Adaptive Search for Sparse Dynamic Targets

Gregory Newstadt, Dennis Wei, and Alfred O. Hero III,

University of Michigan, Dept. of Electrical Engineering and Computer Science, Ann Arbor, MI 48109, USA

Abstract—We consider the problem of energy constrained and noise-limited search for targets that are sparsely distributed over a large area. We propose a multiple-stage search algorithm that accounts for complex time-varying target behavior such as transitions among neighboring cells and varying target amplitudes. This work extends the adaptive resource allocation policy (ARAP) introduced in [Bashan et al, 2008] to policies with $T \gg 2$ stages. The proposed search strategy is driven by minimization of a surrogate function for energy constrained mean-squared error within locations containing targets. Exact optimization of the multi-stage objective function is infeasible, but myopic optimization yields a closed-form solution. We extend the myopic solution with non-myopic considerations that save a percentage of resources for exploring the scene at large. Empirical evidence suggests that the non-myopic policy performs significantly better than the myopic solution in terms of estimation error, probability of detection, and robustness to model mismatch. Moreover, the provided search policy has low computational complexity compared to state-of-the-art dynamic programming solutions.

I. INTRODUCTION

This work considers localizing and estimating moving targets in noise under resource constraints. For stationary targets, there has been significant work in adaptive localization and estimation [1]–[4], viewing the targets as a sparse signal. In these methods, past observations are used to shape future measurements of the scenes, which can result in stronger signal-to-noise-ratios (SNR).

In many applications such as wide-area surveillance, the targets of interest are not likely to remain stationary during the sensing period. Rather, they may exhibit complex dynamic behavior, such as movement, entering/leaving the scene, and obscuration. Dynamic programming (DP) [5] provides general methods for solving sensing problems with dynamic targets. However, exact DP methods generally are intractable when the state space is large as in the work considered here. Chong [6] shows that many adaptive sensing problems can be formulated as partially observable Markov decision processes (POMDP) and provides several approximate techniques. These POMDP approximate solutions allow for non-myopic sensing, but also suffer from large computational burdens for large state spaces.

In this paper, we provide a Bayesian formulation akin to [1] and [7], namely the Dynamic Adaptive Resource Allocation Policy (D-ARAP). This approach can simultaneously account for multiple targets as well as continuous allocation of sensing resources. D-ARAP has several interesting properties in comparison to alternative frameworks for allocating limited resources to localizing moving targets. First, it has significant performance gains over the baseline policy that uniformly allocates resources across the scene, in terms of estimation error and detection probability. Second, D-ARAP empirically

performs at a similar level to POMDP approximate solutions, albeit at a fraction of the computational cost. Finally, D-ARAP has increased robustness to noise and model mismatch as compared to myopic policies. The current paper improves upon our preliminary work [7] on a similar problem by further developing the optimization methods, comparing to gold-standard methods in POMDP approximation, and extending the performance analysis.

We formalize the problem in Section II and describe the adaptive sensing policy in Section III. Numerical performance analysis is given in Section IV. In Section V, we conclude and point to future work.

II. PROBLEM FORMULATION

We consider a space $\mathcal{X} = \{1, 2, \ldots, Q\}$ containing Q cells and a time-varying region of interest (ROI) $\Psi(t) \subset \mathcal{X}$, $t = 1, \ldots, T$. Let $I_i(t)$ be an indicator such that $I_i(t) = 1$ if i is in the ROI at time t ($i \in \Psi(t)$) and $I_i(t) = 0$ otherwise. We use a probabilistic signal model in which $I_i(1) = 1$ with prior probability $p_i(1)$, independently of the other indicators. For $I_i(t) = 0$, the corresponding signal amplitude $\theta_i(t)$ is taken to be zero, while for $I_i(t) = 1$, the amplitude $\theta_i(t)$ is modeled as a Gaussian random variable. The initial amplitudes $\theta_i(1)$, $i \in \Psi(1)$ are drawn independently with means $\mu_i(1)$ and variances $\sigma_i^2(1)$. As in previous work [1], [2], a uniform prior for targets is assumed with $p_i(1) = p_0$, $\mu_i(1) = \mu_0$ and $\sigma_i^2(1) = \sigma_0^2$ for all i.

We generalize previous work to include a dynamic target state model. To describe the model, we index the targets by target number instead of by cell: Let $s^{(n)}(t) \in \Psi(t)$, $n = 1, \ldots, |\Psi(t)|$ be the position of the n-th target at time t and $\vartheta^{(n)}(t) = \theta_{s^{(n)}(t)}(t)$ be its associated amplitude. Let α be the probability that each target is removed from the scene at each time. Conditioned on remaining in the scene, the target transition and amplitude model is

$$\Pr(s^{(n)}(t+1) = i | s^{(n)}(t) = j)$$
$$= \begin{cases} (1-\alpha)\pi_0, & i = j \\ \dfrac{(1-\alpha)(1-\pi_0)}{|G(j)|}, & i \in G(j) \end{cases} \quad (1)$$

$$\vartheta^{(n)}(t+1) = \vartheta^{(n)}(t) + \mathcal{N}(0, \Delta_\theta^2) \quad (2)$$

where π_0 is the probability that a target remains in the same location, and $G(j)$ is the set of cells that are neighbors of cell j. Note that we use a simple Gauss-Markov model [8] for the target amplitude, but this is easily generalizable to richer models. Moreover, while the target variance increases with t, the tracking variance does not. Let $B(t)$ be the event that a

single new target enters the scene at time t with probability β. Then conditioned on $B(t)$,

$$s^{(|\Psi(t)|+1)}(t+1)|B(t) \sim \text{Uniform}\{1, 2, \ldots, Q\},$$
$$\vartheta^{(|\Psi(t)|+1)}(t+1)|B(t) \sim \mathcal{N}(\mu_0, \sigma_0^2). \quad (3)$$

We restrict our attention to the case where at most one target occupies a cell at any instant. In the sparse situations considered here (i.e. $p_0 \ll 1$), this occurs with high probability.

Observations are made in T stages with effort levels $\lambda_i(t)$ that vary with location i and time t. In general, effort might be computing power, complexity, cost, or energy that is allocated to probing a particular cell. It is assumed that the quality of an observation increases with effort. Given $\lambda_i(t)$, the corresponding observation $y_i(t)$ takes the form

$$y_i(t) = \sqrt{\lambda_i(t)} I_i(t) \theta_i(t) + n_i(t), \quad t = 1, \ldots, T, \quad (4)$$

where $n_i(t) \sim \mathcal{N}(0, \sigma^2)$ is i.i.d. noise. The effort in each stage is constrained as $\sum_{i=1}^{Q} \lambda_i(t) \leq \Lambda(t)$.

In this paper, the goal is to estimate $\left\{ \Psi(t), \{\theta_i(t)\}_{i \in \Psi(t)} \right\}_{t=1,\ldots,T}$ over T stages. It follows that the objective function should depend on the posterior distribution of the target state vector given the measurements. When the targets are static, the posterior distribution factors by cell and can be exactly represented by the posterior mean/variance of the target amplitude and the posterior probability of a target existing in each cell. In the dynamic case, there is no simple factorization that allows for efficient exact estimation of the posterior distribution, partly due to the fact that the posterior distribution of the amplitudes becomes a Gaussian mixture distribution (due to nonzero transition probabilities to neighboring cells) rather than a simple univariate Gaussian. There are several signal processing algorithms that could be used to approximately estimate the posterior distribution, including particle filters, extended Kalman filters, and Unscented Kalman filters, with varying tradeoffs between accuracy and computational burden.

A computationally efficient approximate method is provided in [9] based on assumptions of sparsity (i.e. targets are well-separated) and approximations to Gaussian mixtures using a maximum likelihood approach. While full details are available in [9], here we just note that given these assumptions, the posterior distribution can be represented with the following:

$$\boldsymbol{x}(t) = \left\{ p_i(t), \mu_i(t), \sigma_i^2(t) \right\}_{i=1}^{Q} \quad (5)$$
$$p_i(t) = \Pr(I_i(t) = 1 | \boldsymbol{Y}(t-1)), \quad (6)$$
$$\mu_i(t) = \mathbb{E}[\theta_i(t) | I_i(t) = 1, \boldsymbol{Y}(t-1)], \quad (7)$$
$$\sigma_i^2(t) = \text{var}[\theta_i(t) | I_i(t) = 1, \boldsymbol{Y}(t-1)]. \quad (8)$$

where $\boldsymbol{Y}(T) = \{y_i(t)\}_{i,t=1,\ldots,T}$.

III. SEARCH POLICIES

This work considers a time-separable objective function as follows:

$$J(\boldsymbol{\lambda}; T) = \mathbb{E}\left[\sum_{t=1}^{T} \gamma(t) \sum_{i=1}^{Q} \frac{p_i(t)}{\sigma^2/\sigma_i^2(t) + \lambda_i(t)} \right], \quad (9)$$

where $\{\gamma(t)\}_{t=1}^{T}$ is a set of known weights on different stages. When the stage weights are given by $\gamma(T) = 1$ and $\gamma(t) = 0$ for $t < T$, this cost function is exactly the MSE in estimating $\{\theta_i(T)\}_{i \in \Psi(T)}$ in two cases: (a) when targets are static; and (b), when the target locations may change but are known exactly (i.e $p_i(T) = I_i(T)$). In practical cases when the ROI is not known precisely, we use (9) as a surrogate objective.

The optimization problem can be stated as $\{\hat{\lambda}_i(t)\}_{i,t} = \arg\min_{\boldsymbol{\lambda}} J(\boldsymbol{\lambda}; T)$ where $\sum_{i=1}^{Q} \lambda_i(t) \leq \Lambda(t)$ for $t = 1, 2, \ldots, T$. As discussed in previous work on a similar problem [2], it is possible in principle to use dynamic programming to find the optimal solution. However, for $T > 2$, this exact solution is computationally intractable due to the large sizes of the state $\boldsymbol{x}(t)$ and the action space $\boldsymbol{\lambda}(t)$. As an alternative, one may consider the myopic solution which optimizes

$$M(\boldsymbol{\lambda}; t) = \sum_{i=1}^{Q} \frac{p_i(t)}{\sigma^2/\sigma_i^2(t) + \lambda_i(t)}, \quad \sum_{i=1}^{Q} \lambda_i(t) \leq \Lambda(t). \quad (10)$$

The optimal solution as given in [2] begins by defining π to be an index permutation that sorts the quantities $\sqrt{p_i(t)}\sigma_i^2(t)$ in non-increasing order:

$$\sqrt{p_{\pi(1)}(t)}\sigma_{\pi(1)}^2(t) \geq \cdots \geq \sqrt{p_{\pi(Q)}(t)}\sigma_{\pi(Q)}^2(t). \quad (11)$$

Let $c_i(t) = \sigma^2/\sigma_i^2(t)$. Then define $g(k)$ to be the monotonically non-decreasing function of $k = 0, 1, \ldots, Q$ with $g(0) = 0$, $g(Q) = \infty$, and

$$g(k) = \frac{c_{\pi(k+1)}(t)}{\sqrt{p_{\pi(k+1)}(t)}} \sum_{i=1}^{k} \sqrt{p_{\pi(i)}(t)} - \sum_{i=1}^{k} c_{\pi(i)}(t) \quad (12)$$

for $k = 1, \ldots, Q - 1$. Then the solution to (10) is

$$\lambda_i^m(t) = \left(\Lambda(t) + \sum_{j=1}^{k^*} c_{\pi(j)}(t) \right) \frac{\sqrt{p_{\pi(i)}(t)}}{\sum_{j=1}^{k^*} \sqrt{p_{\pi(j)}(t)}} - c_{\pi(i)}(t), \quad (13)$$

for $i = 1, \ldots, k^*$ and $\lambda_i^m(t) = 0$ else. The number of nonzero components is determined by the interval $(g(k-1), g(k)]$ to which the budget parameter $\Lambda(t)$ belongs. Since $g(k)$ is monotonic, the mapping from $\Lambda(t)$ to k^* is well-defined.

Myopic policies can incur several drawbacks by being overly aggressive in the allocation of resources. This may lead to missed or lost targets as well as a lack of robustness to model mismatch. Chong [6] shows that there are significant gains to be had by using non-myopic policies, which trade off short-term performance gains for long-term benefits.

We propose a simple improvement to the myopic policy that uses a convex combination between exploitation of the current belief state and exploration of the scene at large. The proposed non-myopic allocation policy is termed the Dynamic Adaptive Resource Allocation Policy, or D-ARAP, and defined by

$$\lambda_i^{DARAP}(t; \kappa) = [\kappa(t)]\lambda^{uni}(t) + [1 - \kappa(t)]\lambda_i^m(t), \quad (14)$$

where $\kappa(t) \in [0, 1]$, $\lambda^{uni}(t) = \Lambda(t)/Q$ is the uniform allocation policy, and $\lambda_i^m(t)$ is given by (13).

```
procedure {κ(τ)}ᵀ_{τ=1} = NestedPolicy
    Set κ(1) = 1, κ̃(1) = ∅, κ̃(2) = {0}.
    for τ = 3, . . . , T do
        for each κ(2) ∈ [0, 1] do
            Set κ(τ) = {κ(1), κ(2), κ̃(τ − 1)}.
            Calculate C(κ(2)) = J(λ^{DARAP}(κ(τ)); τ).
        end for
        Choose κ̂(2) = arg min C(κ(2)).
                         κ(2)
        Set κ̃(τ) = {κ̂(2), κ̃(τ − 1)}.
    end for
    Return {κ(τ) = {κ(1), κ̃(τ)}}ᵀ_{τ=1}.
end procedure
```

Fig. 1. Nested policy pseudocode

(a) Nested policy (vs. SNR)   (b) Heuristic policy (vs. SNR)

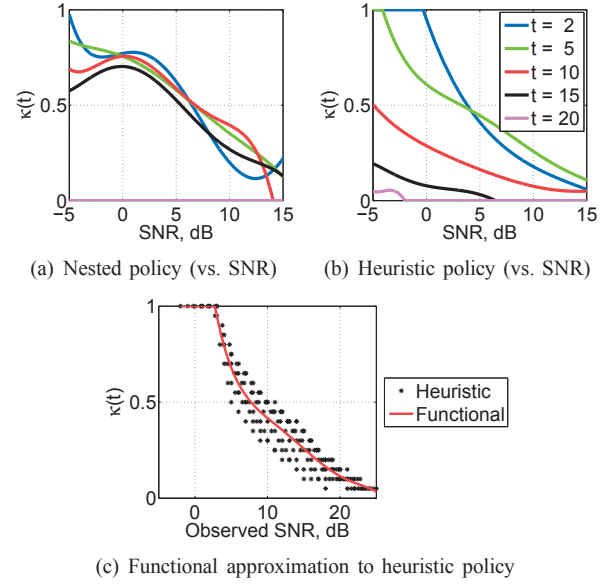(c) Functional approximation to heuristic policy

Fig. 2. Parameters $\kappa(t)$ selected according to the (a) nested and (b) heuristic strategies for policies of $T = 20$ stages, plotted against SNR per stage. In (c), a plot of the heuristic policy parameters versus cumulative SNR motivates a functional approximation to the heuristic strategy, shown by the red line.

Without prior knowledge on the location of targets, the first stage should be purely exploratory, i.e., $\kappa(1) = 1$. In addition, since the last stage should be purely exploitative or myopic, $\kappa(T) = 0$. To determine the full set of parameters $\boldsymbol{\kappa}(T) = \{\kappa(t)\}_{t=1}^T$, we consider offline policies, which are determined prior to collecting observations, and online policies, which are determined adaptively as measurements are collected. Note that $\lambda_i^m(t)$ is a function of previous measurements. Thus, the offline policies can still be adaptive as long as $\kappa(t) < 1$.

We first describe an offline policy termed the "nested policy" that is recursive in the sense that a T-stage policy is created by building upon a previously defined $(T-1)$-stage policy. The pseudocode for the nested policy is given in Fig. 1 and yields policies for $\{\boldsymbol{\kappa}(\tau)\}_{\tau=1}^T$ from 1 to T stages inclusive. In each iteration, a τ-stage policy is constructed by searching over a single parameter $\kappa(2)$ followed by using the previously defined $(\tau-1)$-stage policy. The values of $\kappa(2)$ are chosen to minimize the full non-myopic cost in (9).

The nested approach requires $\mathcal{O}(T^2)$ Monte Carlo simulations to determine policies for $\{\boldsymbol{\kappa}(\tau)\}_{\tau=1}^T$. To lessen this computational burden, we consider a "heuristic policy" which requires $\mathcal{O}(T)$ simulations. For $t = 2, 3, \ldots, T$, $\hat{\kappa}(t)$ is determined by the following $(1+\rho)$-optimality criterion:

$$\hat{\kappa}(t) = \max_{\kappa} \left\{ \kappa : \mathbb{E}[M^D(\kappa; t)] < (1+\rho)\mathbb{E}[M^D(0; t)] \right\},$$
(15)

where $M^D(\kappa; t)$ is the myopic cost (10) using $\boldsymbol{\lambda}^{DARAP}(t; \kappa)$, $\rho > 0$ is a tolerance, and expectations are taken over $\boldsymbol{Y}(t)$ through Monte Carlo simulations. Since $\kappa(t) = 0$ optimizes the myopic cost by definition, (15) results in a policy that is within $(1+\rho)$ of the expected minimum myopic cost at each stage. The overall policy is then given by $\boldsymbol{\kappa}(T) = \{\kappa(1) = 1, \hat{\kappa}(2), \ldots, \hat{\kappa}(T-1), \kappa(T) = 0\}$.

The nested and heuristic policy parameters are shown in Fig. 2 for various values of SNR[1] and $T = 20$. For $t > 2$, the heuristic policy parameters are nearly monotonically decreasing in both SNR and stage t. This motivates the "functional"

policy (Fig. 2(c)), which uses a polynomial fit to the heuristic policy parameters as a function of cumulative SNR[2].

We also consider online policies, which select $\kappa(t)$ independently for each instantiation of the problem. Generally this requires significantly more computation but may lead to better results in some cases. We propose using a T_0-stage rollout policy [5], which requires a known base policy. In each stage t, the rollout policy minimizes the total cost over the next T_0 stages by searching over $\kappa(t)$ and then using the base policy for the $T_0 - 1$ remaining stages. Similar to the nested/heuristic policies, minimization is done through Monte Carlo estimation of the cost function and a line search over $\kappa(t)$.

IV. PERFORMANCE ANALYSIS

We use numerical comparisons to quantify the performance gains and differences among the proposed policies. In these simulations, we let $Q = 10^4$, $p_0 = 10^{-3}$, $\mu_0 = 1$, $\sigma_0 = 1/6$, $\pi_0 = 0.4$, $G = 2$, $\alpha = \beta = 0$, $\Delta_\theta = 1/20$, $\sigma^2 = 1$, $\gamma(t) = (0.9)^{T-t}$ in (9), and $\rho = 0.1$ for the heuristic D-ARAP policy. We start by comparing performance of the proposed policies for various values of SNR and policy lengths T. Performance is compared against (a) a non-adaptive uniform policy which evenly allocates resources across the scene and (b) an oracle policy that knows the immediately previous locations $\Psi(t-1)$ when planning for stage t.

Figs. 3 (a) and (b) show the MSE gains (with respect to a uniform policy) in estimating $\{\theta_i(t)\}_{i \in \Psi(t)}$ as a function of stage t. The nested policy generally has the highest gains in MSE among non-oracle policies. Figs. 3 (c) and (d) show the probability of detection for a fixed probability of false alarm ($P_{fa} = 10^{-4}$) as a function of t. Note that the probabilities

---

[1]SNR is defined in terms of the budget per stage $\Lambda(t)$ and the noise variance σ^2 as $\text{SNR}(\Lambda(t)) = 10 \log_{10}(\Lambda(t)/(Q\sigma^2))$.

[2]Cumulative SNR is defined as $\text{SNR}(\sum_{\tau=1}^t \Lambda(\tau))$.

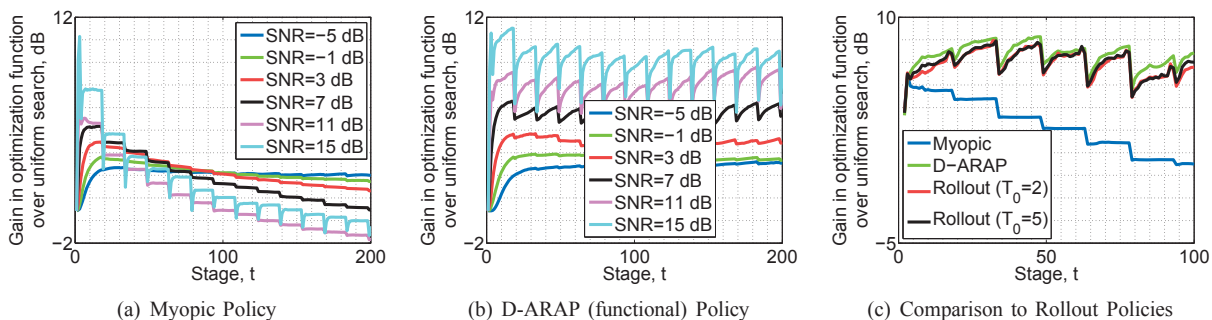(a) Myopic Policy   (b) D-ARAP (functional) Policy   (c) Comparison to Rollout Policies

Fig. 4. Gains (dB) in non-myopic cost (9) with respect to a uniform policy that evenly allocates resources everywhere. In this scenario, faulty measurements every 15 stages cause corresponding drops in performance. The myopic policy shown in (a) misses targets, causing performance to trend downwards as t and/or SNR increase. On the other hand, the D-ARAP (functional) policy shown in (b) shows robustness to the faulty measurements. In (c), the performance is compared across the myopic/functional/rollout policies for SNR = 10 dB. The performance of the rollout policies closely parallels the performance of D-ARAP, although the former incur a much larger computational burden.
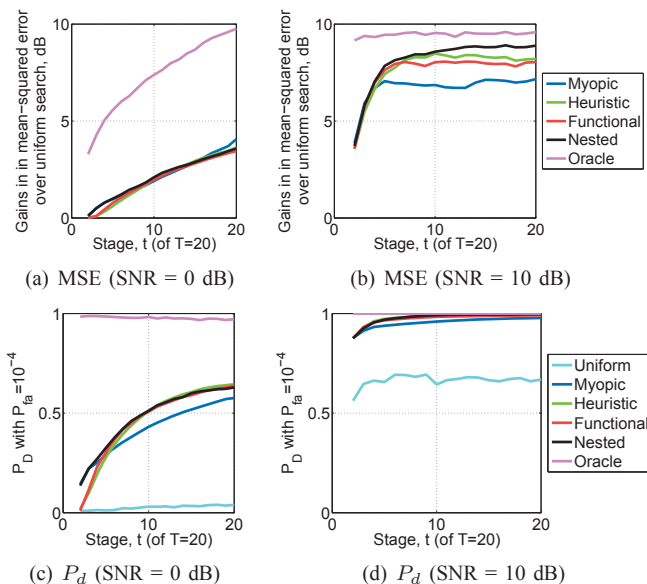

(a) MSE (SNR = 0 dB)   (b) MSE (SNR = 10 dB)

(c) P_d (SNR = 0 dB)   (d) P_d (SNR = 10 dB)

Fig. 3. Comparison of estimation and detection performance in low and high SNR scenarios. In (a)/(b), gains in MSE with respect to a uniform allocation policy are plotted on a dB scale as a function of t. The nested policy generally results in the highest gains among non-oracle policies. In (c)/(d), probabilities of target detection are plotted as a function of t for fixed probability of false alarm $P_{fa} = 10^{-4}$. The detection probabilities for D-ARAP (nested, heuristic, functional) approach 1 faster than for the myopic/uniform policies.

of detection for D-ARAP (nested, heuristic, and functional policies) consistently increase as t gets large and do so faster than for the uniform and myopic policies.

In the next simulation, we test the policies when the sensor has faults every 15 stages for a randomly selected subset of targets, causing highly noisy measurements for those targets. Figure 4 shows gains in terms of the non-myopic cost (9) with respect to the uniform policy for the myopic, functional D-ARAP, and rollout policies ($T_0 = 2$ and $T_0 = 5$). The performance of the myopic policy trends downwards as either SNR or t increases due to missed or lost targets. On the other hand D-ARAP, which always allocates some resources to exploration, provides robustness to the noisy measurements and maintains positive gains over a uniform allocation policy.

The rollout policies are compared only for SNR = 10 dB due to computational constraints. They have very similar performance to D-ARAP although at a much larger computational cost.

V. DISCUSSION AND FUTURE WORK

In this paper, we have proposed a framework, namely D-ARAP, that significantly extends previous work [1] to incorporate moving targets and long horizons, while providing computationally tractable solutions as compared to standard dynamic programming approximations. Non-myopic policies are presented which show robustness over myopic alternatives and significant gains over non-adaptive strategies. Future research paths include deriving analytical results such as convergence rates to the ROI and/or minimum detectable amplitudes. We would also like to continue investigating online policies that are computed as measurements are collected.

REFERENCES

[1] E. Bashan, R. Raich, and A. Hero, "Optimal Two-Stage Search for Sparse Targets Using Convex Criteria," *IEEE Transactions on Signal Processing*, vol. 56, pp. 5389–5402, 2008.

[2] D. Wei and A. O. Hero, "Multistage adaptive estimation of sparse signals," *IEEE Journal of Selected Topics in Signal Processing*, 2013, in press.

[3] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6222–6235, 2011.

[4] D. A. Castanon, "Optimal search strategies in dynamic hypothesis testing," *IEEE Transaction on Systems, Man and Cybernetics*, vol. 25, no. 7, pp. 1130–1138, July 1995.

[5] D. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Nashua, NH: Athena Scientific, 2005, vol. 1.

[6] E. Chong, C. Kreucher, and A. Hero, "Monte-Carlo-based partially observable Markov decision process approximations for adaptive sensing," in *Discrete Event Systems, 2008. WODES 2008. 9th International Workshop on*. IEEE, 2008, pp. 173–180.

[7] G. Newstadt, E. Bashan, and A. Hero, "Adaptive search for dynamic targets under resource constraints," in *Signals, Systems and Computers, Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 1235–1239.

[8] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[9] G. E. Newstadt, "Adaptive sensing techniques for dynamic target tracking and detection with applications to synthetic aperture radars," Ph.D. dissertation, University of Michigan, May 2013.