracking-error variance and measurement costs. The policy is to be used by a sensor system composed of M phased-array radars coordinated to track the positions of N targets moving according to independent scalar Gauss–Markov linear dynamics, which therefore allows for the use of the Kalman filter for track estimation. The paper exploits the natural problem formulation as a multiarmed restless bandit problem (MARBP) with real-state projects subject to deterministic dynamics by deploying Whittle's (1988) index policy for the MARBP. The challenging issues of indexability (existence of the index) and index evaluation are resolved by applying a method recently introduced by the first author for the analysis of real-state restless bandits. Computational results are reported demonstrating the tractability of index evaluation, the substantial performance gains that the Whittle's marginal productivity (MP) index policy achieves against myopic policies advocated in previous work and the resulting index policies suboptimality gaps. Further, a preliminary small–scale computational study shows that the (MP) index policy exhibits a nearly optimal behavior as the number of distinct objective targets grows with the number of radars per target constant.

KEY WORDS: multitarget tracking; sensor management; phased array radar; radar scheduling; scaled track-error variance (STEV); Kalman Filter; index policy; marginal productivity (MP) index; real-state multiarmed restless bandit problems (MARBP)

*Work funded by the Spanish Ministry of Science and Innovation project MTM2007–63140 and an I3 faculty endowment grant, and by the Autonomous Community of Madrid grant CCG08–UC3M1ESP–4162

†J. Niño-Mora and Sofía S. Villar are with the Department of Statistics, Carlos III University of Madrid, 28911 Leganés (Madrid), Spain. Email: jnino@est-econ.uc3m.es, svillar@est-econ.uc3m.es

Contents

List of Figures

1 Introduction

1.1 Background and Motivation

Recent advances in sensor technology have provided modern multi-sensor systems with an increased operating flexibility to achieve given performance objectives. Such an unprecedented flexibility provides new systems with the possibility of rapidly adapting its functioning to suit a variety of highly dynamic environments. Yet, fully exploiting this benefit calls for the development of appropriate *scheduling algorithms*. The widespread adoption of these cutting-edge technologies has stimulated this demand and ultimately matured into an emerging field of research: ***sensor management*** (*SM*).

A concrete example of SM 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 schedulling of transmission parameters so as to optimize the system's performance. We refer the reader to [16] for a survey of the substantial literature in the area.

The design of such scheduling policies should take into consideration the distinctive features of each transmission parameter as well as the *dynamics* and *uncertainty* which characterizes the environment for the utilization of shared, and hence usually scarce, system's resources. This fact accounts for the growing surge of research on modern radar scheduling that seek to optimize a concrete system's objective, such as target tracking or target detection, by formulating stochastic dynamic programming models known as *Markov Decision Problems* (MDP). Unfortunately, optimal scheduling strategies for these problems are often computational intractable in all but a few simple problems. Thus, the design of *tractable* and *near-optimal* SM policies represents a considerable research challenge.

For the inherent benefits of these flexible systems to be fully realized, the following issues have to play a prominent role in the design of SM policies: (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 result in a poor performance in another environment) and; (iv) since idle radar time can be allocated to other tasks in multi-function radars, policy design should take into account that a low system utilization may become highly advantageous.

This paper addresses the *SM* problem faced by the coordinator of a set of M *phased-array* radars whose objective is the dynamic *tracking* over an indeterminate time horizon of a fixed number N of well-separated moving targets. *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.

1.2 Prior Work: MDP Models for Multitarget Tracking

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., [15], [14], [3]. 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 [5], a beam scheduling algorithm is derived from a discrete-time and discrete-state *partially observed Markov decision process* (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 *multiarmed bandit problem* (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 and target track measurements follow scalar, linear Gauss–Markov dynamics, and target *track-error variances* (TEVs) are updated via Kalman filter's equations is introduced in [4]. The authors seek to to optimize the sum of the targets' track error variances over a finite horizon and propose a *greedy* 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 *greedy-index* policy to be optimal in for the case of two symmetric targets.

A promising approach to the design of policies for dynamic prioritization of target track updates, as well as for other related SM problems, draws on the formulation of *multiarmed restless bandit problems* (MARBPs) with real-state projects. The MARBP is a powerful modeling framework which concerns the optimal sequential allocation of a scarce resource to a collection of N stochastic projects, out of which at most M can be engaged at a given time. Each project (or *bandit*) is modeled as a binary-action (active or passive) *Markov decision process* (MDP). The goal is to find a scheduling policy that maximizes the expected total discounted (ETD) or the long-run expected time-average reward earned over an infinite horizon. In the special classic case where only one project can be engaged at each period and passive projects do not change state, there exists an optimal policy of index type, the *Gittins index* [1] is attached to each project as a function of its state, and then a project of largest index is engaged at each time.

Yet, bandit formulations of SM problems call for the use of MARBP models, since projects (where a "project" is a target and the active action is to deploy sensing resources to the target) change state when passive. Indeed, Whittle used multitarget tracking as one of his motivating applications for introducing the MARBP, in the example of M aircraft trying to track the positions of N enemy submarines, where, as he put it, "the bandits are restless in the most literal sense." This inadequacy of the classic model is also pointed out in [6], where the authors extend the results in [4] 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. Despite remarking that such a problem falls within the framework of the MARBP presented by Whittle in [18], they do not use the indexation approach proposed there.

Index policies are generally suboptimal for the MARBP, yet Whittle introduced in [18] a heuristic index policy based on a particular index for restless bandits, which emerges from a Lagrangian relaxation and decomposition approach that also yields a bound on the optimal problem value. The *Whittle index*, which has been extended in [10] into the more general concept of *marginal productivity* (MP) *index* —named after its economic interpretation— raises substantial research challenges as (i) *indexability* (i.e., existence of the index) needs to be established for the model at hand; and (ii) the index needs to be evaluated in a tractable fashion.

Over the last decade, the first author has developed a methodology for resolving such issues on *restless bandit indexation* in the discrete-state case, in a stream of work starting in [7–9], which is reviewed in [11]. More recently he has announced in [12] extensions to real-state restless bandits, which are the cornerstone of the present paper's approach, and are also deployed in an opportunistic spectrum access model in [13]. The potential of real-state MARBP models to effectively address SM problems resides at the possibility of resolving the previously mentioned prominent issues on SM scheduling policy design by deploying such an *indexation methodology*.

1.3 Goals and Contributions

The model we consider here is based on and extends that formulated in [4], in which targets and target track measurements follow scalar, linear Gauss–Markov dynamics, and target *track-error variances* (TEVs) are updated via Kalman filter's equations. This paper extends such a line of work by investigating an MARBP formulation of dynamic tracking of multiple asymmetric targets with scalar linear Gauss–Markov dynamics, which incorporates both tracking-error and measurement (energy) costs, the main goal being to obtain a tractable index policy that performs well based on restless bandit indexation. The paper deploys the methodology for real-state restless bandit indexation announced in [12] to establish indexability and evaluate the MP index in an efficient fashion for the model of concern. The resulting beam scheduling rule is both non-myopic and depends on the target's initial TEV and on its movement and measurement dynamics. Computational results obtained demonstrate the tractability of index evaluation, the substantial performance gains that the marginal productivity (MP) Whittle's index policy achieves against myopic policies advocated in previous work as well as the resulting index policies suboptimality gaps. Moreover, preliminary computational results suggest that the resulting index policy is nearly optimal for the case in which the total number of *distinct* targets grows as the proportion of radars to targets remains constant. Proofs will be included in the full journal version of this work, along with extensive large-scale computational experiments.

1.4 Organization of the Paper

The remainder of the paper is organized as follows. Section 2 describes the multitarget tracking model and the MARBP formulation. Section 3 discusses the restless bandit indexation methodology for real-state restless bandits introduced in [12] as it applies to the design of index policies for multitarget tracking. Section 4 discusses how to deploy such a methodology in the present model to verify indexability and to provide a tractable evaluation of the Whittle's MP index. Alternative index policies and their relation to the Whittle's MP index policy are summarized in Section 5. Section 6 reports the results of some small-scale computational studies to assess the tractability and the computational cost of index evaluation as well as the relative and absolute performance of the Whittle's MP index policy.

2 Multitarget Tracking and Restless Bandit Formulation

2.1 Multitarget Tracking Kalman Filter Model

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 labeled by $m \in \mathsf{M} \triangleq \{1, \ldots, M\}$. 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 [4] and [6], 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 $x_t^{(n)}$ 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* $y_t^{(n)}$ is obtained. Decisions on which target tracks to update at each time are formulated by binary *action* processes $a_t^{(n)} \in \{0, 1\}$, where $a_t^{(n)} = 1$ if target n is measured in slot t and $a_t^{(n)} = 0$ otherwise.

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

$$x_t^{(n)} = F^{(n)} x_{t-1}^{(n)} + \omega_t^{(n)}, \quad t \geq 1, \tag{1}$$

where the *position-noise process* $\omega_t^{(n)}$ 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 $y_t^{(n)}$ is generated by the following linear Gauss–Markov dynamics

$$y_t^{(n)} = H^{(n)} x_t^{(n)} + \nu_t^{(n)}, \tag{2}$$

which is target specific but independent of the radar being used, and where the *measurement-noise process* $\nu_t^{(n)}$ 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 $\widehat{x}_0^{(n)}$ and $p_0^{(n)}$, respectively, are given for each target n, then the optimal minimum-variance predicted estimates are given by the Kalman filter. The TEV $p_t^{(n)}$, which describes the uncertainty in target n's track at the beginning of slot t, is recursively updated by the Kalman equations

$$p_t^{(n)} = \begin{cases} p_{t-1}^{(n)} + q^{(n)}, & \text{if } a_t^{(n)} = 0 \\[2ex] \dfrac{p_{t-1}^{(n)} + q^{(n)}}{p_{t-1}^{(n)}/r^{(n)} + q^{(n)}/r^{(n)} + 1}, & \text{if } a_t^{(n)} = 1 \end{cases}$$

We shall take the *state* of each target n to be its *scaled TEV* (STEV) $s_t^{(n)} \triangleq p_t^{(n)}/r^{(n)}$, which follows the dynamics

$$s_t^{(n)} = \begin{cases} \phi^{0,(n)}\left(s_{t-1}^{(n)}\right), & \text{if } a_t^{(n)} = 0 \\[2ex] \phi^{1,(n)}\left(s_{t-1}^{(n)}\right), & \text{if } a_t^{(n)} = 1, \end{cases}$$

where

$$\phi^{0,(n)}(s) \triangleq \theta^{(n)} + s, \quad \phi^{1,(n)}(s) \triangleq \frac{\theta^{(n)} + s}{1 + \theta^{(n)} + s} \tag{3}$$

and $\theta^{(n)} \triangleq q^{(n)}/r^{(n)}$ is the *position to measurement noise variance ratio* for target n.

Such a STEV state, being a scaled variability measure of target's n current position estimate, naturally moves over the *state space* $\mathsf{S}^{(n)} \triangleq [0, \infty)$. Hence, for some $r^{(n)} \in (0, \infty)$, $s_0^{(n)} = 0$ corresponds to exact knowledge of the targets' initial positions and $s_0^{(n)} = \infty$ to complete uncertainty of the targets' initial positions.

Note that, for any initial state $s_0^{(n)} = s$, the t-th iterate of $\phi^{1,(n)}(s)$, $\phi_t^{1,(n)}(s)$, which is generated by $\phi_0^{1,(n)}(s) \triangleq s$ and $\phi_t^{1,(n)}(s) \triangleq \phi^{1,(n)}\Big(\phi_{t-1}^{1,(n)}(s)\Big)$, converges to the limit

$$\phi_\infty^{1,(n)} \triangleq \lim_{t\to\infty} \phi_t^{1,(n)}(s) = \frac{1}{2}\left(\sqrt{\theta^{(n)}(4+\theta^{(n)})} - \theta^{(n)}\right),$$

which is the unique nonnegative root of $\phi^{1,(n)}(s) = s$ and is an attractive fixed point.

Also, notice that, for any initial state $s_0^{(n)} = s$, the tth iterate of $\phi^{0,(n)}(s)$, $\phi_t^{0,(n)}(s)$, which is generated by $\phi_0^{0,(n)}(s) \triangleq s$ and $\phi_t^{0,(n)}(s) \triangleq \phi^{0,(n)}\Big(\phi_{t-1}^{0,(n)}(s)\Big)$, converges to the limit

$$\phi_\infty^{0,(n)} \triangleq \lim_{t\to\infty} \phi_t^{0,(n)}(s) = \Big(s + \theta^{(n)} t\Big) = \infty,$$

which is an attractive fixed point.

Notice that the subset of states $\mathsf{S}^{(n)} \triangleq [\phi_\infty^{1,(n)}, \infty)$ is *absorbing* for target n. Note further that $\phi_\infty^{1,(n)} \leq \theta^{(n)}$ iff $\theta^{(n)} \geq 1/2$, which will be the case if, for instance, radar's measurements on target n are precise enough, while $\phi_\infty^{1,(n)} \geq \theta^{(n)}$ iff $\theta^{(n)} \leq 1/2$. We assume henceforth that $\theta^{(n)} \geq 1/2$ for each target n.

As alleged by Whittle in [18] when describing the *submarine surveillance* example, the passive and active dynamics ($\phi_t^{0,(n)}(s)$ and $\phi_t^{1,(n)}(s)$) result in contrary movements in the state space $\mathsf{S} \triangleq [0, \infty)$, which respectively correspond to loss and gain of precision on targets' location estimates.

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

$$\sum_{n\in\mathsf{N}} a_t^{(n)} \leq M, \quad t = 0, 1, 2, \ldots \tag{4}$$

We assume that a radar which updates the target n's track in a time slot incurs a *measurement cost* $h^{(n)} \geq 0$, representing the cost of beam energy expended for the track's update. Further, we take the *tracking-error cost* at slot t to be $d^{(n)} p_{t+1}^{(n)} = d^{(n)} r^{(n)} s_{t+1}^{(n)}$, 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 STEV state s is $C^{(n)}(s, a) \triangleq d^{(n)} r^{(n)} \phi^{a,(n)}(s) + h^{(n)} a$.

2.2 Multiarmed Restless Bandit Formulation

Consider the following dynamic optimization problems: (1) find a discount-optimal policy,

$$V_D^*(\mathbf{s}) = \min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(M)} \mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \sum_{n\in\mathsf{N}} \beta^t C^{(n)}\Big(s_t^{(n)}, a_t^{(n)}\Big)\right], \tag{5}$$

which minimizes the *expected total discounted* (ETD) cost, where $0 < \beta < 1$ is the discount factor, $\mathbf{s}_0 = \mathbf{s} = (s^{(n)})$ is the initial joint STEV state, and $\mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}[\cdot]$ denotes expectation under policy $\boldsymbol{\pi}$ conditional on $\mathbf{s}_0 = \mathbf{s}$; and (2) find an average-optimal policy,

$$V_A^*(\mathbf{s}) = \min_{\boldsymbol{\pi} \in \boldsymbol{\Pi}(M)} \overline{\lim_{T \to \infty}} \frac{1}{T} \mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{T} \sum_{n \in \mathsf{N}} C^{(n)}\left(s_t^{(n)}, a_t^{(n)}\right) \right], \tag{6}$$

which minimizes the expected long-run average cost.

Problems (5) and (6) are discrete-time multiarmed restless bandit problems 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 $s_t^{(n)}$ lives on the Borel state space $\mathsf{S}^{(n)}$. Note that, taking action $a_t^{(n)}$ on target n, with $a_t^{(n)} = 1$: a beam is steered toward target n to measure its position; $a_t^{(n)} = 0$: 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 $C^{(n)}\left(s_t^{(n)}, a_t^{(n)}\right)$ per PRI, which describes the tracking accuracy for a given resource consumption $a_t^{(n)}$; and (ii) the target's next state $s_{t+1}^{(n)}$ is given by (3), which implies that, given $a_t^{(n)}$, state transitions are deterministic and independent across projects.

The existence of an optimal solution for MARBP such as (5) is ensured under appropriate conditions on $C^{(n)}$ and $a_t^{(n)}$,(cf. [2]). 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 equations* (DPEs). Nonetheless, exact numerical solution to such DPEs is generally intractable, due both to the curse of dimensionality and to problem specific difficulties introduced by its continuous state space. This computational infeasibility is also the case for the average-cost MARBP (6).

In view of the above, 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. Such policies attach an index $\lambda^{(n)}(s^{(n)})$ to each target n as a function of its STEV state $s^{(n)}$, depending on target parameters. At time t, the resulting index policy selects at most M targets to measure, using $\lambda^{(n)}(s_t^{(n)})$ as a priority index for measuring target n (where a larger index value means a higher priority), among those targets, if any, for which the index exceeds the measurement cost, i.e., $\lambda^{(n)}\left(s_t^{(n)}\right) > h^{(n)}$, breaking ties arbitrarily.

In the sequel, we shall focus for concreteness on discounted-cost problem (5), although our approach also applies to average-cost problem (6).

3 Real-state Restless Bandit Indexation

3.1 Relaxed Problem, Lagrangian Relaxation, and Decomposition

Along the lines introduced in [18] for the equality-constrained case, we shall deploy a Lagrangian relaxation and decomposition approach. We thus start by *relaxing* problem (5) replacing the sample-path *peak resource-usage* constraint (4) that at most M targets are measured at each time by the averaged version of such a requirement that the ETD number of measured targets does not exceed $M/(1-\beta)$, i.e.,

$$\mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty}\sum_{n\in\mathsf{N}}\beta^t a_t^{(n)}\right] \leq \frac{M}{1-\beta}. \tag{7}$$

Denoting by $\Pi(\infty)$ the class of nonanticipative scheduling policies (which can measure any number of targets at any time), the *relaxed primal problem* is

$$V^{\mathrm{R}}(\mathbf{s}) = \min_{(7),\boldsymbol{\pi}\in\boldsymbol{\Pi}(\infty)} \mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty}\sum_{n\in\mathsf{N}}\beta^t C^{(n)}\left(s_t^{(n)}, a^{(n)}\right)\right]. \tag{8}$$

Note that the optimal value (cost) of (8) $V^{\mathrm{R}}(\mathbf{s})$ gives a *lower bound* on the optimal value of (5) $V_D^*(\mathbf{s})$.

To address such a constrained MDP (8) we deploy a Lagrangian approach, including coupling constraint (7) and attaching a multiplier $\lambda \geq 0$ to it. The resulting problem

$$V^{\mathrm{L}}(\mathbf{s};\lambda) = \min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(\infty)} \mathsf{E}_{\mathbf{s}}^{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty}\sum_{n\in\mathsf{N}}\beta^t\left\{C^{(n)}\left(s_t^{(n)}, a_t^{(n)}\right) + \lambda a_t^{(n)}\right\}\right] - \frac{M\lambda}{1-\beta} \tag{9}$$

is a *Lagrangian relaxation* of (8), whose optimal value $V^{\mathrm{L}}(\mathbf{s};\lambda)$ gives a lower bound on $V^{\mathrm{R}}(\mathbf{s})$. Next, given the fact that target's state transitions are independent, we *decompose* problem (9) as

$$V^{\mathrm{L}}(\mathbf{s};\lambda) = \sum_{n\in\mathsf{N}} V_{(n)}^{\mathrm{L}}(s^{(n)};\lambda) - \frac{M\lambda}{(1-\beta)}, \tag{10}$$

where

$$V_{(n)}^{\mathrm{L}}(s^{(n)};\lambda) = \min_{\pi^{(n)}\in\Pi^{(n)}} \mathsf{E}_{s^{(n)}}^{\pi^{(n)}}\left[\sum_{t=0}^{\infty}\beta^t\left\{C^{(n)}\left(s_t^{(n)}, a_t^{(n)}\right) + \lambda a_t^{(n)}\right\}\right], \tag{11}$$

is target n's subproblem optimal value and $\Pi^{(n)}$ is the class of nonanticipative tracking policies for target n *in isolation*. In terms of these individual problems, multiplier λ represents an *additional cost*, to be added to the target's regular measurement cost $h^{(n)}$, that will be paid per time slot a beam is measuring target n. Note that, for a given *charge* λ, target n's subproblem (11) can be interpreted as the optimal control problem faced by a manager exclusively responsible of tracking target n. We will hence refer to (11) as target's n's λ-*charge subproblem.*

The *Lagrangian dual problem* is to find an optimal value $\lambda^*(\mathbf{s})$ of λ giving the best lower bound on $V^{\mathrm{R}}(\mathbf{s})$, which we denote by $V^{\mathrm{D}}(\mathbf{s})$:

$$V^{\mathrm{D}}(\mathbf{s}) = \max_{\lambda \geq 0} V^{\mathrm{L}}(\mathbf{s}; \lambda) \tag{12}$$

Such a problem can be interpreted in economic terms as the central coordinator's problem of selecting a critical λ^* value for the *charge* paid by each target manager so that by independently solving their individual λ-*charge subproblem* the best possible system performance $V^{\mathrm{L}}(\mathbf{s}, \lambda^*(\mathbf{s}))$ is achieved. Note that such a λ^* solves (12) which is a scalar convex optimization problem, since $V^{\mathrm{L}}(\mathbf{s}; \lambda)$ is concave in λ. Clearly, by decoupling the whole problem into n individual subproblems, (10) is significantly easier to solve than (5), yet its computational tractability depends on that of individual subproblems (11).

Notice that although *weak duality* ($V^{\mathrm{R}}(\mathbf{s}) \geq V^{\mathrm{D}}(\mathbf{s})$) is ensured, satisfaction of *strong duality*, i.e. $V^{\mathrm{R}}(\mathbf{s}) = V^{\mathrm{D}}(\mathbf{s})$, calls for further investigation.

3.2 Indexability; Whittle's Marginal Productivity Index Policy

Focus now on the optimal control problem faced by a dedicated manager concentrated exclusively on the tracking target n in isolation as described by subproblem (11). We shall henceforth treat measurement charge λ as a scalar parameter taking values in $\mathbb{R}$. Thus, negative values of λ can be viewed as a *subsidy for measuring target* n, just as positive values of λ were interpreted as an additional *measurement cost for measuring target* n. In light of this economic interpretation of multiplier λ, consider the definition of the following key structural property of such a parametric restless bandit subproblem, termed *indexability*.

Definition 1 We say that subproblem (11) is *indexable* if there exists an *index* $\lambda^{*,(n)}(s)$ which is a scalar function of the target's STEV state $s \in \mathsf{S}$ such that, for any value of multiplier (*measurement charge/subsidy*) $\lambda \in \mathbb{R}$, the active action $a_t^{(n)} = 1$ (measuring the target) is optimal in state $s_t^{(n)} = s$ iff $\lambda^{*,(n)}(s) \geq \lambda$.

If the above definition holds, the solution to an *indexable* restless bandit subproblem can be simplified, and hence also that of Lagrangian dual (12). Exploiting this special structure, a reduced class of admissible policies in $\Pi^{(n)}$ needs to be considered in order to solve it. Clearly, when the optimal policy of (11) can be expressed in terms of such a scalar function $\lambda^{*,(n)}(s)$ it suffices to consider *deterministic stationary policies* to find its solution.

Further, *indexable* subproblems result in a resource allocation rule which is in accordance with the traditional microeconomics profit maximization principle by which a resource should be exploited up the point in which the *marginal profit* of employing an extra unit of it equals zero. From definition 1, the beam is allocated to measure target

n in a PRI only as long as the measurement charge λ, its *marginal cost*, does not exceed function $\lambda^{*,(n)}(s)$, which can thus be viewed as the *marginal revenue* of measuring target n when it occupies STEV state s. Thus, optimal solution of indexable subproblems are such that the *marginal profit* of allocating a beam at each active PRI is at least 0, i.e. $\lambda^{*,(n)}(s) - \lambda \geq 0$.

The indexability property of restless bandits was introduced by Whittle in [18], being later extended by the author (cf. the survey [11]), leading to the unifying concept of *marginal productivity* (MP) *index* after its above mentioned natural economic interpretation.

If each single-target subproblem (11) were indexable and a tractable procedure were available to evaluate index $\lambda^{*,(n)}(s)$, then this would readily yield a computationally tractable algorithm to solve Lagrangian dual problem (12) —provided the objective of (11) could also be efficiently evaluated— and thus compute the lower bound $V^{\mathrm{D}}(\mathbf{s})$ referred to above. Further, we could then use for multitarget problem (5) the resulting Whittle's MP index policy, based on using $\lambda^{*,(n)}\left(s_t^{(n)}\right)$ as target n's priority index.

3.3 Sufficient Indexability Conditions and Index Evaluation

Whittle's indexability ensures that optimal policies for restless bandit problems can be characterized by a scalar priority index, yet this structural property needs to be established for the model at hand. For such a purpose, the first author introduced in work reviewed in [11] sufficient indexability conditions for discrete-state restless bandits based on satisfaction on *partial conservation laws* (PCLs), along with an index algorithm.

The first author has extended the scope of such conditions to real-state restless bandits in results announced in [12], as reviewed next. The ensuing discussion focuses on a single-project restless bandit problem modeling the optimal tracking of an individual target, whose label n is henceforth dropped from the notation. We thus write, e.g., the target's state and action processes as $s_t \in \mathsf{S} \triangleq [0, \infty)$ and $a_t \in \{0, 1\}$, respectively.

We shall evaluate the performance of an admissible tracking policy $\pi \in \Pi$ along two dimensions: the *work measure*

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

giving the ETD number of times the target is measured under policy π starting at $s_0 = s$; and the *cost measure*

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

giving the corresponding ETD cost incurred.

The target's optimal tracking problem (11) is then reexpressed in terms of these measures as

$$V^*(s; \lambda) = \min_{\pi \in \Pi} f(s, \pi) + \lambda g(s, \pi). \tag{13}$$

Once again, consider problem (13), which is a real-state MDP, as target's λ-*charge subproblem*.

In order to show indexability of (13), we shall study the conditions under which it suffices to consider *deterministic stationary policies*, which are naturally represented by their *active (state) sets*, i.e., the set of STEV states where they prescribe the active action (measure the target). For an active set $B \subseteq \mathsf{S}$, we shall refer to the B-*active policy*.

More precisely, 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 $B(z) \triangleq \{s \in \mathsf{S} \colon s > z\}$. Note that $B(z) = (z, \infty)$ for $s \geq 0$, $B(z) = \mathsf{S} = [0, \infty)$ 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

$$g(s, z) = \begin{cases} 1 + \beta g\big(\phi^1(s), z\big), & s > z \\ \beta g\big(\phi^0(s), z\big), & s \leq z, \end{cases} \tag{14}$$

whereas cost measure $f(s, z)$ is characterized by

$$f(s, z) = \begin{cases} C(s, 1) + \beta f\big(\phi^1(s), z\big), & s > z \\ C(s, 0) + \beta f\big(\phi^0(s), z\big), & s \leq z. \end{cases} \tag{15}$$

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*

$$w(s, z) \triangleq g(s, \langle 1, z \rangle) - g(s, \langle 0, z \rangle), \tag{16}$$

and the *marginal cost measure*

$$c(s, z) \triangleq f(s, \langle 0, z \rangle) - f(s, \langle 1, z \rangle). \tag{17}$$

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

$$\lambda(s, z) \triangleq \frac{c(s, z)}{w(s, z)}. \tag{18}$$

The following definition extends to the real-state setting a corresponding definition introduced by the first author in [7] for discrete-state restless bandits.

Definition 2 We say that subproblem (13) is *PCL-indexable* (with respect to threshold policies) if:

(i) *positive marginal work*: $w(s, z) > 0, s \in \mathsf{S}, z \in \overline{\mathbb{R}}$;

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

$$\lambda^*(s) \triangleq \lambda(s, s), \quad s \in \mathsf{S}. \tag{19}$$

is monotone nondecreasing in s

The next result, which extends the scope of a corresponding result in [7] for discrete-state restless bandits to the real-state setting, states the validity of the PCL-based sufficient indexability conditions deployed in this paper. It further shows how to evaluate the Whittle's MP index.

Theorem 1 *If subproblem* (13) *is PCL-indexable, then it is indexable and the* $\lambda^*(s)$ *in* (19) *is its Whittle's MP index.*

4 Indexability Analysis

This section reports the results of the analysis required to establish that the single-target tracking restless bandit model is PCL($\mathscr{F}$)-indexable, so that Theorem 1 can be invoked. We also describe how the MP Index can be computed in a tractable fashion.

The indexability analysis of the present model is based on the evaluation and analysis of work and cost measures $g(s,z)$ and $f(s,z)$, from which their marginal counterparts $w(s,z)$ and $c(s,z)$ are immediately obtained. Under a z-threshold policy, the state variable process s_t starting at some initial state $s_0 = s$ in S determines the evolution of the associated action process a_t. Thus, we next focus on the study of the s_t process for every possible threshold level $z \in \overline{\mathbb{R}}$ and every possible initial state $s \in \mathsf{S}$.

Consider the iterates $\phi_t(s,z)$ and $a_t(s,z)$, which are the STEV and action processes s_t and a_t generated under the z-threshold policy starting at s. They can be recursively computed as follows. Letting

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

where $1_B(s)$ is the indicator of set B and $B^c(z) \triangleq \mathsf{S} \setminus B(z)$, $\phi_0(s,z) \triangleq s$ and $\phi_t(s,z) \triangleq \phi\Big(\phi_{t-1}(s,z), z\Big)$ for $t \geq 1$. Further, $a_0(s,z) \triangleq 1_{B(z)}(s)$, and $a_t(s,z) \triangleq 1_{B(z)}\Big(\phi_t(s,z), z\Big)$ 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 dynamical systems: $(\mathbb{N}_0, \mathsf{S}, \phi)$ and $(\mathbb{N}_0, \{0,1\}, a)$. Such orbits determine 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 or they are closed and 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.

Based on properties of such discrete dynamical systems, evaluation measures can be studied so that sufficient indexability conditions in Definition 1 can be verified by algebraic means. This section outlines how to do so, and further shows how to use such properties to evaluate the index $\lambda^*(s)$.

In the sequel we focus for the sake of simplicity yet without loss of generality, on the case where the target's tracking cost is $h = 0$.

4.1 Case I: Threshold $z < 0$

In the case $z < 0$, henceforth denoted as z^-, under the z-threshold policy s_t starts and remains above threshold for every possible level of the STEV, since the active set includes every possible level of the STEV, i.e. $B(z^-) = \mathsf{S} = [0, \infty)$. Thus, there are no passive initial states to consider.

To obtain the total evaluation measures note that:

$$s_t = \phi_t(s, z) = \phi_t^1(s) \text{ and } a_t = 1 \quad t \geq 0,\ \forall s : s \in \mathsf{S}.$$

Hence, in this case both a_t and s_t converge to constant orbits whose fixed points are 1 and ϕ_∞^1 respectively.

Elementary arguments give that for any s in S and any $z < 0$ the total work and cost measure have the following evaluation:

$$\begin{aligned} g(s, z^-) &= \frac{1}{(1-\beta)} \\ f(s, z^-) &= d\,r \sum_{t=0}^{\infty} \phi_{t+1}^1(s)\beta^t \end{aligned}$$

Note that a closed form solution to $f(s, z^-)$ cannot be obtained, yet the infinite sum in $f(s, z^-)$ converges to a finite limit, since $\phi_t^1(s) \leq 1$ for any θ, $t \geq 0\,, s \in \mathsf{S}$. Thus, we have that $f(s, z^-) \leq \frac{d\,r}{(1-\beta)} \quad \forall s : s \in \mathsf{S}$.

Using the above total measures, we readily conclude that for any s in S:

$$w(s, z^-) = 1 \tag{20}$$

$$c(s, z^-) = d\,r \left[\frac{(s+\theta)^2}{(1+s+\theta)} + \beta \sum_{t=0}^{\infty} \beta^t \left(\phi_{t+1}^1(s+\theta) - \phi_{t+1}^1\left(\frac{s+\theta}{s+\theta+1} \right) \right) \right] \tag{21}$$

From this, it is readily obtained that for $s \to 0^-$ the index in (19) $\lambda^*(s)$ has the evaluation:

$$\lim_{s \to 0^-} \lambda^*(s) = d\,r \left[\frac{\theta^2}{(1+\theta)} + \beta \left(f\left(\theta, z^-\right) - f\left(\frac{\theta}{\theta+1}, z^- \right) \right) \right] \tag{22}$$

4.2 Case II: Threshold $z \in [0, \phi^1_\infty]$

In this case, under the z-threshold policy once s_t gets above the threshold z, it stays so thereafter, given that the passive set is a subset of the *non-absorbing* states in S. Also, for any $\theta > 0$ if $s \leq z$, s_t gets above threshold after a finite number of time slots. Let $t^*_0(s,z) \triangleq \{t \geq 1 : \phi^0_t(s) > z\}$. Note that for $\theta \geq 1/2$, it holds that $\phi^1_\infty \geq \frac{1}{2}$. It is easy to see that for $\theta > 1/2$ and any $s \in \mathsf{S} : s \leq z$, $t^*_0(s,z) = 1$ whereas if $\theta = 1/2$, $t^*_0(s,z) = 1$ for any $s \in \mathsf{S}/0 : s \leq z$ and $t^*_0(0,z) = 2$. In general we have that $t^*_0(s,z) < \infty$ as long as $\theta > 0$.

Then, to obtain the total evaluation measures note that for the case $\theta > 1/2$ or the case $\theta = 1/2$ and $z < 1/2$ it holds that:

$$s_t = \begin{cases} \phi_t(s,z) = \ \phi^0_1(s)\ ,\ \phi^1_t(s+\theta) \text{ for } t \geq 2 & \text{if } s \leq z \\ \\ \phi_t(s,z) = \ \phi^1_t(s) \text{ for } t \geq 1 & \text{if } s > z, \end{cases}$$

Elementary arguments give that for those cases the total work measure has the following evaluation. For any $s \in \mathsf{S}$, $g(s,z) = \frac{1}{(1-\beta)}$ if $s > z$, and $g(s,z) = \frac{\beta}{(1-\beta)}$ if $s \leq z$. An analogous argument can be applied for the case $\theta = z = 1/2$, to conclude that for $s \in \mathsf{S}$, $g(s,1/2) = \frac{1}{(1-\beta)}$ if $s > 1/2$, $g(s,1/2) = \frac{\beta}{(1-\beta)}$ if $0 < s \leq 1/2$ and $g(0,1/2) = \frac{\beta^2}{(1-\beta)}$.

Hence, also in the case $z \in [0, \phi^1_\infty]$ both a_t and s_t converge to constant orbits whose fixed points are 1 and ϕ^1_∞ respectively.

To compute the marginal work measure note that, if $s_0 > z$ we have that $w(s,z) = 1$, since $s_1 > z$ regardless of the selected action. On the other hand, if $s_0 \leq z$, under the threshold policy $s_1 > z$ except for $s = 0$, $z = \theta = 1/2$ in which case $s_1 = z$. Yet if $s_0 \leq z$ and the target is measured, for $\theta \geq 1/2$ it holds that $s_1 \geq 1/3$. This implies that it can either occur that $s_1 \leq z$ or $s_1 > z$, depending on the threshold value. Therefore, it is easy to see that $w(s,z) = 1 - \beta$ if $s \leq z$ and $s_1 \leq z$, while $w(s,z) = 1$ if either $s \leq z$ and $s_1 > z$. Note that if $\theta = z = 1/2$, $w(0,1/2) = 1$. Such results allow us to conclude that for any $s \in \mathsf{S}$, $z \in [0, \phi^1_\infty]$ and $1/2 \leq \theta < \infty$ it holds that $w(s,z)$ is either 1 or $(1-\beta)$. Following a similar argument, it can be concluded that for the more general case $\theta > 0$, $w(s,z)$ is either 1 or $(1 - \beta^{t^*_0(s,z)})$.

From these results it is readily concluded that $w(s,s) = 1$ for all $s \in [0, \phi^1_\infty)$ while $w(s,s) = 1 - \beta$ for $s = \phi^1_\infty$ and some $\theta > 0$.

It follows from the previously stated evolution of the process s_t that the total cost measure under a threshold policy will have the following evaluation:

$$f(s,z) \ = \begin{cases} d\,r\left[\Big(s+\theta\Big) + \beta\left(\sum_{t=0}^{\infty} \phi^1_{t+1}(s+\theta)\beta^t\right)\right] & s \leq z \\ \\ d\,r\left[\sum_{t=0}^{\infty} \phi^1_{t+1}(s)\beta^t\right] & s > z \end{cases}$$

Except for the case $\theta = z = 1/2$ and $s = 0$ in which it holds that

$$f(s,z) = \, d\, r\Bigg[\Big(\frac{1+\beta}{2}\Big) + \beta^2\Big(\sum_{t=0}^{\infty}\phi^1_{t+1}\Big(\frac{1+\beta}{2}\Big)\beta^t\Big)\Bigg].$$

Therefore, it is readily obtained that the index in (19) $\lambda^*(s)$ has the following evaluation

$$\lambda^*(s) \;=\; d\, r\Bigg[\frac{(s+\theta)^2}{(1+s+\theta)} + \beta\Bigg(f\Big(s+\theta, s\Big) - f\Big(\frac{s+\theta}{s+\theta+1}, s\Big)\Bigg)\Bigg],\; s \in [0, \phi^1_\infty) \quad (23)$$

$$\lambda^*(s) \;=\; \frac{d\, r}{1-\beta}\Bigg[\theta + \beta\Big(\frac{\phi^1_\infty + 2\theta}{\phi^1_\infty + 2\theta + 1} - (\phi^1_\infty + \theta)\Big) + \beta^2\Bigg(f\Big(\frac{\phi^1_\infty + 2\theta}{\phi^1_\infty + 2\theta + 1}, \phi^1_\infty\Big) - f\Big(\phi^1_\infty + \theta, \phi^1_\infty\Big)\Bigg)\Bigg],\; s = \phi^1_\infty \quad (24)$$

Notice that the resulting index evaluation $\lambda^*(s)$ for $s \in [0, \phi^1_\infty)$ includes that of case 4.1 as a special case.

4.3 Case III: Threshold $z \in \left(\phi^1_\infty, (\theta + \sqrt{2\theta + \theta^2})\right)$

In the case $z \in (\phi^1_\infty, \infty)$, it holds that the STEV process s_t under a z-threshold policy, starting at any $s \in \mathsf{S}$, hits the set $(z, z+\theta]$ after a finite number of time slots. Further, after first hitting such set the process s_t jumps infinitely often above and below threshold within the interval $(\phi^1_1(z), z+\theta]$, which is the absorbing subset of states in S under a z-threshold policy for any s , $z \in (\phi^1_\infty, \infty)$. Thus, for such z-threshold values, iterates of the process s_t become arbitrarily close to a periodic orbit (i.e. an asymptotically periodic orbit) while iterates of the a_t process settle into a periodic orbit.

The composition of such periodic orbits, in terms of the number active and passive PRI, clearly depends on the threshold level z. Measuring a target only if its STEV exceeds a level z requires a *low/high* radar activity level if the threshold z value is *high/low*. More precisely, measuring the target only if $s_t > z$ makes the system action process infinitely oscillate in periodic orbits of $c(z)$ time slots composed of $a(z)$ active PRI and $p(z)$ passive PRI.

It can be shown that if $z < \left(\theta + \sqrt{2\theta + \theta^2}\right)$ only one passive PRI is necessary to make s_t jump above threshold, whereas if $z \geq \left(\theta + \sqrt{2\theta + \theta^2}\right)$ only one active PRI is enough to make s_t fall to a level at most equal to threshold z. In terms of the above stated results, z-threshold such that $z \in \left(\phi^1_\infty, \theta + \sqrt{2\theta + \theta^2}\right)$ can be thought of as *low* values of the z-threshold while $z \in \left(\theta + \sqrt{2\theta + \theta^2}, \infty\right)$ as *high* values.

Let us now study the case in which $z \in (\phi^1_\infty, \left(\theta + \sqrt{2\theta + \theta^2}\right)$. In such a case it holds that, measuring the target only if s_t exceeds z requires the system action process to infinitely oscillate in periodic orbits of $c(z)$ time slots, in which the proportion of passive PRI with respect to the total number of PRI in the orbit period is less than $\frac{1}{2}$, i.e. $p(z)/c(z) < 1/2$. That is, for such *low* z-threshold values the system spends at least 50% of the time measuring the target.

For $s > z$, let $t_1^*(s,z) \triangleq \{t \geq 1 : \phi_t^1(s) \leq z\}$. It can be shown that for any $z > \phi_\infty^1$, $t_1^*\big((z+\theta), z\big) - t_1^*\big(z^+, z\big) \in \{0,1\}$ with $t_1^*\big(z^+, z\big)$ standing for $\lim_{s \to z+} t_1^*\big(s, z\big)$. In this case, the number of time slots in an orbit $c(z)$ is computed as the number of time slots in which the system returns to a STEV level in which $t_1^*(s,z) = t_1^*\big((z+\theta), z\big)$ starting from an initial state in which $t_1^*(s,z) = t_1^*\big(z+, z\big)$, for some $s \in (z, z+\theta]$. Note that the orbit duration depends on the z-threshold value but not on the initial STEV level. For the special case in which $t_1^*\big((z+\theta), z\big) - t_1^*\big(z^+, z\big) = 0$, the system oscillates in *regular* periodic orbits of $a(z) = t_1^*\big((z+\theta), z\big)$ active time slots followed by 1 passive time slot, hence the orbit's period length $c(z)$ is equal to $a(z)+1$. Yet if $t_1^*((z+\theta), z) - t_1^*((z^+), z)) = 1$ the periodic orbits have an *irregular* composition in terms of active and passive PRI.

For all $s \in \mathsf{S} : s \leq z$ and for some threshold $z \in \Big(\phi_\infty^1, (\theta + \sqrt{2\theta + \theta^2})\Big)$ it holds that $t_0^*(s,z) = 1$. Hence, elementary arguments result in the following evaluation for the total work and cost measure:

Proposition 1 *For $s \in \mathsf{S}$ and $z \in (\phi_\infty^1,\ (\theta + \sqrt{2\theta + \theta^2}))$,*

$$g(s,z) = \begin{cases} \frac{1-\beta^{t_1^*(s,z)}}{1-\beta} + \beta^{t_1^*(s,z)} \sum_{t=0}^{\infty} a_t\Big(\phi_{t_1^*(s,z)}^1(s), z\Big)\beta^t & s > z \\ \beta \sum_{t=0}^{\infty} a_t\Big(\ (s+\theta), z\Big)\beta^t & s \leq z, \end{cases}$$

$$f(s,z) = \begin{cases} \sum_{t=0}^{t_1^*(s,z)-1} \phi_{t+1}^1(s,z)\beta^t + \beta^{t_1^*(s,z)} \sum_{t=0}^{\infty} \phi_t\Big(\phi_{t_1^*(s,z)}^1(s), z\Big)\beta^t & s > z \\ (s+\theta) + \beta \sum_{t=0}^{\infty} \phi_t\Big(\ (s+\theta), z\Big)\beta^t & s \leq z \end{cases}$$

For a given z, Proposition 1 together with the stated properties of the a_t and s_t forward orbits result in the following evaluation of the marginal work measure $w(s,z)$ in (16):

$$w(s,z) \;=\; \begin{cases} (1-\beta)g(s,z) & s > z,\, d = 0 \\ 1 - \beta^{t_1^*(s,z)} \frac{(1-\beta^2)}{(1-\beta^{c(z)})} & s > z,\, d = 1 \\ 1 - \Big(\ (1-\beta)g(s,z)\ \Big) & s \leq z \end{cases}$$

with $d \triangleq t_1^*\big((s+\theta), z\big) - t_1^*\big(s, z\big)$, and $c(z) \geq 2$.

Note that $w(s,z) > 0$ for all $s \in \mathsf{S}, z \in \overline{\mathbb{R}}$. Also, from the previous analysis it can be readily obtained that for $s \in [(\theta + \sqrt{2\theta + \theta^2}), \infty)$ the marginal work measure in s has the following closed form evaluation $w(s,s) = \frac{1-\beta}{1-\beta^{c(s)}}$, where $c(z)$ is the number of time slots defining the periodic orbit of the process $a(t)$ under a z-threshold policy. Notice that $w(s,s) = \frac{1}{c(z)}$ for $\beta = 1$.

Further, arguing along the lines used for the marginal work measure $w(s,z)$, analogous formulae for the evaluation of $c(s,z)$ can be derived. From such formulae it is readily obtained the following evaluation of index in (19) $\lambda^*(s)$:

$$\begin{aligned} \lambda^*(s) \quad &= \quad \frac{d\,r(1-\beta^{c(z)})}{(1-\beta)}\left[\frac{(s+\theta)^2}{1+s+\theta} - \beta\Big(\frac{(s+\theta)}{1+s+\theta}+\theta\Big)\right] + \\ & \quad\; d\,r(1-\beta^{c(z)})\left[f\Big((s+\theta),s\Big) - f\Big(\frac{(s+\theta)}{(s+\theta+1)}+\theta,s\Big)\right],\ s \in \Big(\phi_\infty^1, (\theta+\sqrt{2\theta+\theta^2})\Big). \end{aligned} \tag{25}$$

4.4 Case IV: Threshold $z \in [(\theta + \sqrt{2\theta+\theta^2}), \infty)$

Following the argument invoked in the previous case, for $z \in [(\theta + \sqrt{2\theta+\theta^2}), \infty)$ the STEV process s_t under a z-threshold policy hits the set $(z, z+\theta]$ after a finite number of time slots thereafter jumping infinitely often above and below threshold within the interval $(\phi_1^1(z), z+\theta]$. As well, iterates of the process s_t become arbitrarily close to a periodic orbit (i.e. an asymptotically periodic orbit) while iterates of the a_t process settle into a periodic orbit. However, the composition of such periodic orbits differs from that of case 4.3 in terms of active and passive action periods.

Measuring a target only if its STEV exceeds a *high* z-threshold level requires the system action process to infinitely oscillate in periodic orbits of $c(z)$ time slots composed of $a(z)$ active PRI and $p(z)$ passive PRI. For $z \geq \Big(\theta + \sqrt{2\theta+\theta^2}\Big)$ it holds that $a(z) = 1$ for all z, i.e. only one active PRI is necessary to make s_t fall to a level at most equal to threshold z. Thus, it can be shown that the proportion of active PRI with respect to the total number of PRI in the orbit is at most equal to $\frac{1}{2}$, i.e. $a(z)/c(z) \leq 1/2$. That is, for such *high* z-threshold values the system spends at most 50% of the time measuring the target.

It can be shown that for any $z \geq (\theta+\sqrt{2\theta+\theta^2})$, $t_0^*\Big(\phi_1^1(z+\theta), z\Big) - t_0^*\Big(\phi_1^1(z^-), z\Big) \in \{0,1\}$ with $t_0^*\Big(\phi_1^1(z^-), z\Big)$ standing for $\lim_{s\to z_-} t_0^*\Big(\phi_1^1(s,z\Big)$. In this case, the number of time slots in an orbit $c(z)$ is also computed as the number of time slots in which the system returns to a STEV level in which $t_0^*(s,z) = t_0^*\Big(\phi_1^1(z+\theta), z\Big)$ starting from an initial state in which $t_0^*(s,z) = t_0^*\Big(\phi_1^1(z)-, z\Big)$, for some $s \in (z, z+\theta]$. Once more, orbit duration depends on the z-threshold value but not on the initial STEV level. For the special case in which $t_0^*\Big(\phi_1^1(z+\theta), z\Big) - t_0^*\Big(\phi_1^1(z^-), z\Big) = 0$, the system oscillates in *regular* periodic orbits of 1 active time slot followed by $p(z) = t_0^*\Big(\phi_1^1(z+\theta), z\Big)$ passive time slots, hence the orbit's period length $c(z)$ is equal to $1 + p(z)$. Yet if $t_0^*\Big(\phi_1^1(z+\theta), z\Big) - t_0^*\Big(\phi_1^1(z^-), z\Big) = 1$ the periodic orbits have an *irregular* composition in terms of active and passive PRI.

It can be shown that it holds that $t_1^*(s,z) = 1$. Hence, elementary arguments result in the following evaluation for the total work and cost measure:

Proposition 2 *For $s \in \mathsf{S}$ and $z \in [(\theta + \sqrt{2\theta + \theta^2}), \infty)$,*

$$g(s,z) = \begin{cases} 1 + \beta \sum_{t=0}^{\infty} a_t(\phi^1_{t_1^*(s,z)}(s), z)\beta^t & s > z \\ \beta^{t_0^*(s,z)} \sum_{t=0}^{\infty} a_t(\phi^0_{t_0^*(s,z)}(s), z)\beta^t & s \le z, \end{cases}$$

$$f(s,z) = \begin{cases} \frac{s+\theta}{s+\theta+1} + \beta \sum_{t=0}^{\infty} \phi_t(\phi^1_{t_1^*(s,z)}(s), z)\beta^t & s > z \\ \sum_{t=0}^{t_0^*(s,z)-1} \phi^0_{t+1}(s)\beta^t + \beta^{t_0^*(s,z)+1} \sum_{t=0}^{\infty} \phi_t(\phi^0_{t_0^*(s,z)}(s), z)\beta^t & s \le z \end{cases}$$

where $\sum_{t=0}^{t_0^*(s,z)-1} \phi^0_{t+1}(s)\beta^t$ admits the following closed form solution:

$$s\frac{(1-\beta^{t_0^*(s,z)})}{(1-\beta)} + \theta\frac{1-\beta^{t_0^*(s,z)+1}(2 + t_0^*(s,z)(1-\beta) - \beta)}{(1-\beta^2)}.$$

For a given z, Proposition 2 together with the stated properties of the a_t and s_t forward orbits, allow us to obtain the following evaluation of the marginal work measure $w(s,z)$ in (16):

$$w(s,z) = \begin{cases} (1-\beta)g(s,z) & s > z \\ 1 - \beta^{t_0^*(s,z)} \frac{(1-\beta^2)}{(1-\beta^{c(z)})} & s \le z,\, d_1 = 1 \\ 1 - (1-\beta)g(s,z) & s \le z,\, d_1 = 0 \end{cases}$$

with $d \triangleq t_0^*\Big(\frac{(s+\theta)}{s+\theta+1}, z\Big) - t_0^*\Big(s, z\Big)$ and $c(z) \ge 2$.

Note that $w(s,z) > 0$ for all $s \in \mathsf{S}, z \in \overline{\mathbb{R}}$. From the previous analysis, it is readily obtained that for $s \in [(\theta + \sqrt{2\theta + \theta^2}), \infty)$ the marginal work measure in s has the following closed form evaluation $w(s,s) = \frac{1-\beta}{1-\beta^{c(s)}}$. Note that this evaluation coincides with that of case 4.3, hence for $\beta = 1$, $w(s,s) = \frac{1}{c(z)}$.

Further, arguing along the lines used for the marginal work measure $w(s,z)$ evaluation, $c(s,z)$ can be derived via (17). From this it is readily obtained that the index $\lambda^*(s)$ in (19) has the evaluation

$$\lambda^*(s) = \frac{d\,r(1-\beta^{c(z)})}{1-\beta}\Bigg[\frac{(s+\theta)^2}{1+s+\theta} - \beta\sum_{t=0}^{t_0^*-1}\phi_{t+1}^0(s)\beta^t +$$
$$\beta(1-\beta^{t_0^*})\left(f\left((s+\theta), s\right) - f\left(s_{t_0^*}, s\right)\right)\Bigg],\ s \in [(\theta + \sqrt{2\theta+\theta^2}), \infty). \quad (26)$$

4.5 Case V: Threshold $z = \infty$

In the case $z = \infty$, under the z-threshold policy s_t will never be above threshold starting from any possible initial level of the STEV, given that the active set is the null set, i.e. $B(z) = \emptyset$. Hence, in this case every possible initial state is a passive initial state.

Once more, to obtain the total evaluation measures note that:

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

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 evaluation:

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

Hence, for any s in S the corresponding marginal measures are:

$$w(s,\infty) = 1 \quad (27)$$
$$c(s,\infty) = d\,r\left(\frac{(s+\theta)^2}{(1-\beta)(1+s+\theta)}\right) \quad (28)$$

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

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

Therefore the index (19) $\lambda^*(s)$ has the evaluation:

$$\lambda^*(s) = d\,r\left(\lim_{s\to\infty}\frac{(s+\theta)^2}{(1-\beta)(1+s+\theta)}\right) = \infty \quad (29)$$

4.6 Verification of PCL-indexability and Index Evaluation

Based on the results in Sections 4.1-4.5, conditions stated in 2 can be verified and we therefore obtain the following result.

Proposition 3 *The single-target tracking problem is PCL-indexable with respect to threshold policies under the β-discounted criterion, for $0 \leq \beta < 1$. Therefore, it is indexable, and the index $\lambda^*(s)$ previously calculated is its Whittle's MP index.*

We can also extend the result of Proposition 3 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 $\lambda^*(s)$ as $\beta \nearrow 1$.

We can thus evaluate work measure $g(s,z)$ and cost measure $f(s,z)$ by computing the infinite series

$$\begin{aligned} g(s,z) &= \sum_{t=0}^{\infty} \beta^t a_t(s,z) \\ f(s,z) &= d\, r \sum_{t=0}^{\infty} \beta^t \phi_{t+1}(s,z), \end{aligned} \tag{30}$$

truncating them to a finite number T of terms.

From these, we can readily compute the marginal work and cost measures $w(s,z)$ and $c(s,z)$ via (16)–(17). In turn, we can use the latter to obtain the index $\lambda^*(s)$ via (19). Alternatively, $\lambda^*(s)$ can be evaluated using $w(s,s)$ previously derived closed form formulae and an approximation of $c(s,s)$ based on the properties of s_t asymptotically periodic orbits.

5 The MP Index and Other Index Policies

5.1 The Myopic Index and the TEV 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 $s \in \mathsf{S}$ and some $\theta > 0$. Therefore, it is straightforward that both conditions in Definition 2 hold and thus, by Theorem 1, the target's optimal tracking problem for $\beta = 0$ is indexable and $\lambda^*(s) = \lambda^{\text{myopic}}(s)$ is its Whittle's MP index.

Such a myopic index policy is optimal in the multi-target model for $\beta = 0$, as it minimizes the total cost function, i.e. the sum of the N targets' tracking errors and energy expanded for the next PRI. Notice that, for $\beta = 0$ the optimal policy is such that for all $n \in \mathsf{N}$ we choose $a_t^{(n)}$ such that:

$$\min_{\pi \in \mathbf{\Pi}(M)} \Big\{ d^{(n)}\, r^{(n)} \phi_1^{a_t^{(n)}=0}(s^{(n)})\ ;\ d^{(n)}\, r^{(n)} \phi_1^{a_t^{(n)}=1}(s^{(n)}) \Big\}$$

The above stated condition is equivalent to choosing $a_t^{(n)}$ such that:

$$\max_{\pi \in \mathbf{\Pi}(M)} \Big\{ a_t^{(n)} d^{(n)}\, r^{(n)} \big[\phi_t(s,0) - \phi_t(s^{(n)},1)\big] \Big\} \Longleftrightarrow \max_{\pi \in \mathbf{\Pi}(M)} \Big\{ a_t^{(n)} \lambda^{\text{myopic}}(s^{(n)}) \Big\}$$

Further, in the completely symmetric case in which all targets $n \in \mathsf{N}$ 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 TEV $\lambda^{\text{TEV}}(s^{(n)}) = d^{(n)}\, r^{(n)} s^{(n)}$ result 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 *TEV* index. Thus, for $\beta = 0$ and in a completely symmetric scenario the MP index policy, the *TEV* index policy and the myopic index policy yield an identical tracking performance which is also optimal.

Both in [6] and [4] authors claim to optimize 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. Yet, notice that for the general case of asymmetric targets such a heuristic is not optimal nor does the above mentioned index policy equivalence hold.

5.2 The MP Index and the Gittins index: case $\theta = 0$

An interesting case to consider is when $\theta = 0$, under which active and passive dynamics are reduced 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:

$$g(s,z) = \begin{cases} \frac{1-\beta^{t_1^*(s,z)}}{1-\beta}, & s > z \\ 0, & s \leq z. \end{cases} \tag{31}$$

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

$$f(s,z) = \begin{cases} d\,r\Big[\sum_{t=0}^{t_1^*(s,z)-1} \frac{s}{s(t+1)+1}\beta^t + \frac{\beta^{t_1^*(s,z)}}{(1-\beta)} \frac{s}{s t_1^*(s,z)+1}\Big], & s > z \\ \\ d\,r\Big[\frac{s}{(1-\beta)}\Big], & s \leq z. \end{cases} \tag{32}$$

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 $s \leq z$. Further, $w(s,s) = 1$ while $c(s,s) = \frac{d\,r}{(1-\beta)}\Big[\phi^0(s) - \phi^1(s)\Big]$. Hence $\lambda(s,z) = c(s,z)$ and $\lambda^*(s) = d\,r s^2/(1+s)$. Since $(d/ds)\lambda^*(s) = \frac{d\,r}{(1-\beta)} s(2+s)/((1+s)^2 > 0$, the index $\lambda^*(s)$ is non decreasing for $s \in \mathsf{S}$ (and strictly increasing for $s \in \mathsf{S}\backslash 0$). Therefore, both conditions in Definition 2 hold and, by Theorem 1, the target's optimal tracking problem is indexable and $\lambda^*(s)$ is its Whittle's MP index. Moreover in this case, $\lambda^*(s)$ is also its Gittins index, since the model formulation under $\theta = 0$ is *classic*, and it can be conveniently expressed as: $\frac{\lambda^{\text{myopic}}(s)}{(1-\beta)}$.

Notice that the case $\theta = 0$ occurs either when the target's movement process is deterministic, i.e. $q = 0$, or when its measurement process is such that $r = \infty$. Also, note that in the latter case, its Whittle's MP index $\lambda^*(s) = \infty$ for all $s \in \mathsf{S}$ while in the former

$\lambda^*(s)$ depend only of r and s. Further, for the case of any N objective targets and $M = 1$ we can expect such an index to be optimal [1].

5.3 The MP Index: case $\theta = \infty$

It is also interesting to consider the model when $\theta = \infty$, under which active and passive dynamics are reduced to: $\phi_t^1(s) = \phi_\infty^1 = 1$ while $\phi_t^0(s) = \infty$ for all $s, z \in \mathsf{S}$. Hence, staring from any initial STEV in S, the process s_t under a threshold policy infinitely alternates between 1, the minimum STEV level, and ∞, the maximum STEV level, for all $t = 1, 2, ...$ Following the previous section argument, it can easily be seen that for all $z \in \mathsf{S}$:

$$g(s,z) = \begin{cases} \frac{1}{1-\beta^2}, & s > z \\ \frac{\beta}{1-\beta^2}, & s \leq z. \end{cases} \tag{33}$$

whereas cost measure $f(s, z)$ tends to infinite irrespective of the initial state and threshold value. Thus, it can be shown that in this case for any $s, z \in \mathsf{S}$ it holds that $w(s, z) = w(s, s) = \frac{1}{1+\beta}$. Further, $c(s, s)$ can be conveniently expressed as

$$c(s,s) = d\, r\Big[\Big(\phi^0(s) - \phi^1(s)\Big) + \beta\Big(f(\phi^0(s), s) - f(\phi^1(s), s)\Big)\Big].$$

Thus, $\lambda^*(s) = c(s, s)(1 + \beta)$. For this case it holds that $\lambda^*(s) = \lim_{\theta\to\infty} \lambda^{\text{myopic}}(s)(1 + \beta)$. From where it follows $\lambda^*(s)$ is increasing for $s \in \mathsf{S}$, therefore the target's optimal tracking problem is indexable with $\lambda^*(s)$ is its Whittle's MP index.

Notice that the case $\theta = \infty$ occurs either when the target's movement process is such that $q = \infty$ or when its measurement process is exact, i.e. $r = 0$. Also, note that in the former case, its Whittle's MP index $\lambda^*(s) = \infty$ for all $s \in \mathsf{S}$ while in the latter case $\lambda^*(s) = d\, q\, (1 + \beta)$ for all $s \in \mathsf{S}$.

6 Computational Experiments

6.1 MP Index Evaluation

We have implemented a MATLAB script for index evaluation using the above results. The MP index was then computed for a target instance with parameters $d = 1$, $r = 1$, and $q = 5$, so $\theta = 5$, $\phi_\infty^1 = 0.8541$ and $(\theta + \sqrt{2\theta + \theta^2}) = 0.9161$. The series in (30) 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 $\lambda^*(s)$ 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 periodic cycles that govern the evolution of the total work measures under a threshold policy. (See Appendix A.)

Fig. 1 shows the results. As required by the PCL-indexability conditions, in each case the index $\lambda^*(s)$ was monotone nondecreasing (in fact, strictly increasing) in s. Note that the index $\lambda^*(s)$ is continuous in s, being also piecewise differentiable. Further, for fixed s the index $\lambda^*(s)$ is increasing in β, converging as $\beta \nearrow 1$ to a limiting index that can be used for average-criterion problem (6), which we have approximated by taking $\beta = 0.9999$. For each s, the time to compute $\lambda^*(s)$ was negligible.

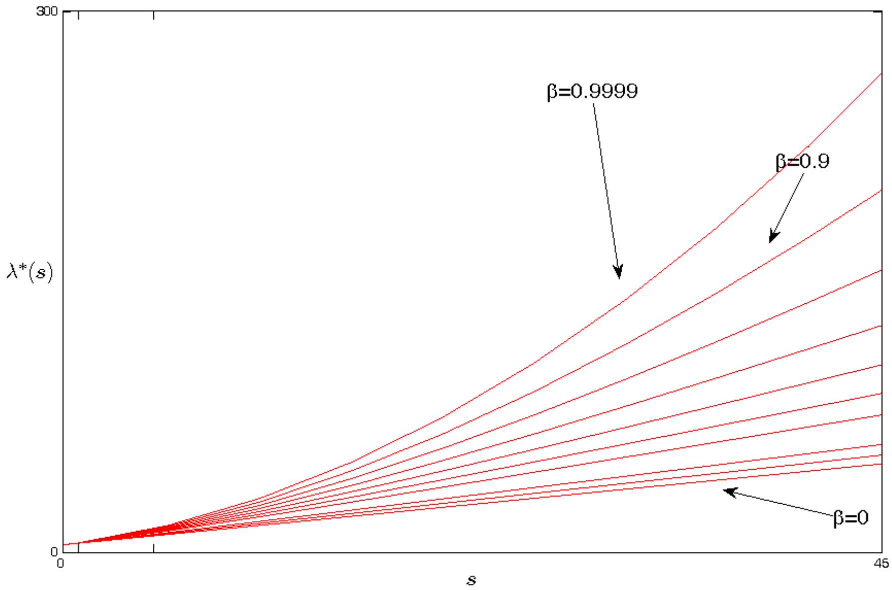

Figure 1: The Whittle's MP index for different discount factors β.

6.2 Numerical Convergence of the MP Index Evaluation

The convergence rate of the above implemented MP 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 MP index. Staring from a target instance with parameters as those of section 6.1, we implemented a script that computed the MP 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 $\beta \leq 0.9$ numerical convergence of such series is achieved at some $T_{max} \leq 10^2$. Thus, defining $\lambda_L^*(s)$ as $\lim_{T\to\infty} \lambda_T^*(s)$, 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 $\frac{e(T+1)}{e(T)}$.

Fig. 2 shows the results. The MP index approximate evaluation appears to converge linearly. Further, the convergence rate seems to be equal to the discount factor β. In fact, in appendix A we analytically derive such a result for the marginal work measure $w(s,s)$, for which a closed form expression is available. Extending the proof for the MP approximate index evaluation calls for further investigation.

Notice that under such conditions, the limiting index for average-criterion problem tends to converge sublinearly. Therefore, precise enough approximations for the case when $\beta \nearrow 1$ will result computationally more expensive 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 follow from the indexability analysis of section 4 and the study of the STEV dynamics under a threshold policy.

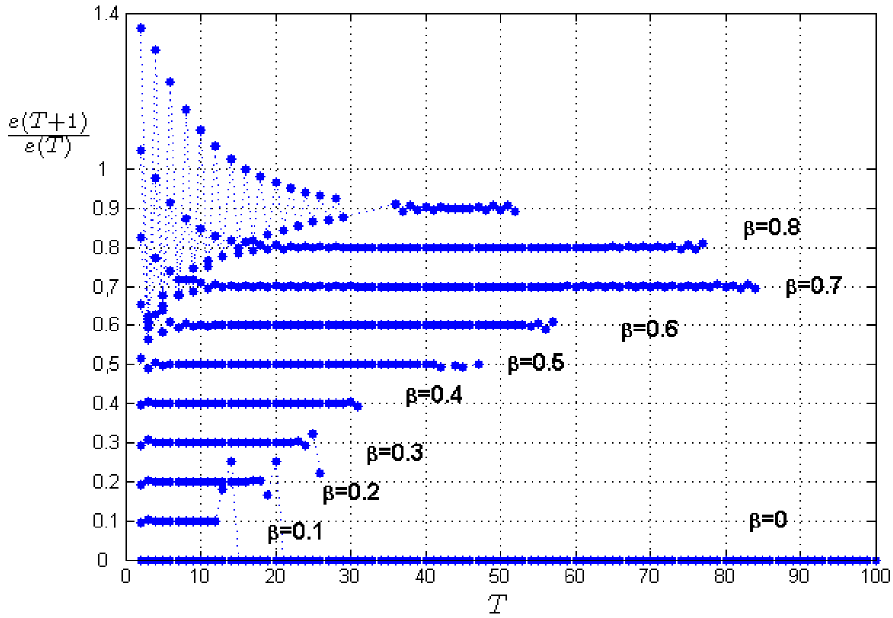

Figure 2: The Whittle's MP index convergence rate for different discount factors β.

6.3 Benchmarking the MP Index Policy

We have performed some small-scale preliminary computational studies to assess the relative performance of the *MP-index policy* against the alternative reviewed policies: the *TEV-index policy*, based on index $\lambda^{\mathrm{TEV}}(s) = d\,r\,s$, which has been proposed in [4,6] and is called the *greedy policy* there, and the *myopic-index policy*, based on the MP index $\lambda^{\mathrm{myopic}}(s) = d\,r[\phi^0(s) - \phi^1(s)]$ corresponding to $\beta = 0$.

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* 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, at each new instance, $q^{(1)}$ assumed values over the range $q^{(1)} \in \{0.5, 1, 2, \dots, 10\}$. The discount factor is $\beta = 0.99$.

The MP index was computed on-line for each target, truncating the corresponding infinite series to 10^3 terms based on the results of section 6.2. 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 1 reports the resulting TEV performance objective 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 MP index policy outperforms both the myopic and the TEV index policy. As for the MP index policy's suboptimality gap, we can bound it above using the relaxation's lower bound. Moreover, we observe that the MP index suboptimality gap is at least 2 % and at most 5 %. The MP index policy 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 MP 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 MP 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 MP index policy does. However, the MP 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 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 target's movement processes remain invariant, the movement process for target 1 becomes less volatile. In particular, at each new instance, $q^{(1)}$ assumed values over the range $q^{(1)} \in \{0.5, 1, 2, \dots, 10\}$. The discount factor is again $\beta = 0.99$.

Table 2 reports the resulting TEV performance objective achieved under each policy

Table 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

for each value of parameter $q^{(1)}$. The results show that also in this case the MP index policy outperforms both the myopic and TEV 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 MP index policy over the TEV index policy is 8.58 %, which is a significant amount. Among the TEV 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 MP 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 MP index policy's suboptimality gap, bounding it above by means of the relaxation's lower bound, we note that the MP index suboptimality gap is al least 2.31 % and at most 11.68 %.

6.4 Asymptotic Optimality of the MP Index Policy

Together with the restless bandit indexability property introduced in [18], Whittle conjectured that for a population with N projects, the policy of being active in the M projects of greatest MP index is asymptotically optimal as M and N tend to ∞ in constant ratio p with $p = M/N$.

Such a conjecture can be formulated in terms of the problem under study as follows. Denote as π_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.

Proposition 4 *For population of fixed composition in the sense that* $\pi_j \to \pi$ *as* $N \to \infty$,

Table 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

with all N targets being indexable, Whittle conjectured that

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

In [17] 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. $p = 1/4$) for tracking a population of $N = 4$ different targets (i.e. $\pi = 1/N$), 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 N vary over the range $N \in 4 * \{1, 2, \ldots, 40\}$. For each instance the MP 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 $s_0^{(n)} = 0$.

Based on the resulting TEV performance objective achieved under the MP index policy and on the lower bound provided by the Lagrangian relaxation apporach discussed above, an upper bound for the MP index policy suboptimality gap is computed for each population size N. Results, illustrated in Figure 3, show that the upper bound of the MP 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 MP policy to be nearly optimal for cases in which, given a constant beam per target ratio, target heterogeneity grows as the total number of objective targets goes to infinite. 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 suboptimality gap initially increases fast as N gets larger, tending to stabilize around 13.5 % for the largest values of N considered.

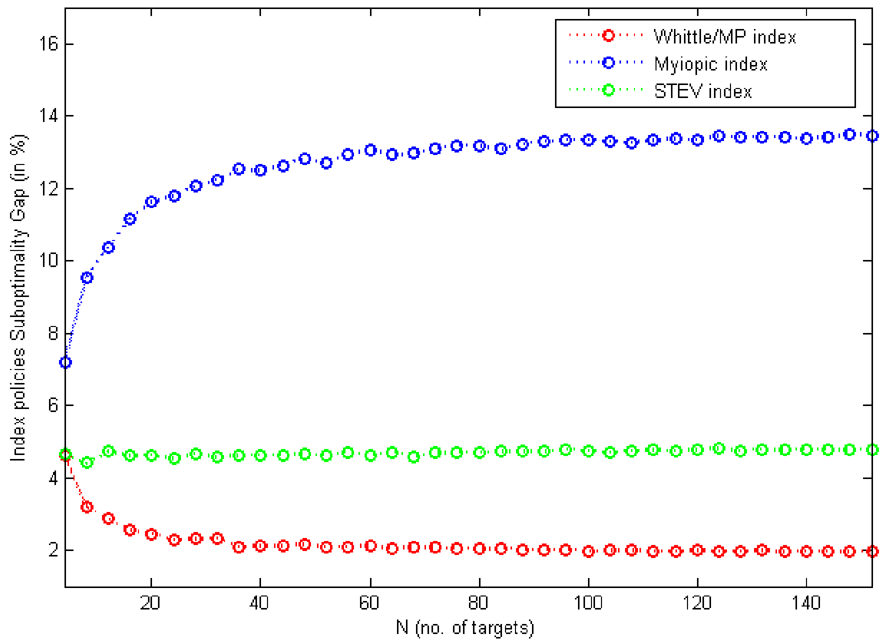

Figure 3: The Whittle's MP index suboptimality gap as $m, n \to \infty$ with $M = pN$.

7 Conclusion and Future Work

We have designed a novel tractable priority–index policy for scheduling target updates in a discrete-time multitarget tracking model based on the MARBP indexation methodology developed by the first author. Such MP index policy successfully addresses all key issues in the design of SM polices. Computational studies demonstrate the tractability of the MP index, suggesting that an on-line scheduling algorithm based on it is implementable. Moreover, the MP policy accounts for the long term effects of current actions and it exhibits performance advantages over other previously suggested policies when implemented in fairly distinct scenarios. In addition, the MP scheduling rule not only efficiently allocates the system's resources among objective targets but also it indicates when resources should idle.

Future work is required to extend these results to the more general model in which targets move in a multidimensional space. Also, a natural extension of this model is the case in which probability of target detection is no longer assumed to be unity, and hence probabilities of misdetection or false alarm are to be considered.

References

[1] J. C. Gittins. Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B*, 41(2):148–177, 1979. with discussion.

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

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

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

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

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

[7] J. Niño-Mora. Restless bandits, partial conservation laws and indexability. *Adv. Appl. Probab.*, 33(1):76–98, 2001.

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

[9] J. Niño-Mora. Marginal productivity index policies for scheduling multiclass wireless transmissions. In *Proc. 2nd Euro-NGI Conf. Next Generation Internet Networks — Design and Engineering (NGI 2006, Valencia, Spain)*, pages 342–349, Piscataway, NJ, 2006. IEEE.

[10] J. Niño-Mora. Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock M/G/1 queues. Math. Oper. Res., vol. 31(no. 1):50–84, 2006.

[11] J. Niño-Mora. Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):161–198, 2007. followed by discussions by I. J. B. F. Adan and O. J. Boxma, E. Altman, O. Hernández-Lerma, R. Weber, P. Whittle, and D. D. Yao.

[12] J. Niño-Mora. An index policy for dynamic fading-channel allocation to heterogeneous mobile users with partial observations. In *Proc. NGI 2008, 4th Euro-NGI Conf. Next Generation Internet Networks (Krakow, Poland)*, pages 231–238. IEEE, 2008.

[13] J. Niño-Mora. A restless bandit marginal productivity index for opportunistic spectrum access with sensing errors. In *Proc. NET-COOP 2009, 3rd Euro-NG Conf. Network Control and Optimization (Eindhoven, The Netherlands)*, Lecture Notes Comput. Sci., page in press, Berlin, 2009. Springer.

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

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

[16] S. Suvorova W. Moran and S. Howard. *"Application of sensor scheduling concepts to radar"*. Eds. New York, NY: Springer, 2008.

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

[18] P. Whittle. Restless bandits: Activity allocation in a changing world. 25A:287–298, 1988.

A Marginal Work Convergence Rate

From results reviewed in 4 we have that:

$$w(s,s) = \begin{cases} 1 & s \in [0, \phi_\infty^1) \\ (1-\beta) & s = \phi_\infty^1 \\ \frac{1-\beta}{1-\beta^{c(s)}}, & s \in (\phi_\infty^1, \infty) \\ 1, & s = \infty. \end{cases} \tag{34}$$

Note that from definition 16 it follows that:

$$w(s,s) = 1 + \beta\Big(g(\phi_1^1(s), s) - g(\phi_1^0(s), s)\Big) \tag{35}$$

For the target instance $s = 1$, $\theta = 0.5$ and $d = r = 1$ it holds that $a(1) = 1$ and also $p(1) = t_0^*(1, \phi_1^1(1)) = 1$, thus $c(1) = 2$. Then, it follows from 1 that: $g(\phi_1^1(1), 1) = \sum_{t=0}^{\infty} \beta^{2t+1} = \dfrac{\beta}{(1-\beta^2)}$ and $g(\phi_1^1(1), 1) = \sum_{t=0}^{\infty} \beta^{2t} = \dfrac{1}{(1-\beta^2)}$. Also, from 34 we have that

$w(1,1) = \frac{1}{1+\beta}$ for $0 \le \beta \le 1$, since $w(1,1) = \frac{1}{c(1)} = 1/2$ for $\beta = 1$. Denote the marginal work in STEV s computed by truncating the infinite series up to time slot T as $\hat{w}(s,s)_T$, and notice that:

$$\hat{w}(s,s)_T = 1 + \beta\Big[\sum_{t=0}^{T} \beta^t \Big(a_t(\phi_1^1(s), s) - a_t(\phi_1^0(s), s)\Big)\Big] \tag{36}$$

Proposition 5 *For $s = 1$, $\theta = 0.5$, $d = r = 1$, the following holds:*

$$\lim_{T\to\infty} \left| \frac{\hat{w}(s,s)_{T+1} - w(s,s)}{\hat{w}(s,s)_T - w(s,s)} \right| = \beta \tag{37}$$

Proof. From 36 we have that for $c(s) = 2$ 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,

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

Thus,

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

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

Notice that, despite the fact that the proof has been done for $s = 1$, $\theta = 0.5$, $d = r = 1$, it can be extended to the general case $s \in \mathsf{S}$ and some parameter specification $\theta, d, r \ge 0$.

B Möbius Transformations for Multitarget Tracking

We have considered two iterated mappings of the form $s \mapsto \phi^i(s)$ where s denotes the initial STEV and $i = 0, 1$ stands for passive and active actions respectively. Letting $\phi_0^i(s) \triangleq s$ and $\phi_t^i(s) \triangleq \phi^i(\phi_{t-1}^i(s))$ for $t \ge 1$, and defining:

$$\phi^0(s) = s + \theta \tag{38}$$

$$\phi^1(s) = \frac{s+\theta}{s+\theta+1} \tag{39}$$

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 this it is convenient to visualize them as Möbius Transformations, also known as Linear Fractional Transformations (LFT).

Theorem 3 *A Möbius transformation is a function* $m\colon \mathbb{C} \to \mathbb{C}$ *of the form*

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

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

Möbius transformations have the following useful property. A given Möbius transformations has the following associated matrix representation:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

And the composition of two Möbius transformations $n \circ m\,(z)$, with associated matrix N and M respectively, is also a Möbius transformations whose associated matrix is equal to the matrix product NM.
Note that the condition above expressed that $ad - cb \neq 0$ is equivalent to saying that $|M| \neq 0$. If $ad - cb = 0$ the $m(x) = c$ where c is a constant.
Every Möbius transformation whose associated matrix is not the identity matrix, has two fixed points that can be obtained by solving the following fixed point equation: $m(x) = x$. Denote these fixed points as γ_1 and γ_2, then:

$$\gamma_{1,2} = \frac{(a-d) \pm \sqrt{(a-d)^2 + 4bc}}{2c} \tag{40}$$

Möbius transformations can be written in terms of these fixed points in a so called **normal** form with the following associated matrix:

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

where $k = \frac{\lambda_2}{\lambda_1}$ and $\lambda_{1,2}$ are the eigenvalues of the H matrix and can be shown to be equal to: $\lambda_i = c\gamma_i + d$.
This representation will be of use in order to obtain the closed from expression corresponding to the t^{th} iterate, since it can be shown that: if the transformation of matrix H has fixed points γ_1,γ_2 and characteristic constant k, then $M' = M^n$ will have the same fixed points and characteristic constant equal to $k' = k^n$.
Thus, the t^{th} iterate of a Möbius transformation has the following associated matrix representation:

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

This expression of the LFT allows us to distinguish between attractive and repulsive fixed points of the transformation. Note that in the t^{th} iterate $m'(x)$ is equal to:

$$m'(x) = \frac{\gamma_1 x - \gamma_1\gamma_2 + k^n(\gamma_1\gamma_2 - \gamma_2 x)}{x - \gamma_2 + k^n(\gamma_1 - x)} \tag{41}$$

Thus, continuous iteration of the transformation leads us to:

$$\lim_{n\to\infty} m'(x) = \begin{cases} \frac{\gamma_1(x-\gamma_2)}{x-\gamma_2} = \gamma_1 & \text{if } |k| < 1 \\ \frac{\gamma_2(\gamma_1-x)}{\gamma_1-x} = \gamma_2 & \text{if } |k| > 1 \end{cases}$$

Therefore, whenever $|k| < 1$ we can say that γ_1 is an attractive fixed point while γ_2 is a repulsive fixed point, and for $|k| > 1$ roles are reversed.
Now, given all these elements we have that equation (38) and (39) define two Möbius transformations 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 (38), the corresponding LFT is a pure translation (since in this case $c = 0$ and $a = d$) and thus, both fixed points are at infinity. In this case we can not re-express the function in terms of the normal form because this form is only valid for LFT with two distinct fixed points. The solution to recursion (38) is easy to obtain and it is equal to:

$$\phi_t^0(s) \quad = \quad s + \theta t \tag{42}$$

Note that:

$$\frac{\partial \phi_t^0(s)}{\partial t} = \theta \geq 0 \qquad \lim_{t\to\infty} \phi_t^0(s) = \infty$$

From (42) and the above results we know that the attractive fixed point of the passive dynamics is at infinity, and that as t increases (i.e. as we systematically do not observe a target), the resulting STEV $\phi_t^0(s)$ increases.
Now, for any $s \leq z$ there is a first $t \geq 1$ for which $\phi_{t-1}^0(s) \leq z$ and $\phi_t^0(s) > z$, let us denote that critical iteration as $t_0^*(s,z)$, and note that in this case it holds that $t_0^*(s,z)$ is an integer such that:

$$\frac{z-s}{\theta} < t_0^*(s,z) \leq \left(\frac{z-s}{\theta}\right) + 1$$

This leads us to conclude that:

$$t_0^*(s,z) = \left\lfloor \frac{z-s}{\theta} \right\rfloor + 1$$

Thus, given the value of $\theta > 0$, $t_0^*(s,z)$ will be finite and greater as the initial value is further away from the threshold. Now let us analyze the Möbius transformation generated by equation (39). Its fixed points are equal to:

$$\gamma_{1,2} \quad = \quad \frac{1}{2}\left(-\theta \pm \sqrt{\theta(4+\theta)}\right) \tag{43}$$

The eigenvalues of matrix Φ^1 are the following:

$$\begin{aligned}\lambda_{1,2} &= \frac{(2+\theta) \pm \sqrt{\theta(4+\theta)}}{2} \\ \lambda_{1,2} &= \gamma_i + (1+\theta)\end{aligned} \tag{44}$$

Therefore, we have that the closed formula for the t-th iterate of the active recursion is an LFT whose associated matrix representation is the following:

$$\Phi'^1(\gamma_2, \gamma_1, k') = \begin{pmatrix} \gamma 1 - k^n \gamma_2 & (k^n - 1)\gamma_1\gamma_2 \\ 1 - k^n & k^n \gamma_1 - \gamma_2 \end{pmatrix}$$

Which indicates that γ_1 will be the fixed point of the recursion (due to the fact that $|k| < 1$).
Note also that, the associated matrix can be obtained as the t^{th} power of the matrix Φ^1, thus, it holds that:

$$(\Phi^1)^t = \begin{pmatrix} \frac{\lambda_1 - d}{c} & \frac{\lambda_2 - d}{c} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix} \begin{pmatrix} \frac{\lambda_1 - d}{c} & \frac{\lambda_2 - d}{c} \\ 1 & 1 \end{pmatrix}^{-1}$$

Now using the fact that: $\lambda_i = c\gamma_i + d$ We reexpress this as:

$$(\Phi^1)^t = \begin{pmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix} \begin{pmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{pmatrix}^{-1}$$

From where we can conclude that:

$$(\Phi^1)^t = \tfrac{1}{\gamma_1 - \gamma_2} \begin{pmatrix} \gamma_1 \lambda_1^t & \gamma_2 \lambda_2^t \\ \lambda_1^t & \lambda_2^t \end{pmatrix} \begin{pmatrix} 1 & -\gamma_2 \\ -1 & \gamma_1 \end{pmatrix}$$

Then,

$$(\Phi^1)^t = \tfrac{\lambda_2}{\gamma_1 - \gamma_2} \begin{pmatrix} \gamma_1 k^t - \gamma_2 & \gamma_1 \gamma_2 (1 - k^t) \\ k^t - 1 & \gamma_1 - \gamma_2 k^t \end{pmatrix}$$

Where $k = \frac{\lambda_2}{\lambda_1}$.
Another useful representation of the t-th iterate of a Möbius Transformation with distinct fixed points is the following:

$$\phi_t^1(s) = \frac{\alpha}{1 - k^n (1 - \frac{\alpha}{s - \gamma_2})} + \gamma_2 \tag{45}$$

with $\alpha = \sqrt{\left(\frac{a-d}{c}\right)^2 + \frac{4b}{c}}$, and for this particular case we have that: $\alpha = \sqrt{\theta(4+\theta)}$.
Note that:

$$\lim_{t \to \infty} \phi_t^1(s) = \alpha + \gamma_2 = \gamma_1 = \frac{1}{2}\left(\sqrt{\theta(4+\theta)} - \theta\right) \geq 0$$

Thus, let us define the minimum value attainable by continued measurement of a target ϕ^1_∞ as $\lim_{t\to\infty} \phi^1_t(s)$ and thus, conclude that:

$$\phi^1_\infty \triangleq \gamma_1 = \frac{1}{2}\left(\sqrt{\theta(4+\theta)} - \theta\right)$$

Also, notice that:

$$\frac{\partial \phi^1_t(s)}{\partial t} = \left[\frac{\alpha}{1 - k^n(1 - \frac{\alpha}{s-\gamma_2})}\right]^{-2} \left(1 - \frac{\alpha}{s - \gamma_2}\right) k^n (\log k)$$

Given that $\left(1 - \frac{\alpha}{s-\gamma_2}\right) = \frac{s-\phi^1_\infty}{s-\gamma_2}$, we conclude that:

$$\frac{\partial \phi^1_t(s)}{\partial t} \begin{cases} \leq 0 & \text{for } s \geq \phi^1_\infty \\ = 0 & \text{for } s = \phi^1_\infty \\ \geq 0 & \text{for } s \leq \phi^1_\infty \end{cases}$$

We will express the solution to active recursion (39) as in (45) for the sake of simplicity. Analogously to the passive recursion case, for any $s > z$ there is a first $t \geq 1$ for which $\phi^1_{t-1}(s) > z$ and $\phi^1_t(s) \leq z$, let us denote that critical iteration as $t^*_1(s, z)$, and note that in this case it holds that:

$$\frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} \leq t^*_1(s, z) < \frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} + 1$$

This leads us to conclude that:

$$t^*_1(s, z) = \left\lceil \frac{1}{\log k} \log \left\{ \frac{\left[1 - \frac{\alpha}{z-\gamma_2}\right]}{\left[1 - \frac{\alpha}{s-\gamma_2}\right]} \right\} \right\rceil$$

Thus, given the value of k, α, γ_2 associated to a certain θ value and a given threshold level z, $t^*_1(s, z)$ will be greater as the initial value is further away from the threshold.