Sequential Detection with Mutual Information Stopping Cost

Vikram Krishnamurthy, Robert Bitmead, Michel Gevers and Erik Miehling

Abstract

This paper formulates and solves a sequential detection problem that involves the mutual information (stochastic observability) of a Gaussian process observed in noise with missing measurements. The main result is that the optimal decision is characterized by a monotone policy on the partially ordered set of positive definite covariance matrices. This monotone structure implies that numerically efficient algorithms can be designed to estimate and implement monotone parametrized decision policies. The sequential detection problem is motivated by applications in radar scheduling where the aim is to maintain the mutual information of all targets within a specified bound. We illustrate the problem formulation and performance of monotone parametrized policies via numerical examples in fly-by and persistent-surveillance applications involving a GMTI (Ground Moving Target Indicator) radar.

Index Terms

Sequential detection, stopping time problem, mutual information, Kalman filter, radar tracking, monotone decision policy, lattice programming

Vikram Krishnamurthy (vikramk@ece.ubc.ca) and Erik Miehling (erikm@ece.ubc.ca) are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. Robert Bitmead (rbitmead@ucsd.edu) is with the Department of Mechanical and Aerospace Engineering, University of California San Diego, CA 92093-0411, USA. Michel Gevers (gevers@csam.ucl.ac.be) is with the Department of Mathematical Engineering, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

I. INTRODUCTION

Consider the following sequential detection problem. L targets (Gaussian processes) are allocated priorities $\nu_1, \nu_2, \ldots, \nu_L$. A sensor obtains measurements of these L evolving targets with signal to noise ratio (SNR) for target l proportional to priority ν_l. A decision maker has two choices at each time k: If the decision maker chooses action $u_k = 2$ (continue) then the sensor takes another measurement and accrues a measurement cost c_ν. If the decision maker chooses action $u_k = 1$ (stop), then a stopping cost proportional to the mutual information (stochastic observability) of the targets is accrued and the problem terminates. What is the optimal time for the decision maker to apply the stop action? Our main result is that the optimal decision policy is a monotone function of the target covariances (with respect to the positive definite partial ordering). This facilitates devising numerically efficient algorithms to compute the optimal policy.

The sequential detection problem addressed in this paper is non-trivial since the decision to continue or stop is based on Bayesian estimates of the targets' states. In addition to Gaussian noise in the measurement process, the sensor has a non-zero probability of missing observations. Hence, the sequential detection problem is a partially observed stochastic control problem. Targets with high priority are observed with higher SNR and the uncertainty (covariance) of their estimates decreases. Lower priority targets are observed with lower SNR and their relative uncertainty increases. The aim is to devise a sequential detection policy that maintains the stochastic observability (mutual information or conditional entropy) of all targets within a specified bound.

Why stochastic observability? As mentioned above, the stopping cost in our sequential detection problem is a function of the mutual information (stochastic observability) of the targets. The use of mutual information as a measure of stochastic observability was originally investigated in [1]. In [2], determining optimal observer trajectories to maximize the stochastic observability of a single target is formulated as a stochastic dynamic programming problem – but no structural results or characterization of the optimal policy is given; see also [3]. We also refer to [4] where a nice formulation of sequential waveform design for MIMO radar is given using a Kullback-Leibler divergence based approach. As described in Section III-C, another favorable property of stochastic observability is that its monotonicity with respect to covariances does not require stability of the state matrix of the target (eigenvalues strictly inside the unit circle). In target models, the state matrix for the dynamics of the target has eigenvalues at 1 and thus is not stable.

Organization and Main Results:

(i) To motivate the sequential detection problem, Section II presents a GMTI (Ground moving

target indicator) radar with macro/micro-manager architecture and a linear Gaussian state space model for the dynamics of each target. A Kalman filter is used to track each target over the time scale at which the micro-manager operates. Due to the presence of missed detections, the covariance update via the Riccati equation is measurement dependent (unlike the standard Kalman filter where the covariance is functionally independent of the measurements).

(ii) In Section III, the sequential detection problem is formulated. The cost of stopping is the stochastic observability which is based on the mutual information of the targets. The optimal decision policy satisfies Bellman's dynamic programming equation. However, it is not possible to compute the optimal policy in closed form.[1] Despite this, our main result (Theorem 1) shows that the optimal policy is a monotone function of the target covariances. This result is useful for two reasons: (a) Algorithms can be designed to construct policies that satisfy this monotone structure. (b) The monotone structural result holds without stability assumptions on the linear dynamics. So there is an inherent robustness of this result since it holds even if the underlying model parameters are not exactly specified.

(iii) Section IV exploits the monotone structure of the optimal decision policy to construct finite dimensional parametrized policies. Then a simulation-based stochastic approximation (adaptive filtering) algorithm (Algorithm 1) is given to compute these optimal parametrized policies. The practical implication is that, instead of solving an intractable dynamic programming problem, we exploit the monotone structure of the optimal policy to compute such parametrized policies in polynomial time.

(iv) Section V presents a detailed application of the sequential detection problem in GMTI radar resource management. By bounding the magnitude of the nonlinearity in the GMTI measurement model, we show that for typical operating values, the system can be approximated by a linear time invariant state space model. Then detailed numerical examples are given that use the above monotone policy and stochastic approximation algorithm to demonstrate the performance of the radar management algorithms. We present numerical results for two important GMTI surveillance problems, namely, the target fly-by problem and the persistent surveillance problem. In both cases, detailed numerical examples are given and the performance is compared with periodic

---

[1]For stochastic control problems with continuum state spaces such as considered in this paper, apart from special cases such as linear quadratic control and partially observed Markov decision processes, there are no finite dimensional characterizations of the optimal policy [5]. Bellman's equation does not translate into practical solution methodologies since the state space is a continuum. Quantizing the space of covariance matrices to a finite state space and then formulating the problem as a finite-state Markov decision process is infeasible since such quantization typically would require an intractably large state space.

stopping policies. Persistent surveillance has received much attention in the defense literature [6], [7], since it can provide critical, long-term surveillance information. By tracking targets for long periods of time using aerial based radars, such as DRDC-Ottawa's XWEAR radar [6] or the U.S. Air Force's Gorgon Stare Wide Area Airborne Surveillance System, operators can "rewind the tapes" in order to determine the origin of any target of interest [7].

(v) The appendix presents the proof of Theorem 1. It uses lattice programming and supermodularity. A crucial step in the proof is that the conditional entropy described by the Riccati equation update is monotone. This involves use of Theorem 2 which derives monotone properties of the Riccati and Lyapunov equations. The idea of using lattice programming and supermodularity to prove the existence of monotone policies is well known in stochastic control, see [8] for a textbook treatment of the countable state Markov decision process case. However, in our case since the state space comprises covariance matrices that are only partially ordered, the optimal policy is monotone with respect to this partial order. The structural results of this paper allow us to determine the nature of the optimal policy without brute force numerical computation.

Motivation – GMTI Radar Resource Management: This paper is motivated by GMTI radar resource management problems [9], [10], [11]. The radar macro-manager deals with priority allocation of targets, determining regions to scan, and target revisit times. The radar *micro-manager* controls the target tracking algorithm and determines how long to maintain a priority allocation set by the macro-manager. In the context of GMTI radar micro-management, the sequential detection problem outlined above reads: Suppose the radar macro-manager specifies a particular target priority allocation. How long should the micro-manager track targets using the current priority allocation before returning control to the macro-manager? Our main result, that the optimal decision policy is a monotone function of the targets' covariances, facilitates devising numerically efficient algorithms for the optimal radar micro-management policy.

II. RADAR MANAGER ARCHITECTURE AND TARGET DYNAMICS

This section motivates the sequential detection problem by outlining the macro/micro-manager architecture of the GMTI radar and target dynamics. (The linear dynamics of the target model are justified in Section V-A where a detailed description is given of the GMTI kinematic model).

A. Macro- and Micro-manager Architecture

(The reader who is uninterested in the radar application can skip this subsection.) Consider a GMTI radar with an agile beam tracking L ground moving targets indexed by $l \in \{1, \ldots, L\}$.

In this section we describe a two-time-scale radar management scheme comprised of a micro-manager and a macro-manager.

a) Macro-manager: At the beginning of each scheduling interval n, the radar macro-manager allocates the target priority vector $\nu_n = (\nu_n^1, \ldots, \nu_n^L)$. Here the priority of target l is $\nu_n^l \in [0,1]$ and $\sum_{l=1}^L \nu_n^l = 1$. The priority weight ν_n^l determines what resources the radar devotes to target l. This affects the track variances as described below. The choice ν_n is typically rule-based, depending on several extrinsic factors. For example, in GMTI radar systems, the macro-manager picks the target priority vector ν_{n+1} based on the track variances (uncertainty) and threat levels of the L targets. The track variances of the L targets are determined by the Bayesian tracker as discussed below.

b) Micro-manager: Once the target priority vector ν is chosen (we omit the subscript n for convenience), the micro-manager is initiated. The clock on the fast time scale k (which is called the decision epoch time scale in Section V-A) is reset to $k = 0$ and commences ticking. At this decision epoch time scale, $k = 0, 1, \ldots$, the L targets are tracked/estimated by a Bayesian tracker. Target l with priority ν^l is allocated the fraction ν^l of the total number of observations (by integrating $\nu^l \Delta$ observations on the fast time scale, see Section V-A) so that the observation noise variance is scaled by $1/(\nu_l \Delta)$. The question we seek to answer is: *How long should the micro-manager track the L targets with priority vector ν before returning control to the macro-manager to pick a new priority vector?* We formulate this as a sequential decision problem.

Note that the priority allocation vector ν and track variances of the L targets capture the interaction between the micro- and macro-managers.

B. Target Kinematic Model and Tracker

We now describe the target kinematic model at the epoch time scale k: Let $s_k^l = \left[x_k^l, \dot{x}_k^l, y_k^l, \dot{y}_k^l \right]^T$ denote the Cartesian coordinates and velocities of the ground moving target $l \in \{1, \ldots, L\}$. Section V-A shows that on the micro-manager time scale, the GMTI target dynamics can be approximated as the following linear time invariant Gaussian state space model

$$s_{k+1}^l = F s_k^l + G w_k^l,$$

$$z_k^l = \begin{cases} H s_k^l + \frac{1}{\sqrt{\nu^l \Delta}} v_k^l, & \text{with probability } p_d^l, \\ \emptyset, & \text{with probability } 1 - p_d^l. \end{cases} \tag{1}$$

The parameters F, G, H are defined in Section V. They can be target (l) dependent; to simplify notation we have not done this. In (1), z_k^l denotes a 3-dimensional observation vector of target l

at epoch time k. The noise processes w_k^l and $v_k^l/\sqrt{\nu^l \Delta}$ are mutually independent, white, zero-mean Gaussian random vectors with covariance matrices Q^l and $R^l(\nu^l)$, respectively. (Q and R are defined in Section V). Finally, p_d^l denotes the probability of detection of target l, and $\emptyset$ represents a missed observation that contains no information about state s.[2]

Define the one-step-ahead predicted covariance matrix of target l at time k as

$$P_k^l = \mathbb{E}\left\{ \left(s_k^l - \mathbb{E}\{s_k^l|z_{1:k-1}^l\}\right)\left(s_k^l - \mathbb{E}\{s_k^l|z_{1:k-1}^l\}\right)^T \right\}.$$

Here the superscript T denotes transpose. Based on the priority vector ν and model (1), the covariance of the state estimate of target $l \in \{1, \ldots, L\}$ is computed via the following measurement dependent Riccati equation

$$P_{k+1}^l = \mathcal{R}(P_k^l, z_k) \overset{\text{def}}{=} F P_k^l F^T + Q^l - I(z_k^l \neq \emptyset) F P_k^l H^T \left(H P_k^l H^T + R^l(\nu^l)\right)^{-1} H P_k^l F^T. \quad (2)$$

Here $I(\cdot)$ denotes the indicator function. In the special case when a target l is allocated zero priority ($\nu^{(l)} = 0$), or when there is a missing observation ($z_k^l = \emptyset$), then (2) specializes to the Kalman predictor updated via the Lyapunov equation

$$P_{k|k-1}^l = \mathcal{L}(P_k^l) \overset{\text{def}}{=} F P_{k-1}^l F^T + Q^l. \quad (3)$$

III. SEQUENTIAL DETECTION PROBLEM

This section presents our main structural result on the sequential detection problem. Section III-A formulates the stopping cost in terms of the mutual information of the targets being tracked. Section III-B formulates the sequential detection problem. The optimal decision policy is expressed as the solution of a stochastic dynamic programming problem. The main result (Theorem 1 in Section III-C) states that the optimal policy is a monotone function of the target covariance. As a result, the optimal policy can be parametrized by monotone policies and estimated in a computationally efficient manner via stochastic approximation (adaptive filtering) algorithms. This is described in Section IV.

Notation: Given the priority vector ν allocated by the macro-manager, let $a \in \{1, \ldots, L\}$ denote the highest priority target, i.e, $a = \arg\max_l \nu^l$. Its covariance is denoted P^a. We use the notation P^{-a} to denote the set of covariance matrices of the remaining $L - 1$ targets. The sequential decision problem below is formulated in terms of (P^a, P^{-a}).

---

[2]With suitable notational abuse, we use '$\emptyset$' as a label to denote a missing observation. When a missing observation is encountered, the track estimate is updated by the Kalman predictor with covariance update (3).

A. Formulation of Mutual Information Stopping Cost

As mentioned in Section II-A, once the radar macro-manager determines the priority vector ν, the micro-manager switches on and its clock $k = 1, 2, \ldots$ begins to tick. The radar micro-manager then solves a sequential detection problem involving two actions: At each slot k, the micro-manager chooses action $u_k \in \{1 \text{ (stop)}, 2 \text{ (continue)}\}$. To formulate the sequential detection problem, this subsection specifies the costs incurred with these actions.

Radar Operating cost: If the micro-manager chooses action $u_k = 2$ (continue), it incurs the radar operating cost denoted as c_ν. Here $c_\nu > 0$ depends on the radar operating parameters,

Stopping cost – Stochastic Observability: If the micro-manager chooses action $u_k = 1$ (stop), a stopping cost is incurred. In this paper, we formulate a stopping cost in terms of the stochastic observability of the targets, see also [1], [2]. Define the stochastic observability of each target $l \in \{1, , \ldots, L\}$ as the mutual information

$$I(s_k^l; z_{1:k}^l) = \alpha^l h(s_k^l) - \beta^l h(s_k^l | z_{1:k}^l). \tag{4}$$

In (4), α^l and β^l are non-negative constants chosen by the designer. Recall from information theory [12], that $h(s_k^l)$ denotes the differential entropy of target l at time k. Also $h(s_k^l | z_{1:k}^l)$ denotes the conditional differential entropy of target l at time k given the observation history $z_{1:k}^l$. The mutual information $I(s_k^l; z_{1:k}^l)$ is the average reduction in uncertainty of the target's coordinates s_k^l given measurements $z_{1:k}^l$. In the standard definition of mutual information $\alpha^l = \beta^l = 1$. However, we are also interested in the special case when $\alpha^l = 0$, in which case, we are considering the conditional entropy for each target (see Case 4 below).

Consider the following stopping cost if the micro-manager chooses action $u_k = 1$ at time k:

$$\bar{C}(s_k, z_{1:k}) = -I(s_k^a; z_{1:k}^a) + \mathbf{F}(\{I(s_k^l, z_{1:k}^l); l \neq a\}). \tag{5}$$

Recall a denotes the highest priority target. In (5), $\mathbf{F}(\cdot)$ denotes a function chosen by the designer to be monotone increasing in each of its $L - 1$ variables (examples are given below).

The following lemma follows from straightforward arguments in [12].

Lemma 1: Under the assumption of linear Gaussian dynamics (1) for each target l, the mutual information of target l defined in (4) is

$$I(s_k^l, z_{1:k}^l) = \alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|, \tag{6}$$

where $\bar{P}_k^l = \mathbb{E}\{(s_k^l - \mathbb{E}\{s_k^l\})(s_k^l - \mathbb{E}\{s_k^l\})^T\}$, $P_k^l = \mathbb{E}\{(s_k^l - \mathbb{E}\{s_k^l | z_{1:k}\})(s_k^l - \mathbb{E}\{s_k^l | z_{1:k}\})^T\}$. Here $\bar{P}_k^l$ denotes the predicted (a priori) covariance of target l at epoch k given no observations.

It is computed using the Kalman predictor covariance update (3) for k iterations. Also, P_k^l is the posterior covariance and is computed via the Kalman filter covariance update (2). ∎

Using Lemma 1, the stopping cost $\bar{C}(\cdot, \cdot)$ in (5) can be expressed in terms of the Kalman filter and predictor covariances. Define the four-tuple of sets of covariance matrices

$$P_k = (P_k^a, \bar{P}_k^a, P_k^{-a}, \bar{P}_k^{-a}). \tag{7}$$

Therefore the stopping cost (5) can be expressed as

$$\bar{C}(P_k) = -\alpha^a \log |\bar{P}_k^a| + \beta^a \log |P_k^a| + \mathbf{F}\left(\{\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|; \ l \neq a\}\right). \tag{8}$$

Examples: We consider the following examples of $\mathbf{F}(\cdot)$ in (8) :

Case 1. Maximum mutual information difference stopping cost: $\bar{C}(s_k, z_{1:k}) = -I(s_k^a, z_{1:k}^a) + \max_{l \neq a} I(s_k^l, z_{1:k}^l)$ in which case,

$$\bar{C}(P_k) = -\alpha^a \log |\bar{P}_k^a| + \beta^a \log |P_k^a| + \max_{l \neq a}\left[\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|\right]. \tag{9}$$

The stopping cost is the difference in mutual information between the target with highest mutual information and the target with highest priority. This can be viewed as a stopping cost that discourages stopping too soon.

Case 2. Minimum mutual information difference stopping cost: $\bar{C}(s_k, z_{1:k}) = -I(s_k^a, z_{1:k}^a) + \min_{l \neq a} I(s_k^l, z_{1:k}^l)$ in which case,

$$\bar{C}(P_k) = -\alpha^a \log |\bar{P}_k^a| + \beta^a \log |P_k^a| + \min_{l \neq a}\left[\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|\right]. \tag{10}$$

The stopping cost is the difference in mutual information between the target with lowest mutual information and the target with highest priority. This can be viewed as a conservative stopping cost in the sense that preference is given to stop sooner.

Case 3. Average mutual information difference stopping cost: $\bar{C}(s_k, z_{1:k}) = -I(s_k^a, z_{1:k}^a) + \sum_{l \neq a} I(s_k^l, z_{1:k}^l)$ in which case,

$$\bar{C}(P_k) = -\alpha^a \log |\bar{P}_k^a| + \beta^a \log |P_k^a| + \sum_{l \neq a}\left[\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|\right]. \tag{11}$$

This stopping cost is the difference between the average mutual information of the $L-1$ targets (if α^l and β^l include a $1/(L-1)$ term) and the highest priority target.

Case 4. Conditional differential entropy difference stopping cost: We are also interested in the following special case which involves scheduling between a Kalman filter and $L-1$ measurement-free Kalman predictors, see [13]. Suppose the high priority target a is allocated a Kalman

filter and the remaining $L - 1$ targets are allocated measurement-free Kalman predictors. This corresponds to the case where $\nu^a = 1$ and $\nu^l = 0$ for $l \neq a$ in (1), that is, the radar assigns all its resources to target a and no resources to any other target. Then solving the sequential detection problem is equivalent to posing the following question: What is the optimal stopping time τ when the radar should decide to start tracking another target? In this case, the mutual information of each target $l \neq a$ is zero (since $\bar{P}_k^l = P_k^l$ in (6)). So it is appropriate to choose $\alpha^l = 0$ for $l \neq a$ in (8). Note from (4), that when $\alpha^l = 0$, the stopping cost of each individual target becomes the negative of its conditional entropy. That is, the stopping cost is the difference in the conditional differential entropy instead of the mutual information.

B. Formulation of Sequential Decision Problem

With the above stopping and continuing costs, we are now ready to formulate the sequential detection problem that we wish to solve. Let μ denote a stationary decision policy of the form

$$\mu : P_k \rightarrow u_{k+1} \in \{1 \text{ (stop)}, 2 \text{ (continue)}\}. \tag{12}$$

Recall from (7) that P_k is a 4-tuple of sets of covariance matrices. Let $\boldsymbol{\mu}$ denote the family of such stationary policies. For any prior 4-tuple P_0 (recall notation (7)) and policy $\mu \in \boldsymbol{\mu}$ chosen by the micro-manager, define the stopping time $\tau = \inf\{k : u_k = 1\}$. The following cost is associated with the sequential decision procedure:

$$J_\mu(P) = \mathbb{E}^\mu\{(\tau - 1)c_\nu + \bar{C}(P_\tau)|P_0 = P\}. \tag{13}$$

Here c_ν is the radar operating cost and $\bar{C}$ the stopping cost introduced in Section III-A. Also, $\mathbb{E}^\mu$ denotes expectation with respect to stopping time τ and initial condition P. (A measure-theoretic definition of $\mathbb{E}^\mu$, which involves an absorbing state to deal with stopping time τ, is given in [14]).

The goal is to determine the optimal stopping time τ with minimal cost, that is, compute the optimal policy $\mu^* \in \boldsymbol{\mu}$ to minimize (13). Denote the optimal cost as

$$J_{\mu^*}(P) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(P). \tag{14}$$

The existence of an optimal stationary policy μ^* follows from [5, Prop.1.3, Chapter 3]. Since c_ν is non-negative, for the conditional entropy cost function of Case 4 in Section III-A, stopping is guaranteed in finite time, i.e., τ is finite with probability 1. For Cases (1) to (3), in general τ is not necessarily finite – however, this does not cause problems from a practical point of

view since the micro-manager has typically a pre-specified upper time bound at which it always chooses $u_k = 1$ and reverts back to the macro-manager. Alternatively, for Cases (1) to (3), if one truncates $\bar{C}(P)$ to some upper bound, then again stopping is guaranteed in finite time.

Considering the above cost (13), the optimal stationary policy $\mu^* \in \boldsymbol{\mu}$ and associated value function $\bar{V}(P) = J_{\mu^*}(P)$ are the solution of the following "Bellman's dynamic programming equation" [8] (Recall our notation $P = (P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a})$.)

$$\bar{V}(P) = \min\big\{\bar{C}(P), c_\nu + \mathbb{E}_z\left[V\big(\mathcal{R}(P^a, z^a), \mathcal{L}(\bar{P}^a), \mathcal{R}(P^{-a}, z^{-a}), \mathcal{L}(\bar{P}^{-a})\big)\right]\big\},$$

$$\mu^*(P) = \arg\min\big\{\bar{C}(P), c_\nu + \mathbb{E}_z\left[V\big(\mathcal{R}(P^a, z^a), \mathcal{L}(\bar{P}^a), \mathcal{R}(P^{-a}, z^{-a}), \mathcal{L}(\bar{P}^{-a})\big)\right]\big\}, \qquad (15)$$

where $\mathcal{R}$ and $\mathcal{L}$ were defined in (2) and (3). Here $\mathcal{R}(P^{-a}, z^{-a})$ denotes the Kalman filter covariance update for the $L-1$ lower priority targets according to (2). Our goal is to characterize the optimal policy μ^* and optimal stopping set defined as

$$\mathcal{S}_{\text{stop}} = \{(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}) : \mu^*(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}) = 1\}. \qquad (16)$$

In the special Case 4 of Section III-A, when $\alpha^l = 0$, then $\mathcal{S}_{\text{stop}} = \{(P^a, P^{-a}) : \mu^*(P^a, P^{-a}) = 1\}$.

The dynamic programming equation (15) does not translate into practical solution methodologies since the space of P, 4-tuples of sets of positive definite matrices, is uncountable, and it is not possible to compute the optimal decision policy in closed form.

C. Main Result: Monotone Optimal Decision Policy

Our main result below shows that the optimal decision policy μ^* is a monotone function of the covariance matrices of the targets. To characterize μ^* in the sequential decision problem below, we introduce the following notation:

Let m denote the dimension of the state s in (1). (In the GMTI radar example $m = 4$).

Let $\mathcal{M}$ denote the set of all $m \times m$ real-valued, symmetric positive semi-definite matrices. For $P, Q \in \mathcal{M}$ define the positive definite partial ordering $\succeq$ as $P \succeq Q$ if $x^T P x \geq x^T Q x$ for all $x \neq 0$, and $P \succ Q$ if $x^T P x > x^T Q x$ for $x \neq 0$. Define $\preceq$ with the inequalities reversed. Notice that $[\mathcal{M}, \succeq]$ is a partially ordered set (poset).

Note that ordering positive definite matrices also orders their eigenvalues. Let $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ denote vectors with elements in $\mathbb{R}_+$. Then define the componentwise partial order on $\mathbb{R}^m$ (denoted by $\preceq_l$) as $x \preceq_l y$ (equivalently, $y \succeq_l x$) if $x_i \leq y_i$ for all $i = 1, \ldots, m$.

For any matrix $P \in \mathcal{M}$, let $\lambda_P \in \mathbb{R}_+^m$ denote the eigenvalues of P arranged in decreasing order as a vector. Note $P \succeq Q$ implies $\lambda_P \succeq_l \lambda_Q$. Clearly, $[\mathbb{R}_+^m, \succeq_l]$ is a poset.

Define scalar function f to be increasing[3] if $\lambda_P \preceq_l \lambda_Q$ implies $f(\lambda_P) \leq f(\lambda_Q)$, or equivalently, if $P \preceq Q$ implies $f(P) < f(Q)$. Finally we say that $f(P^{-a})$ is increasing in P^{-a} if $f(\cdot)$ is increasing in each component P^l of P^{-a}, $l \neq a$.

The following is the main result of this paper regarding the policy $\mu^*(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a})$.

Theorem 1: Consider the sequential detection problem (13) with stochastic observability cost (8) and stopping set (16).

1) The optimal decision policy $\mu^*(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a})$ is increasing in P^a, decreasing in $\bar{P}^a$, decreasing in P^{-a}, and increasing in $\bar{P}^{-a}$ on the poset $[\mathcal{M}, \succeq]$. Alternatively, $\mu^*(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a})$ is increasing in λ_{P^a}, decreasing in $\lambda_{\bar{P}^a}$, decreasing in $\lambda_{P^{-a}}$ and increasing in $\lambda_{\bar{P}^{-a}}$ on the poset $[\mathbb{R}^m_+, \succeq_l]$. Here $\lambda_{P^{-a}}$ denotes the $L-1$ vectors of eigenvalues λ_{P^l}, $l \neq a$ (and similarly for $\lambda_{\bar{P}^a}$).

2) In the special case when $\alpha^l = 0$ for all $l \in \{1, \ldots, L\}$, (i.e., Case 4 in Section III-A where stopping cost is the conditional entropy) the optimal policy $\mu^*(P^a, P^{-a})$ is increasing in P^a and decreasing in P^{-a} on the poset $[\mathcal{M}, \succeq]$. Alternatively, $\mu^*(P^a, P^{-a})$ is increasing in λ_{P^a}, and decreasing in $\lambda_{P^{-a}}$ on the poset $[\mathbb{R}^m_+, \succeq_l]$. ■

The proof is in Appendix B. The monotone property of the optimal decision policy μ^* is useful since (as described in Section IV) parametrized monotone policies are readily implementable at the radar micro-manager level and can be adapted in real time. Note that in the context of GMTI radar, the above policy is equivalent to the radar micro-manager *opportunistically* deciding when to stop looking at a target: If the measured quality of the current target is better than some threshold, then continue; otherwise stop.

To get some intuition, consider the second claim of Theorem 1 when each state process has dimension $m = 1$. Then the covariance of each target is a non-negative scalar. The second claim of Theorem 1 says that there exists a threshold switching curve $P^a = g(P^a)$, where $g(\cdot)$ is increasing in each element of P^a, such that for $P^a < g(P^{-a})$ it is optimal to stop, and for $P^a \geq g(P^{-a})$ it is optimal to continue. This is illustrated in Figure 1. Moreover, since g is monotone, it is differentiable almost everywhere (by Lebesgue's theorem).

To prove Theorem 1 we will require the following monotonicity result regarding the Riccati

---

[3]Throughout this paper, we use the term "increasing" in the weak sense. That is "increasing" means non-decreasing. Similarly, the term "decreasing" means non-increasing.
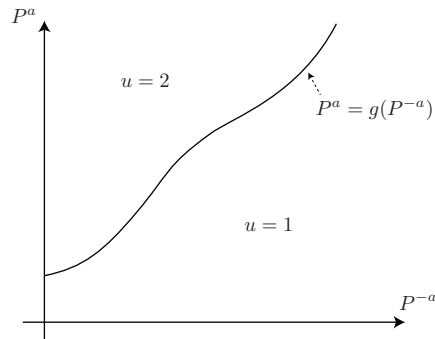
Fig. 1. Threshold switching curve for optimal decision policy $\mu^*(P^a, P^{-a})$. Claim 2 of Theorem 1 says that the optimal decision policy is characterized by a monotone increasing threshold curve $g(\cdot)$ when each target has state dimension $m = 1$.

and Lyapunov equations of the Kalman covariance update. This is proved in Appendix C. Below $\det(\cdot)$ denotes determinant.

Theorem 2: Consider the Kalman filter Riccati covariance update, $\mathcal{R}(P, z)$, defined in (2) with possibly missing measurements, and Lyapunov covariance update, $\mathcal{L}(P)$, defined in (3). The following properties hold for $P \in \mathcal{M}$ and $z \in \mathbb{R}^{m_z}$ (where m_z denotes the dimension of the observation vector z in (1)) :

(i) $\frac{\det(\mathcal{L}(P))}{\det(P)}$ and (ii) $\frac{\det(\mathcal{R}(P,z))}{\det(P)}$ are monotone decreasing in P on the poset $[\mathcal{M}, \succeq]$. ■

Discussion: An important property of Theorem 2 is that stability of the target system matrix F (see (24)) is not required. In target tracking models (such as (1)), F has eigenvalues at 1 and is therefore not stable. By using Theorem 2, Lemma 4 (in Appendix A) shows that the stopping cost involving stochastic observability is a monotone function of the covariances. This monotone property of the stochastic observability of a Gaussian process is of independent interest.

Instead of stochastic observability (which deals with log-determinants), suppose we had chosen the stopping cost in terms of the trace of the covariance matrices. Then, in general, it is not true that $\text{trace}(\mathcal{R}(P, z)) - \text{trace}(P)$ is decreasing in P on the poset $[\mathcal{M}, \succeq]$. Such a result typically requires stability of F.

IV. PARAMETRIZED MONOTONE POLICIES AND STOCHASTIC OPTIMIZATION ALGORITHMS

Theorem 1 shows that the optimal sequential decision policy $\mu^*(P) = \arg\inf_{\mu \in \boldsymbol{\mu}} J_\mu(P)$ is monotone in P. Below, we characterize and compute optimal parametrized decision policies of the form $\mu_{\theta^*}(P) = \arg\inf_{\theta \in \Theta} J_{\mu_\theta}(P)$ for the sequential detection problem formulated in Section III-B. Here $\theta \in \Theta$ denotes a suitably chosen finite dimensional parameter and Θ is a subset of Euclidean space. Any such parametrized policy $\mu_{\theta^*}(P)$ needs to capture the essential feature of Theorem 1: it needs to be decreasing in $P^{-a}, \bar{P}^a$ and increasing in $P^a, \bar{P}^{-a}$. In this section, we derive several examples of parametrized policies that satisfy this property. We then present simulation-based adaptive filtering (stochastic approximation) algorithms to estimate these optimal parametrized policies. To summarize, instead of attempting to solve an intractable dynamic programming problem (15), we exploit the monotone structure of the optimal decision policy (Theorem 1) to estimate a parametrized optimal monotone policy (Algorithm 1 below).

A. Parametrized Decision Policies

Below we give several examples of parametrized decision policies for the sequential detection problem that are monotone in the covariances. Because such parametrized policies satisfy the conclusion of Theorem 1, they can be used to approximate the monotone optimal policy of the sequential detection problem. Lemma 2 below shows that the constraints we specify are necessary and sufficient for the parametrized policy to be monotone implying that such policies $\mu_{\theta^*}(P)$ are an approximation to the optimal policy $\mu^*(P)$ within the appropriate parametrized class Θ.

First we consider 3 examples of parametrized policies that are linear in the vector of eigenvalues λ (defined in Section III-C). Recall that m denotes the dimension of state s in (1). Let θ^l and $\underline{\theta}^l \in \Theta = \mathbb{R}_+^m$ denote the parameter vectors that parametrize the policy μ_θ defined as

$$\mu_\theta(\lambda^a, \lambda^{-a}) = \begin{cases} 1 \text{ (stop)}, & \text{if } -\theta^{aT}\lambda_{P^a} + \underline{\theta}^{aT}\lambda_{\bar{P}^a} + \max_{l \neq a}\left[\theta^{lT}\lambda_{P^l} - \underline{\theta}^{lT}\lambda_{\bar{P}^l}\right] \geq 1, \\ 2 \text{ (continue)}, & \text{otherwise.} \end{cases} \tag{17}$$

$$\mu_\theta(\lambda^a, \lambda^{-a}) = \begin{cases} 1 \text{ (stop)}, & \text{if } -\theta^{aT}\lambda_{P^a} + \underline{\theta}^{aT}\lambda_{\bar{P}^a} + \min_{l \neq a}\left[\theta^{lT}\lambda_{P^l} - \underline{\theta}^{lT}\lambda_{\bar{P}^l}\right] \geq 1, \\ 2 \text{ (continue)}, & \text{otherwise.} \end{cases} \tag{18}$$

$$\mu_\theta(\lambda^a, \lambda^{-a}) = \begin{cases} 1 \text{ (stop)}, & \text{if } -\theta^{aT}\lambda_{P^a} + \underline{\theta}^{aT}\lambda_{\bar{P}^a} + \sum_{l \neq a}\left[\theta^{lT}\lambda_{P^l} - \underline{\theta}^{lT}\lambda_{\bar{P}^l}\right] \geq 1, \\ 2 \text{ (continue)}, & \text{otherwise.} \end{cases} \tag{19}$$

As a fourth example, consider the parametrized policy in terms of covariance matrices. Below θ^l and $\underline{\theta}^l \in \mathbb{R}^m$ are unit-norm vectors, i.e, $\theta^{lT}\theta^l = 1$ and $\underline{\theta}^{lT}\underline{\theta}^l = 1$ for $l = 1, \ldots, L$. Let $\mathcal{U}$ denote the space of unit-norm vectors. Define the parametrized policy μ_θ, $\theta \in \Theta = \mathcal{U}$ as

$$\mu_\theta(P^a, P^{-a}) = \begin{cases} 1 \text{ (stop)}, & \text{if } -\theta^{aT}P^a\theta^a + \underline{\theta}^{aT}\bar{P}^a\underline{\theta}^a + \sum_{l\neq a}\theta^{lT}P^l\theta^l - \underline{\theta}^{lT}\bar{P}^l\underline{\theta}^l \geq 1, \\ 2 \text{ (continue)}, & \text{otherwise.} \end{cases} \quad (20)$$

The following lemma states that the above parametrized policies satisfy the conclusion of Theorem 1 that the policies are monotone. The proof is straightforward and hence omitted.

Lemma 2: Consider each of the parametrized policies (17), (18), (19). Then $\theta^l, \underline{\theta}^l \in \Theta = \mathbb{R}_+^m$ is necessary and sufficient for the parametrized policy μ_θ to be monotone increasing in $P^a, \bar{P}^{-a}$ and decreasing in $P^{-a}, \bar{P}^a$. For (20), $\theta \in \Theta = \mathcal{U}$ (unit-norm vectors) is necessary and sufficient for the parametrized policy μ_θ to be monotone increasing in $P^a, \bar{P}^{-a}$ and decreasing in $P^{-a}, \bar{P}^a$.
∎

Lemma 2 says that since the constraints on the parameter vector θ are necessary and sufficient for a monotone policy, the classes of policies (17), (18), (19) and (20) do not leave out any monotone policies; nor do they include any non monotone policies. Therefore optimizing over Θ for each case yields the best approximation to the optimal policy within the appropriate class. *Remark*: Another example of a parametrized policy that satisfies Lemma 2 is obtained by replacing λ_X with $\log\det(X)$ in (17), (18), (19). In this case, the parameters $\theta^a, \underline{\theta}^a, \theta^l, \underline{\theta}^l$ are scalars. However, numerical studies (not presented here) show that this scalar parametrization is not rich enough to yield useful decision policies.

B. Stochastic Approximation Algorithm to estimate θ^*

Having characterized monotone parameterized policies above, our next goal is to compute the optimal parametrized policy μ_{θ^*} for the sequential detection problem described in Section III-B. This can be formulated as the following stochastic optimization problem:

$$J_{\mu_{\theta^*}} = \inf_{\theta\in\Theta} J_\theta(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}),$$

where $J_\theta(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}) = \mathbb{E}^{\mu_\theta}\{(\tau-1)c_\nu + \bar{C}(P^a_\tau, \bar{P}^a_\tau, P^{-a}_\tau, \bar{P}^{-a}_\tau|P_0 = P, \bar{P}_0 = \bar{P}\}.$ (21)

Recall that τ is the stopping time at which stop action $u = 1$ is applied, i.e. $\tau = \inf\{k : u_k = 1\}$.

The optimal parameter θ^* in (21) can be computed by simulation-based stochastic optimization algorithms as we now describe. Recall that for the first three examples above (namely, (17), (18) and (19)), there is the explicit constraint that θ^l and $\underline{\theta}^l \in \Theta = \mathbb{R}_+^m$. This constraint can be

eliminated straightforwardly by choosing each component of θ^l as $\theta^l(i) = \left[\phi^l(i)\right]^2$ where $\phi^l(i) \in \mathbb{R}$. The optimization problem (21) can then be formulated in terms of this new unconstrained parameter vector $\phi^l \in \mathbb{R}^m$.

In the fourth example above, namely (20), the parameter θ^l is constrained to the boundary set of the m-dimensional unit hypersphere $\mathcal{U}$. This constraint can be eliminated by parametrizing θ^l in terms of spherical coordinates ϕ as follows: Let

$$\theta^l(1) = \cos\phi^l(1), \ \theta^l(i) = \prod_{j=1}^{i-1} \sin\phi^l(j)\cos\phi^l(i), \ i = 2, \ldots, m-1, \quad \theta^l(m) = \prod_{j=1}^{m} \sin\phi^l(j). \tag{22}$$

where $\phi^l(i) \in \mathbb{R}$, $i = 1, \ldots, m$ denote a parametrization of θ. Then it is trivially verified that $\theta^l \in \mathcal{U}$. Again the optimization problem (21) can then be formulated in terms of this new unconstrained parameter vector $\phi^l \in \mathbb{R}^m$.

Algorithm 1 Policy Gradient Algorithm for computing optimal parametrized policy

---

Step 1: Choose initial threshold coefficients ϕ_0 and parametrized policy μ_{θ_0}.

Step 2: For iterations $n = 0, 1, 2, \ldots$

- Evaluate sample cost $\hat{J}_n(\mu_\phi) = (\tau - 1)c_\nu + \bar{C}(P_\tau^a, \bar{P}_\tau^a, P_\tau^{-a}, \bar{P}_\tau^{-a})$.
  Compute gradient estimate $\widehat{\nabla}_\phi \hat{J}_n(\mu_\phi)$ as:

$$\widehat{\nabla}_\phi J_n = \frac{J_n(\phi_n + \omega_n \mathbf{d}_n) - J_n(\phi_n - \omega_n \mathbf{d}_n)}{2\omega_n} \mathbf{d}_n, \quad \mathbf{d}_n(i) = \begin{cases} -1, & \text{with probability } 0.5, \\ +1, & \text{with probability } 0.5. \end{cases}$$

  Here $\omega_n = \frac{\omega}{(n+1)^\gamma}$ denotes the gradient step size with $0.5 \leq \gamma \leq 1$ and $\omega > 0$.
- Update threshold coefficients ϕ_n via (where ϵ_n below denotes step size)

$$\phi_{n+1} = \phi_n - \epsilon_{n+1} \widehat{\nabla}_\phi \hat{J}_n(\mu_\phi), \quad \epsilon_n = \epsilon/(n+1+s)^\zeta, \quad 0.5 < \zeta \leq 1, \text{ and } \epsilon, s > 0. \tag{23}$$

Several possible simulation based stochastic approximation algorithms can be used to estimate μ_{θ^*} in (21). In our numerical examples, we used Algorithm 1 to estimate the optimal parametrized policy. Algorithm 1 is a Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm [15]; see [16] for other more sophisticated gradient estimators. Algorithm 1 generates a sequence of estimates ϕ_n and thus θ_n, $n = 1, 2, \ldots$, that converges to a local minimum θ^* of (21) with policy $\mu_{\theta^*}(P)$. In Algorithm 1 we denote the policy as μ_ϕ since θ is parametrized in terms of ϕ as described above.

The SPSA algorithm [15] picks a single random direction $\mathbf{d}_n$ (see Step 2) along which the derivative is evaluated after each batch n. As is apparent from Step 2 of Algorithm 1, evaluation of the gradient estimate $\widehat{\nabla}_\phi J_n$ requires only 2 batch simulations. This is unlike the well known Kiefer-Wolfowitz stochastic approximation algorithm [15] where $2m$ batch simulations are required to evaluate the gradient estimate. Since the stochastic gradient algorithm (23) converges to a local optimum, it is necessary to retry with several distinct initial conditions.

V. APPLICATION: GMTI RADAR SCHEDULING AND NUMERICAL RESULTS

This section illustrates the performance of the monotone parametrized policy (21) computed via Algorithm 1 in a GMTI radar scheduling problem. We first show that the nonlinear measurement model of a GMTI tracker can be approximated satisfactorily by the linear Gaussian model (1) that was used above. Therefore the main result Theorem 1 applies, implying that the optimal radar micro-management decision policy is monotone. To illustrate these micro-management policies numerically, we then consider two important GMTI surveillance problems – the target fly-by problem and the persistent surveillance problem.

A. GMTI Kinematic Model and Justification of Linearized Model (1)

The observation model below is an abstraction based on approximating several underlying pre-processing steps. For example, given raw GMTI measurements, space-time adaptive processing (STAP) (which is a two-dimensional adaptive filter) is used for near real-time detection, see [6] and references therein. Similar observation models can be used as abstractions of synthetic aperture radar (SAR) based processing.

A modern GMTI radar manager operates on three time-scales (The description below is a simplified variant of an actual radar system.):

- Individual observations of target l are obtained on the fast time-scale $t = 1, 2, \ldots$. The period at which t ticks is typically 1 milli-second. At this time-scale, ground targets can be considered to be static.

- Decision epoch $k = 1, 2, \ldots, \tau$ is the time-scale at which the micro-manager and target tracker operate. Recall τ is the stopping time at which the micro-manager decides to stop and return control to the macro-manager. The clock-period at which k ticks is typically $T = 0.1$ seconds. At this epoch time-scale k, the targets move according to the kinematic model (24), (25) below. Each epoch k is comprised of intervals $t = 1, 2, \ldots, \Delta$ of the fast

time-scale, where Δ is typically of the order of 100. So, 100 observations are integrated at the t-time-scale to yield a single observation at the k-time-scale.

- The scheduling interval $n = 1, 2 \ldots$, is the time-scale at which the macro-manager operates. Each scheduling interval n is comprised of τ_n decision epochs. This stopping time τ_n is determined by the micro-manager. τ_n is typically in the range 10 to 50 – in absolute time it corresponds to the range 1 to 5 seconds. In such a time period, a ground target moving at 50 km per hour moves approximately in the range 14 to 70 meters.

1) GMTI Kinematic Model: The tracker assumes that each target $l \in \{1, \ldots, L\}$ has kinematic model and GMTI observations [9],

$$s_{k+1}^l = F s_k^l + G w_k^l, \tag{24}$$

$$z_k^l = \begin{cases} h(s_k^l, \xi_k) + \frac{1}{\sqrt{\nu^l \Delta}} v_k^l, & \text{with probability } p_d^l, \\ \emptyset, & \text{with probability } 1 - p_d^l. \end{cases} \tag{25}$$

Here z_k^l denotes a 3-dimensional (range, bearing and range rate) observation vector of target l at epoch time k and ξ_k denotes the Cartesian coordinates and speed of the platform (aircraft) on which the GMTI radar is mounted. The noise processes w_k^l and $v_k^l / \sqrt{\nu^l \Delta}$ are zero-mean Gaussian random vectors with covariance matrices Q^l and $R^l(\nu^l)$, respectively. The observation z_k in decision epoch k is the average of the $\nu^l \Delta$ measurements obtained at the fast time scale t. Thus the observation noise variance in (25) is scaled by the reciprocal of the target priority $\nu^l \Delta$. In (24), (25) for a GMTI system,

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix}, \quad R(\nu^l) = \frac{1}{\nu^l \Delta} \begin{bmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_a^2 & 0 \\ 0 & 0 & \sigma_{\dot{r}}^2 \end{bmatrix}, \tag{26}$$

$$Q = \begin{bmatrix} \frac{1}{4} T^4 \sigma_x^2 & \frac{1}{2} T^3 \sigma_x^2 & 0 & 0 \\ \frac{1}{2} T^3 \sigma_x^2 & T^2 \sigma_x^2 & 0 & 0 \\ 0 & 0 & \frac{1}{4} T^4 \sigma_y^2 & \frac{1}{2} T^3 \sigma_y^2 \\ 0 & 0 & \frac{1}{2} T^3 \sigma_y^2 & T^2 \sigma_y^2 \end{bmatrix}, \quad h(s, \xi) = \begin{bmatrix} \sqrt{(x - \xi_x)^2 + (y - \xi_y)^2 + \xi_z^2} \\ \arctan\left(\frac{y - \xi_y}{x - \xi_x}\right) \\ \frac{(x - \xi_x)(\dot{x} - \dot{\xi}_x) + (y - \xi_y)(\dot{y} - \dot{\xi}_y)}{\sqrt{(x - \xi_x)^2 + (y - \xi_y)^2 + \xi_z^2}} \end{bmatrix}.$$

Recall that T is typically 0.1 seconds. The elements of $h(s, \xi)$ correspond to range, azimuth, and range rate, respectively. Also $\xi = (\xi_x, \dot{\xi}_x, \xi_y, \dot{\xi}_y)$ denotes the x and y position and speeds, respectively, and ξ_z denotes the altitude, assumed to be constant, of the aircraft on which the GMTI radar is mounted.

2) Approximation by Linear Gaussian State Space Model: Starting with the nonlinear state space model (24), the aim below is to justify the use of the linearized model (1). We start with linearizing the model (24) as follows; see [17, Chapter 8.3]. For each target l, consider a nominal deterministic target trajectory $\bar{s}_k^l$ and nominal measurement $\bar{z}_k^l$ where $\bar{s}_{k+1}^l = F\bar{s}_k^l$, $\bar{z}_k^l = h(\bar{s}_k^l, \xi)$. Defining $\tilde{s}_k^l = s_k^l - \bar{s}_k^l$ and $\tilde{z}_k^l = z_k^l - \bar{z}_k^l$, a first order Taylor series expansion around this nominal trajectory yields,

$$\tilde{s}_{k+1}^l = F\tilde{s}_k^l + Gw_k^l,$$

$$\tilde{z}_k^l = \begin{cases} \nabla_s h(\bar{s}_k^l, \xi_k)\tilde{s}_k^l + R_1(s_k, \bar{s}_k^l, \xi_k) + \frac{1}{\sqrt{\nu^l \Delta}}v_k^l, & \text{with probability } p_d^l, \\ \emptyset, & \text{with probability } 1 - p_d^l, \end{cases} \tag{27}$$

where $\|R_1(s^l, \bar{s}^l, \xi)\| \leq \frac{1}{2}\|(s^l - \bar{s}^l)^T \nabla^2 h(\zeta, \xi)(s^l - \bar{s}^l)\|$ and $\zeta = \gamma s^l + (1 - \gamma)\bar{s}^l$ for some $\gamma \in [0, 1]$. In the above equation, $\nabla_s h(s, \xi)$ is the Jacobian matrix defined as (for simplicity we omit the superscript l for target l),

$$\nabla_s h(s, \xi)_{ij} = \begin{bmatrix} \frac{\delta_x}{r} & 0 & \frac{\delta_y}{r} & 0 \\ \frac{-\delta_y}{\delta_x^2 + \delta_y^2} & 0 & \frac{\delta_x}{\delta_x^2 + \delta_y^2} & 0 \\ \frac{\dot{\delta}_x}{r} - \frac{\delta_x\delta_y\dot{\delta}_y + \delta_x^2\dot{\delta}_x}{r^3} & \frac{\delta_x}{r} & \frac{\dot{\delta}_y}{r} - \frac{\delta_x\delta_y\dot{\delta}_x + \delta_y^2\dot{\delta}_y}{r^3} & \frac{\delta_y}{r} \end{bmatrix}, \quad r = \sqrt{\delta_x^2 + \delta_y^2 + \xi_z^2}. \tag{28}$$

where $\delta_x = x - \xi_x$, $\delta_y = y - \xi_y$ denotes the relative position of the target with respect to the platform and $\dot{\delta}_x$, $\dot{\delta}_y$ denote the relative velocities. Since the target is ground based and the platform is constant altitude, ξ_z is a constant.

In (27), $\nabla^2 h(\cdot, \cdot)$ denotes the $3 \times 4 \times 4$ Hessian tensor. By evaluating this Hessian tensor for typical operating modes and $k \leq 50$, we show below that

$$\frac{\|R_1(s^l, \bar{s}^l, \xi)\|}{\|\nabla_s h(\bar{s}^l, \xi)\tilde{s}^l\|} \leq 0.02, \quad \frac{\|\nabla_s h(\bar{s}_k^l, \xi_k) - \nabla_s h(\bar{s}_0^l, \xi_0)\|}{\|\nabla_s h(\bar{s}_k^l, \xi_k)\|} \leq 0.06. \tag{29}$$

The first inequality above says that the model is approximately linear in the sense that the ratio of linearization error $R_1(\cdot)$ to linear term is small; the second inequality says that the model is approximately time-invariant, in the sense that the relative magnitude of the error between linearizing around $\bar{s}_0$ and $\bar{s}_k$ is small. Therefore, on the micro-manager time scale, the target dynamics can be viewed as a linear time invariant state space model (1).

Justification of (29): Using typical GMTI operating parameters, we evaluate the bounds in (29). Denote the state of the platform (aircraft) on which the radar is situated as $p_0 = [p_{x,0}, \dot{p}_x, p_{y,0}, \dot{p}_y] = [-35000\text{m}, 100\text{m/s}, -15000\text{m}, 20\text{m/s}]$. Then the platform height is $p_z = \sqrt{p_{x,0}^2 + p_{y,0}^2}\tan\theta_d$, where θ_d is the depression angle, typically between $10°$ to $25°$. We assume

a depression angle of $\theta_d = 15°$ below yielding $p_z = 10203.2$m. Next, consider typical behaviour of ground targets with speed 15m/s (54 km/h) and select the following significantly different initial target state vectors (denoted by superscripts $a - e$)

$$s_0^a = \begin{bmatrix} 100 & 3 & 40 & 7 \end{bmatrix}, \ s_0^b = \begin{bmatrix} -20 & -4 & 200 & 1 \end{bmatrix}, \ s_0^c = \begin{bmatrix} 50 & 2 & 95 & 10 \end{bmatrix}, \quad (30)$$

$$s_0^d = \begin{bmatrix} -70 & 5 & -50 & -6 \end{bmatrix}, \ s_0^e = \begin{bmatrix} 150 & -15 & 10 & 0 \end{bmatrix}. \quad (31)$$

Now, propagate these initial states using the target model with $T = 0.1$s, $\sigma_x = \sigma_y = 0.5$, $\sigma_r = 20$m, $\sigma_{\dot{r}} = 5$m/s, $\sigma_a = 0.5°$ with a true track variability parameter $\sigma_p = 1.5$ (used for true track simulation as σ_x and σ_y). Define (see (29) and recall that $\zeta = \gamma s + (1 - \gamma)\bar{s}$ for some $\gamma \in [0,1]$)

$$D(\bar{s}_k, \xi_k) \equiv \frac{\|\nabla_s h(\bar{s}_k, \xi_k) - \nabla_s h(\bar{s}_0, \xi_0)\|}{\|\nabla_s h(\bar{s}_k, \xi_k)\|}, \quad E(s, \bar{s}, \xi, \gamma) = \frac{\frac{1}{2}\|(s - \bar{s})^T D^2 h(\zeta, \xi)(s - \bar{s})\|}{\|\nabla_s h(\bar{s}, \xi)(s - \bar{s})\|}.$$

Tables I to III show how $D(\cdot)$ and $E(\cdot)$ evolve with iteration $k = 10, 50, 100$. The entries in the tables are small, thereby justifying the linear time invariant state space model (1).

Remark: Since a linear Gaussian model is an accurate approximate model, most real GMTI trackers use an extended Kalman filter. Approximate nonlinear filtering methods such as sequential Markov Chain Monte-Carlo methods (particle filters) are not required.

B. Numerical Example 1: Target Fly-by

With the above justification of the model (1), we present the first numerical example. Consider $L = 4$ ground targets that are tracked by a GMTI platform, as illustrated in Figure 2. The nominal range from the GMTI sensor to the target region is approximately $\tilde{r} = 30$km. For this example, the initial (at the start of the micro-manager cycle) estimated and true target states of the four targets are given in Table IV.

We assume in this example that the most uncertain target is regarded as being the highest priority. Based on the initial states and estimates in Table IV, the mean square error values are, $\text{MSE}(\hat{s}_0^1) = 710.87$, $\text{MSE}(\hat{s}_0^2) = 222.16$, $\text{MSE}(\hat{s}_0^3) = 187.37$, and $\text{MSE}(\hat{s}_0^4) = 140.15$. Thus, target $l = 1$ is the most uncertain and allocated the highest priority. So we denote $a = 1$.

The simulation parameters are as follows: sampling time $T = 0.1$s (see Section V-A1); probability of detection $p_d = 0.75$ (for all targets, so superscript l is omitted); track standard deviations of target model $\sigma_x = \sigma_y = 0.5$m; measurement noise standard deviations $\sigma_r = 20$m, $\sigma_a = 0.5°$, $\sigma_{\dot{r}} = 5$m/s; and platform states $[p_x, \dot{p}_x, p_y, \dot{p}_y] = [10\text{km}, 53\text{m/s}, -30\text{km}, 85\text{m/s}]$. We assume a target priority vector of $\nu = [\nu^1, \nu^2, \nu^3, \nu^4] = [0.6, 0.39, 0.008, 0.002]$. Recall from

	$D(\bar{s}_{10}, \xi_{10})$	$D(\bar{s}_{50}, \xi_{50})$	$D(\bar{s}_{100}, \xi_{100})$
s_0^a	0.0010	0.0052	0.0104
s_0^b	0.0009	0.0049	0.0104
s_0^c	0.0010	0.0059	0.0119
s_0^d	0.0007	0.0040	0.0080
s_0^e	0.0010	0.0053	0.0112

TABLE I

RATE OF CHANGE OF JACOBIAN FOR VARIOUS RUNNING TIMES.

	$E(s_{10}, \bar{s}_{10}, \xi_{10}, 0.1)$	$E(s_{50}, \bar{s}_{50}, \xi_{50}, 0.1)$	$E(s_{100}, \bar{s}_{100}, \xi_{100}, 0.1)$
s_0^a	0.00019091	0.0010597	0.01395
s_0^b	0.00020866	0.0011699	0.014375
s_0^c	0.00019165	0.0010813	0.01453
s_0^d	0.0002008	0.0011065	0.011844
s_0^e	0.0002294	0.0012735	0.015946

TABLE II

RATIO OF SECOND-ORDER TO FIRST-ORDER TERM OF TAYLOR SERIES EXPANSION FOR $\alpha = 0.1$.

	$E(s_{10}, \bar{s}_{10}, \xi_{10}, 0.8)$	$E(s_{50}, \bar{s}_{50}, \xi_{50}, 0.8)$	$E(s_{100}, \bar{s}_{100}, \xi_{100}, 0.8)$
s_0^a	0.00019267	0.0011104	0.014211
s_0^b	0.00021104	0.0012164	0.014633
s_0^c	0.00019447	0.0011228	0.014838
s_0^d	0.00020148	0.0011386	0.012178
s_0^e	0.00023083	0.0013596	0.016603

TABLE III

RATIO OF SECOND-ORDER TO FIRST-ORDER TERM OF TAYLOR SERIES EXPANSION FOR $\alpha = 0.8$.

$$\hat{s}_0^1 = [130, 5.5, 84, 8.1]^{\mathrm{T}}, \qquad s_0^1 = [100, 3, 40, 7]^{\mathrm{T}}$$
$$\hat{s}_0^2 = [-47.88, -2.38, 210.41, 0.418]^{\mathrm{T}}, \quad s_0^2 = [-20, -4, 200, 1]^{\mathrm{T}}$$
$$\hat{s}_0^3 = [55.84, 2.37, 121.74, 9.56]^{\mathrm{T}}, \qquad s_0^3 = [50, 2, 95, 10]^{\mathrm{T}}$$
$$\hat{s}_0^4 = [-55.13, 5.75, -68.41, -6.10]^{\mathrm{T}}, \quad s_0^4 = [-70, 5, -50, -6]^{\mathrm{T}}$$

TABLE IV

INITIAL TARGET STATES AND ESTIMATES.

(25) that the target priority scales the inverse of the covariance of the observation noise. We chose an operating cost of $c_\nu = 0.8$, and the stopping cost of $\bar{C}(P^{-a})$ specified in (11), with constants $\alpha^{1,\dots,4} = \beta^{2,\dots,4} = 0.05$, $\beta^1 = 5$. The parametrized policy chosen for this example was $\mu_\theta(P^a, P^{-a})$ defined in (20). We used the SPSA algorithm (Algorithm 1) to estimate the parameter θ^* that optimizes the objective (21). Since the SPSA converges to a local minimum, several initial conditions were evaluated via a random search.

Figure 3 explores the sensitivity of the sample-path cost (achieved by the parametrized policy) with respect to probability of detection, p_d, and the operating cost, c_ν. The sample-path cost increases with c_ν and decreases with p_d. Larger values of the operating cost, c_ν, cause the radar micro-manager to specify the "stop" action sooner than for lower values of c_ν. As can be seen in the figure, neither the sample-path cost or the average stopping time is particularly sensitive to changes in the probability of detection. However, as expected, varying the operating cost has a large effect on both the sample-path cost and the associated average stopping time.

Figure 4 compares the optimal parametrized policy with periodic myopic policies. Such periodic myopic policies stop at a deterministic pre-specified time (without considering state information) and then return control to the macro-manager. The performance of the optimal parametrized policy is measured using multiple initial conditions. As seen in Figure 4, the optimal parametrized policy is the lower envelope of all possible periodic stopping times, for each initial condition. The optimal periodic policy is highly dependent upon the initial condition. The main performance advantage of the optimal parametrized policy is that it achieves virtually the same cost as the optimal periodic policy for any initial condition.

C. Numerical Example 2: Persistent Surveillance

As mentioned in Section I, persistent surveillance involves exhaustive surveillance of a region over long time intervals, typically over the period of several hours or weeks [18] and is useful in providing critical, long-term battlefield information. Figure 5 illustrates the persistent surveillance setup. Here $\tilde{r}$ is the nominal range from the target region to the GMTI platform, assumed in our simulations to be approximately 30km. The points on the GMTI platform track labeled (1) – (72) correspond to locations[4] where we evaluate the Jacobian (28). Assume a constant platform orbit speed of 250m/s (or approximately 900km/h [19]) and a constant altitude of approximately 5000m. Assuming 72 divisions along the 30km radius orbit, the platform sensor

---

[4]The platform state at location $n \in \{1, 2, ..., 72\}$ is defined as $p = [\tilde{r}\cos(n \cdot 5°), -\tilde{v}\sin(n \cdot 5°), \tilde{r}\sin(n \cdot 5°), \tilde{v}\cos(n \cdot 5°)]$
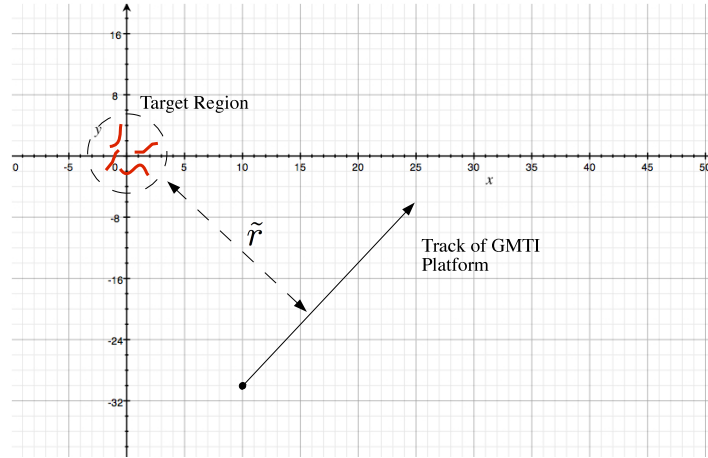
Fig. 2. Target Fly-by Scenario. The GMTI platform (aircraft) moves with constant altitude and velocity at nominal range $r = 30$ km from the target region. (r is defined in (28)). Initial states of the four targets are specified in Table IV.
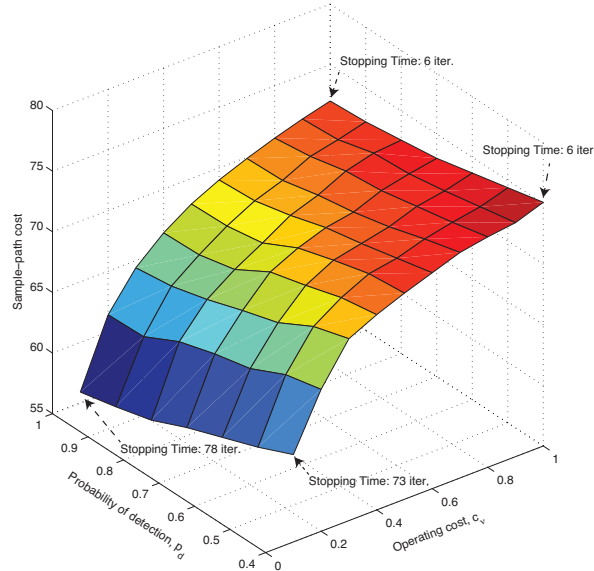

Fig. 3. Dependence of the sample-path cost achieved by the parametrized policy on the probability of detection, p_d, and the operating cost, c_ν. The sample-path cost increases with the operating cost, but decreases with the probability of detection. Note the stopping times associated with the labelled vertices above.
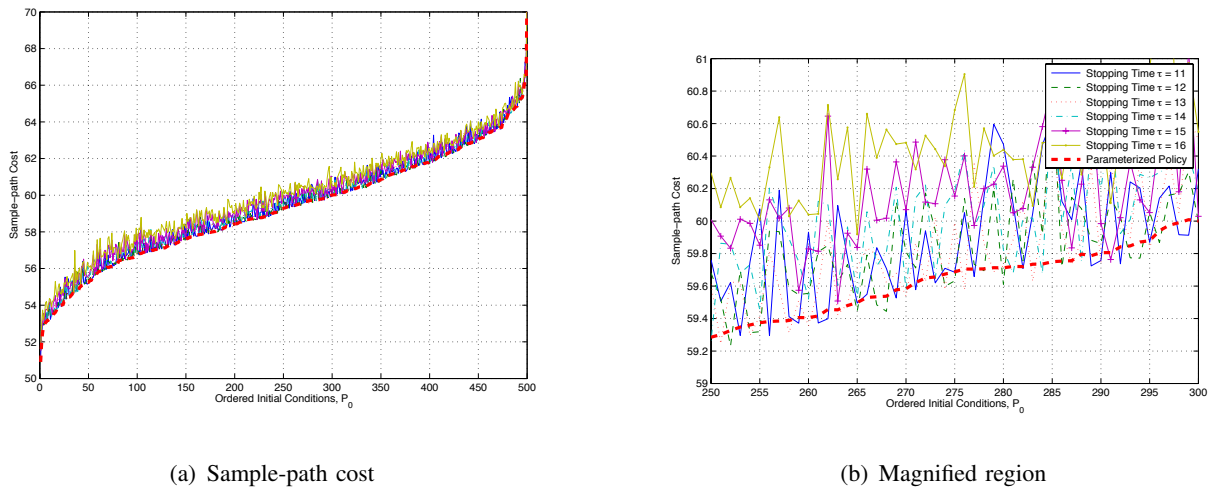
(a) Sample-path cost

(b) Magnified region

Fig. 4. Plot of sample-path cost of periodic policies and the parametrized policy (thick-dashed line) versus initial conditions. These initial conditions are ordered with respect to the cost achieved using the parametrized policy for that particular initial condition. Notice that the sample-path cost is the lower envelope of all deterministic stopping times for any initial condition.

takes 10.4 seconds to travel between the track segments. Using a similar analysis to the Appendix, the measurement model changes less than 5% in l_2-norm in 10.4s, thus the optimal parameter vector is approximately constant on each track segment.

Simulation parameters for this example are as follows: number of targets $L = 4$; sampling time $T = 0.1$s; probability of detection $p_d = 0.9$; track variances of target model $\sigma_x = \sigma_y = 0.5$m; and measurement noise parameters $\sigma_r = 20$m, $\sigma_a = 0.5°$, $\sigma_{\dot{r}} = 5$m/s. The platform velocity is now changing (assume a constant speed of $\tilde{v} = 250$m/s), unlike the previous example, which assumed a constant velocity platform. Since the linearized model will be different at each of the pre-specified points, (1)–(72), along the GMTI track, we computed the optimal parametrized policy at each of the respective locations. The radar manager then switches between these policies depending on the estimated position of the targets.

We consider Case 4 of Section III-A where the radar devotes all its resources to one target, and none to the other targets. That is, we assume a target priority vector of $\nu = [1, 0, 0, 0]$. In this case, the first target is allocated a Kalman filter, with all the other targets allocated measurement-free Kalman predictors. Since the threshold parametrization vectors depend on the target's state and measurement models, the first target $l = a$ has a unique parameter vector, where targets $l \neq a$ all have the same parameter vectors. Also, $\underline{\theta}^l = \theta$, for all $l \in \{1, 2, ..., L\}$.

We chose $\alpha^1 = 0.25$, $\alpha^2 = \alpha^3 = \alpha^4 = 0$, $\beta^1 = 0.25$, $\beta^2 = \beta^3 = \beta^4 = 1$ in stopping cost (11)

(average mutual information difference stopping cost). The parametrized policy considered was $\mu_\theta(P^a, P^{-a})$ in (20). The optimal parametrized policy was computed using Algorithm 1 at each of the 72 locations on the GMTI sensor track. As the GMTI platform orbits the target region, we switch between these parametrized policy vectors, thus continually changing the adopted tracking policy. We implemented the following macro-manager: $a = \arg\max_{l=1,\dots,L}\left\{\log|P^l|\right\}$. The priority vector was chosen as $\nu^a = 1$ and $\nu^l = 0$ for all $l \neq a$. Figure 6. shows log-determinants of each of the targets' error covariance matrices over multiple macro-management tracking cycles.

VI. CONCLUSIONS

This paper considers a sequential detection problem with mutual information stopping cost. Using lattice programming we prove that the optimal policy has a monotone structure in terms of the covariance estimates (Theorem 1). The proof involved showing monotonicity of the Riccati and Lyapunov equations (Theorem 2). Several examples of parametrized decision policies that satisfy this monotone structure were given. A simulation-based adaptive filtering algorithm (Algorithm (1)) was given to estimate the parametrized policy. The sequential detection problem was illustrated in a GMTI radar scheduling problem with numerical examples.

APPENDIX

This appendix presents the proof of the main result Theorem 1. Appendix A presents the value iteration algorithm and supermodularity that will be used as the basis of the inductive proof. The proof of Theorem 1 in Appendix B uses lattice programming [20] and depends on certain monotone properties of the Kalman filter Riccati and Lyapunov equations. These properties are proved in Theorem 2 in Appendix C.

A. Preliminaries

We first rewrite Bellman's equation (15) in a form that is suitable for our analysis. Define

$$C(P) = c_\nu - \bar{C}(P) + \sum_{z^a, z^{-a}} \bar{C}(\mathcal{R}(P^a, z^a), \mathcal{L}(P^a), \mathcal{R}(P^{-a}, z^{-a}), \mathcal{L}(P^{-a}))q_{z^a}q_{z^{-a}},$$

$$V(P) = \bar{V}(P) - \bar{C}(P), \quad \text{where } P = (P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}), \tag{32}$$

$$q_{z^l} = \begin{cases} p_d^l, & \text{if } z^l \neq \emptyset, \\ 1 - p_d^l, & \text{otherwise,} \end{cases}, l = 1, \dots, L.$$
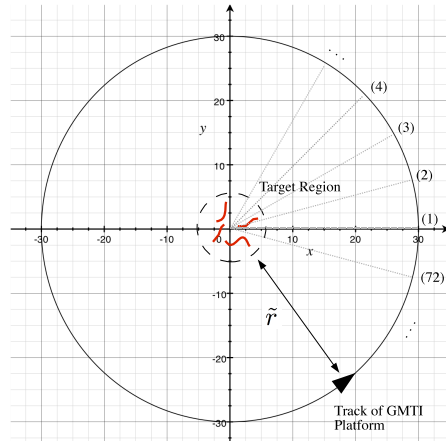
Fig. 5. Representation of the persistent surveillance scenario in GMTI systems. The GMTI platform (aircraft) orbits the target region in order to obtain persistent measurements as long as targets remain within the target region. The nominal range from the platform to the target region is assumed to be 30km.
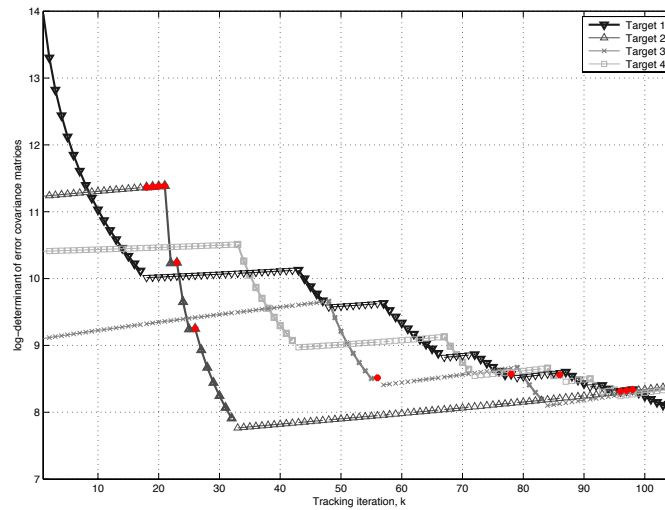

Fig. 6. Plot of log-determinants for each target over multiple scheduling intervals. On each scheduling interval, a Kalman filter is deployed to track one target and Kalman predictors track the remaining 3 targets. The bold line corresponds to the target allocated the Kalman filter by the micro-manager in each scheduling interval. Initially a Kalman filter is deployed on target $l = 1$. Data points marked in red indicate missing observations.

In (32) we have assumed that the missed observation events in (25) are statistically independent between targets, and so $q_{z^{-a}} = \prod_{l \neq a} q_{z^l}$. Actually, the results below hold for any joint distribution of missed observation events (and therefore allow these events to be dependent between targets). For notational convenience we assume (25) and independent missed observation events.

Clearly $V(\cdot)$ and optimal decision policy $\mu^*(\cdot)$ satisfy Bellman's equation

$$V(P) = \min\{\mathcal{Q}(P,1), \mathcal{Q}(P,2)\}, \quad \mu^*(P) = \underset{u \in \{1,2\}}{\arg\min}\{\mathcal{Q}(P,u)\}, \tag{33}$$

$$\mathcal{Q}(P,1) = 0, \quad \mathcal{Q}(P,2) = C(P) + \sum_{z^a, z^{-a}} V\left(\left(\mathcal{R}(P^a, z), \mathcal{L}(P^a), \mathcal{R}(P^{-a}, z^{-a}), \mathcal{L}(P^{-a})\right)\right) q_{z^a} q_{z^{-a}}.$$

Our goal is to characterize the stopping set defined as

$$\mathcal{S}_{\text{stop}} = \{P : \mathcal{Q}(P,1) \leq \mathcal{Q}(P,2)\} = \{P : \mathcal{Q}(P,2) \geq 0\} = \{P : \mu^*(P) = 1\}.$$

Since the $\arg\min$ function is translation invariant (that is, $\arg\min_u f(P,u) = \arg\min_u (f(P,u) + h(P))$ for any functions f and h), both the stopping set $\mathcal{S}_{\text{stop}}$ and optimal policy μ^* in these new coordinates are identical to those in the original coordinate system (16), (15).

Value Iteration Algorithm: The value iteration algorithm will be used to construct a proof of Theorem 1 by mathematical induction. Let $k = 1, 2, \ldots$, denote iteration number. The value iteration algorithm is a fixed point iteration of Bellman's equation and proceeds as follows:

$$V_0(P) = -\bar{C}(P), \quad V_{k+1}(P) = \min_{u \in \{1,2\}} \mathcal{Q}_{k+1}(P,u),$$

$$\mu_{k+1}^*(P) = \underset{u \in \{1,2\}}{\arg\min} \mathcal{Q}_{k+1}(P,u), \quad \text{where } \mathcal{Q}_{k+1}(P,1) = 0,$$

$$\mathcal{Q}_{k+1}(P,2) = C(P) + \sum_{z^a, z^{-a}} V_k\left(\mathcal{R}(P^a, z^a), \mathcal{R}(P^{-a}, z^{-a})\right) q_{z^a} q_{z^{-a}}. \tag{34}$$

Submodularity: Next, we define the key concept of submodularity [20]. While it can be defined on general lattices with an arbitrary partial order, here we restrict the definition to the posets $[\mathcal{M}, \succeq]$ and $[\mathcal{R}_+^m, \succeq_l]$, where the partial orders $\succeq$ and $\succeq_l$ were defined above.

Definition 1 (Submodularity and Supermodularity [20]): A scalar function $\mathcal{Q}(P, u)$ is submodular in P^a if

$$\mathcal{Q}(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, 2) - \mathcal{Q}(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, 1) \leq \mathcal{Q}(R^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, 2) - \mathcal{Q}(R^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, 1),$$

for $P^a \succeq R^a$. $\mathcal{Q}(\cdot, \cdot, u)$ is supermodular if $-\mathcal{Q}(\cdot, \cdot, u)$ is submodular. A scalar function $\mathcal{Q}(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, u)$ is sub/supermodular in each component of P^{-a} if it is sub/supermodular in each component P^l, $l \neq a$. An identical definition holds with respect to $\mathcal{Q}(\lambda^a, \lambda^{-a}, u)$ on $[\mathcal{R}_+^m, \succeq_l]$.

The most important feature of a supermodular (submodular) function $f(x,u)$ is that $\arg\min_u f(x,u)$ decreases (increases) in its argument x, see [20]. This is summarized in the following result.

Theorem 3 ([20]): Suppose $\mathcal{Q}(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, u)$ is submodular in P^a, submodular in $\bar{P}^{-a}$, supermodular in P^{-a} and supermodular in $\bar{P}^a$. Then there exists a $\mu^*(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}) = \arg\min_{u\in\{1,2\}} \mathcal{Q}(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a}, u)$, that is increasing in P^a, decreasing in $\bar{P}^a$, increasing in $\mathbb{P}^{-a}$ and decreasing in P^{-a}.

Next we state a well known result (see [21] for proof) that the evolution of the covariance matrix in the Lyapunov and Riccati equation are monotone.

Lemma 3 ([21]): $\mathcal{R}(\cdot)$ and $\mathcal{L}(\cdot)$ are monotone operators on the poset $[\mathcal{M}, \succeq]$. That is, if $P_1 \succeq P_2$, then $\mathcal{L}(P_1) \succeq \mathcal{L}(P_2)$ and for all z, $\mathcal{R}(P_1, z) \succeq \mathcal{R}(P_2, z)$.

Finally, we present the following lemma which states that the stopping costs (stochastic observability) are monotone in the covariance matrices. The proof of this lemma depends on Theorem 2, the proof of which is given in Appendix C below.

Lemma 4: For $\bar{C}(P)$ in Case 1 (9), Case 2 (10) and Case 3 (11), the cost $C(P^a, \bar{P}^a, P^{-a}, \bar{P}^{-a})$ defined in (32) is decreasing in P^a, $\bar{P}^{-a}$, and increasing in P^{-a}, $\bar{P}^a$. (Case 4 is a special case when $\alpha_l = 0$ for all $l \in \{1, \ldots, L\}$.)

Proof: For Case 1 and Case 2 let $l^* = \arg\max_{l\neq a} \left[\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|\right]$ or $l^* = \arg\min_{l\neq a} \left[\alpha^l \log |\bar{P}_k^l| - \beta^l \log |P_k^l|\right]$, respectively. From (32) with $|\cdot|$ denoting determinant,

$$C(P) = c_\nu - \alpha^a \log \frac{|\mathcal{L}(\bar{P}^a)|}{|\bar{P}^a|} + \beta^a \sum_{z^a} \log \frac{|\mathcal{R}(P^a, z^a)|}{|P^a|} q(z^a) + \alpha^{l^*} \log \frac{|\mathcal{L}(\bar{P}^{l^*})|}{|\bar{P}^{l^*}|} - \beta^{l^*} \log \frac{|\mathcal{R}(P^{l^*}, z^{l^*})|}{|P^{l^*}|} q(z^{l^*}).$$
(35)

For Case 3,

$$C(P) = c_\nu - \alpha^a \log \frac{|\mathcal{L}(\bar{P}^a)|}{|\bar{P}^a|} + \beta^a \sum_{z^a} \log \frac{|\mathcal{R}(P^a, z^a)|}{|P^a|} q(z^a) + \sum_{l\neq a} \alpha^l \log \frac{|\mathcal{L}(\bar{P}^l)|}{|\bar{P}^l|} - \sum_{l\neq a} \beta^l \log \frac{|\mathcal{R}(P^l, z^l)|}{|P^l|} q(z^l).$$
(36)

Theorem 2 shows that $\frac{|\mathcal{L}(\bar{P}^l)|}{|\bar{P}^l|}$ and $\frac{|\mathcal{R}(P^l, z^l)|}{|P^l|}$ are decreasing in $\bar{P}^l$ and P^l for all l. $\blacksquare$

B. Proof of Theorem 1

Proof: The proof is by induction on the value iteration algorithm (34). Note $V_0(P)$ defined in (34) is decreasing in P^a, $\bar{P}^{-a}$ and increasing in P^{-a}, $\bar{P}^a$ via Lemma 3.

Next assume $V_k(P)$ is decreasing in P^a, $\bar{P}^{-a}$ and increasing in P^{-a}, $\bar{P}^a$. Since $\mathcal{R}(P^a, y)$, $\mathcal{L}(\bar{P}^a)$, $\mathcal{R}(P^{-a}, y^{-a})$ and $\mathcal{L}(\bar{P}^{-a})$ are monotone increasing in P^a, $\bar{P}^a$, P^{-a} and $\bar{P}^{-a}$, it follows that the term $V_k\left(\mathcal{R}(P^a, z^a), \mathcal{L}(\bar{P}^a), \mathcal{R}(P^{-a}, z^{-a}), \mathcal{L}(\bar{P}^{-a})\right) q_{z^a} q_{z^{-a}}$ is decreasing in P^a, $\bar{P}^{-a}$ and

increasing in $P^{-a}, \bar{P}^a$ in (34). Next, it follows from Lemma 4 that $C(P^a, , \bar{P}^a, P^{-a}, \bar{P}^{-a})$ is decreasing in $P^a, \bar{P}^{-a}$ and increasing in $P^{-a}, \bar{P}^a$. Therefore from (34), $\mathcal{Q}_{k+1}(P, 2)$ inherits this property. Hence $V_{k+1}(P^a, P^{-a})$ is decreasing in $P^a, \bar{P}^{-a}$ and increasing in $P^{-a}, \bar{P}^a$. Since value iteration converges pointwise, i.e, $V_k(P)$ pointwise $V(P)$, it follows that $V(P)$ is decreasing in $P^a, \bar{P}^{-a}$ and increasing in $P^{-a}, \bar{P}^a$.

Therefore, $\mathcal{Q}(P, 2)$ is decreasing in $P^a, \bar{P}^{-a}$ and increasing in $P^{-a}, \bar{P}^a$. This implies $\mathcal{Q}(P, u)$ is submodular in (P^a, u), submodular in $(\bar{P}^{-a}, u)$, supermodular in (P^{-a}, u) and supermodular in $(\bar{P}^a, u)$. Therefore, from Theorem 3, there exists a version of $\mu^*(P)$ that is increasing in $P^a, \bar{P}^{-a}$ and decreasing in $P^{-a}, \bar{P}^a$. ■

C. Proof of Theorem 2

We start with the following lemma.

Lemma 5: If matrices X and Z are invertible, then for conformable matrices W and Y,

$$\det(Z)\det(X + YZ^{-1}W) = \det(X)\det(Z + WX^{-1}Y). \tag{37}$$

Proof: The Schur complement formulae applied to $\begin{pmatrix} X & Y \\ -W & Z \end{pmatrix}$ yields,

$$\begin{pmatrix} I & YZ^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X + YZ^{-1}W & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} I & 0 \\ -Z^{-1}W & I \end{pmatrix}$$
$$= \begin{pmatrix} I & 0 \\ -WX^{-1} & I \end{pmatrix} \begin{pmatrix} X & 0 \\ 0 & Z + WX^{-1}Y \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ 0 & I \end{pmatrix}.$$

Taking determinants yields (37). ■

Theorem 2(i): Given positive definite matrices Q and $P_1 \succ P_2$ and arbitrary matrix F,

$$\frac{\det(\mathcal{L}(P))}{\det(P)} \text{ is decreasing in } P, \text{ or equivalently, } \frac{\det(FP_1F^T + Q)}{\det(P_1)} < \frac{\det(FP_2F^T + Q)}{\det(P_2)}$$

Proof: Applying (37) with $[X, Y, W, Z] = [Q, F, F^T, P^{-1}]$,

$$\frac{\det(FPF^T + Q)}{\det(P)} = \det(P^{-1} + F^TQ^{-1}F)\det(Q). \tag{38}$$

Since $P_1 \succ P_2 \succ 0$, then $0 \prec P_1^{-1} \prec P_2^{-1}$ and thus $0 \prec P_1^{-1} + F^TQ^{-1}F \prec P_2^{-1} + F^TQ^{-1}F$. Since positive definite dominance implies dominance of determinants, it follows that

$$\det(P_1^{-1} + F^TQ^{-1}F) < \det(P_2^{-1} + F^TQ^{-1}F).$$

Using (38), the result follows. ■

Theorem 2(ii): Given positive definite matrices Q, R and arbitrary matrix F, $\frac{\det(\mathcal{R}(P,z))}{\det(P)}$ is decreasing in P. That is for $P_1 \succ P_2$

$$\frac{\det(FP_1F^T - FP_1H^T(HP_1H^T + R)^{-1}HP_1F^T + Q)}{\det(P_1)}$$
$$< \frac{\det(FP_2F^T - FP_2H^T(HP_2H^T + R)^{-1}HP_2F^T + Q)}{\det(P_2)} \quad (39)$$

Proof: Using the matrix inversion lemma $(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$,

$$(P^{-1} + H^T R^{-1}H)^{-1} = P - PH^T(HPH^T + R)^{-1}HP,$$

$$F(P^{-1} + H^T R^{-1}H)^{-1}F^T + Q = FPF^T - FPH^T(HPH^T + R)^{-1}HPF^T + Q$$

$$\det(FPF^T - FPH^T(HPH^T + R)^{-1}HPF^T + Q) = \det(Q + F(P^{-1} + H^T R^{-1}H)^{-1}F^T). \quad (40)$$

Applying the identity (37) with $[X, Y, W, Z] = [Q, F, F^T, P^{-1} + H^T R^{-1}H]$ we have,

$$\det(P^{-1} + H^T R^{-1}H)\det(Q + F(P^{-1} + H^T R^{-1}H)^{-1}F^T) = \det(Q)\det(P^{-1} + H^T R^{-1}H + F^T Q^{-1}F). \quad (41)$$

Further, using (37) with $[X, Y, W, Z] = [P^{-1}, H^T, H, R]$, we have,

$$\det(P^{-1} + H^T R^{-1}H) = \det(P^{-1})\det(R + HPH^T)/\det(R) \quad (42)$$

Substituting (42) into (41),

$$\det(P^{-1})\det(R + HPH^T)\det(Q + F(P^{-1} + H^T R^{-1}H)^{-1}F^T)$$
$$= \det(Q)\det(P^{-1} + H^T R^{-1}H + F^T Q^{-1}F)\det(R) \quad (43)$$

$$\frac{\det(Q + F(P^{-1} + H^T R^{-1}H)^{-1}F^T)}{\det(P)} = \frac{\det(Q)\det(P^{-1} + H^T R^{-1}H + F^T Q^{-1}F)\det(R)}{\det(R + HPH^T)}. \quad (44)$$

From (44) and (40)

$$\frac{\det(FPF^T - FPH^T(HPH^T + R)^{-1}HPF^T + Q)}{\det(P)} = \frac{\det(Q)\det(P^{-1} + H^T R^{-1}H + F^T Q^{-1}F)\det(R)}{\det(R + HPH^T)}. \quad (45)$$

We are now ready to prove the result. Since $P_1 \succ P_2 \succ 0$,

- $0 \prec P_1^{-1} + H^T R^{-1}H + F^T Q^{-1}F \prec P_2^{-1} + H^T R^{-1}H + F^T Q^{-1}F$,
- $\det(P_1^{-1} + H^T R^{-1}H + F^T Q^{-1}F) < \det(P_2^{-1} + H^T R^{-1}H + F^T Q^{-1}F)$,
- $R + HP_1H^T \succ R + HP_2H^T \succ 0$,

- $\det(R + HP_1H^T) > \det(R + HP_2H^T)$.

Therefore, (39) follows from the following inequality

$$\frac{\det(Q)\det(P_1^{-1} + H^T R^{-1} H + F^T Q^{-1} F)\det(R)}{\det(R + HP_1H^T)} < \frac{\det(Q)\det(P_2^{-1} + H^T R^{-1} H + F^T Q^{-1} F)\det(R)}{\det(R + HP_2H^T)}$$

∎

REFERENCES

[1] R.R. Mohler and C.S. Hwang, "Nonlinear data observability and information," *Journal of Franklin Institute*, vol. 325, no. 4, pp. 443–464, 1988.

[2] A. Logothetis and A. Isaksson, "On sensor scheduling via information theoretic criteria," in *Proc. American Control Conf.*, San Diego, 1999, pp. 2402–2406.

[3] A.R. Liu and R.R. Bitmead, "Stochastic observability in network state estimation and control," *Automatica*, vol. 47, pp. 65–78, 2011.

[4] E. Grossi and M. Lops, "MIMO radar waveform design: a divergence-based approach for sequential and fixed-sample size tests," in *3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009, pp. 165–168.

[5] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1 and 2, Athena Scientific, Belmont, Massachusetts, 2000.

[6] Bhashyam B., A. Damini, and K. Wang, "Persistent GMTI surveillance: theoretical performance bounds and some experimental results," in *Radar Sensor Technology XIV*, Orlando, Florida, 2010, SPIE.

[7] R. Whittle, "Gorgon stare broadens UAV surveillance," Aviation Week, Nov. 2010.

[8] D.P. Heyman and M.J. Sobel, *Stochastic Models in Operations Research*, vol. 2, McGraw-Hill, 1984.

[9] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.

[10] J. Wintenby and V. Krishnamurthy, "Hierarchical resource management in adaptive airborne surveillance radars–a stochastic discrete event system formulation," *IEEE Trans. Aerospace and Electronic Systems*, vol. 20, no. 2, pp. 401–420, April 2006.

[11] V. Krishnamurthy and D.V. Djonin, "Optimal threshold policies for multivariate POMDPs in radar resource management," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, 2009.

[12] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

[13] R. Evans, V. Krishnamurthy, and G. Nair, "Networked sensor management and data rate control for tracking maneuvering targets," *IEEE Trans. Signal Proc.*, vol. 53, no. 6, pp. 1979–1991, June 2005.

[14] O. Hernández-Lerma and J. Bernard Laserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.

[15] J. Spall, *Introduction to Stochastic Search and Optimization*, Wiley, 2003.

[16] G. Pflug, *Optimization of Stochastic Models: The Interface between Simulation and Optimization*, Kluwer Academic Publishers, 1996.

[17] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New Jersey, 1970.

[18] R.D. Rimey, W. Hoff, and J. Lee, "Recognizing wide-area and process-type activities," in *Information Fusion, 2007 10th International Conference on*, July 2007, pp. 1–8.

[19] U.S. Air Force, "E-8c joint stars factsheet," http://www.af.mil/information/factsheets/factsheet.asp?id=100, Sept 2007.

[20] D.M. Topkis, *Supermodularity and Complementarity*, Princeton University Press, 1998.

[21] B.D.O. Anderson and J.B. Moore, *Optimal filtering*, Prentice Hall, Englewood Cliffs, New Jersey, 1979.