Implementing Perception-Action Cycles using Stochastic Optimization

Alexander Charlish
Fraunhofer FKIE
Wachtberg, Germany
alexander.charlish@ieee.org

Kristine Bell
Metron, Inc.
Reston, VA, USA
kristine.bell@ieee.org

Chris Kreucher
Centauri
Ann Arbor, MI, USA
ckreuche@umich.edu

***Abstract*—Cognitive radar problems involve the selection of actions based on the uncertain knowledge of a system state that is partially observed through noisy measurements. This process of sequential decision making under uncertainty can be considered as a stochastic optimization problem. This paper explicitly makes the connection between cognitive radar and stochastic optimization by presenting a framework for describing cognitive radar problems in terms of stochastic optimization, thereby pointing to ways to employ stochastic optimization for designing perception-action cycles in a cognitive radar.**

***Index Terms*—Cognitive radar, stochastic optimization, sensor management, POMDP**

I. Introduction

Cognitive radar problems require the selection of actions based on an uncertain perception that is obtained through inexact measurements. There is a broad variety of different cognitive radar problems that differ in terms of input data and the types of actions selected, for example, waveform selection and optimization, measurement scheduling, resource management, detection, tracking and imaging [1]. A single radar may in fact comprise several individual perception-actions cycles, spread over multiple information abstraction levels [2]. Despite their differences, the variety of cognitive radar problems can be described in terms of a set of similar problem components. Consequently, after identifying the problem components, similar methodologies can be applied for designing perception-action cycles for a cognitive radar.

Cognitive radar problems can be classed as types of stochastic optimization problems. Stochastic optimization is a broad term for techniques that perform decision making under uncertainty, which are currently widely deployed in a range of applications including finance, business, logistics and transportation, and science and engineering. Stochastic optimization methods seek a policy that exploits models to map from all available information at the current time into an optimized action. As this policy is essentially a perception-action cycle, the design of perception-actions for cognitive radar can benefit from applying algorithmic strategies for finding policies from the stochastic optimization field.

There are many communities focusing on stochastic optimization problems, who have established a wide variety of algorithmic solutions. These stochastic optimization communities cover techniques and applications such as decision trees, stochastic search, optimal stopping, optimal control, (partially observable) Markov decision processes (MDPs/POMDPs), approximate dynamic programming, reinforcement learning, model predictive control, stochastic programming, ranking and selection, and multiarmed bandit problems. It has been shown [3] that these problems can be described in a single stochastic optimization framework, and the respective solution methodologies can be grouped into just four classes.

The primary contribution of this paper is to directly connect the cognitive radar problem with the large body of work done in stochastic optimization. This connection makes the methodologies developed in the stochastic optimization communities directly applicable to the cognitive radar problem, promising to lead to improved methods for designing perception-action cycles in cognitive radar.

II. Stochastic Optimization Framework

This section presents a framework, inspired by [3] [1], which enables cognitive radar problems to be described in terms of stochastic optimization problems.

A. General Problem Components

All the problems addressed by the stochastic optimization communities comprise the problem components described in this subsection. The next subsection and the remainder of the paper focus on partially observable problems that are more relevant to cognitive radar.

System State - We are interested in the state of a dynamic system, which can be modelled as a random vector $\boldsymbol{X}_k$ for decision step k with realisation $\mathbf{x}_k \in \mathcal{X}$ where $\mathcal{X}$ is the system state space.

Actions and Action Space - We can select an action or actions at each decision step k, which influences the transition of the system state between time step k and $k+1$. The realisation of an action for decision step k is denoted $\mathbf{a}_k \in \mathcal{A}$ where $\mathcal{A}$ is the action space.

Exogenous Information - Additional information is revealed at each sequential decision step. The information revealed at each time step is modelled as a random vector $\boldsymbol{Z}_k$

[1] Although we adopt the framework in [3], we use the terminology and notation that is established in the signal processing community.

978-1-7281-8942-0/20/$31.00 2020 IEEE

with realisation $\mathbf{z}_k$. For completely observable problems the exogenous information is the system state.

State Transition Function - Between decision steps the system evolves according to a transition function $\mathbf{x}_{k+1} = f_X(\mathbf{x}_k, \mathbf{a}_k, \mathbf{w}_k)$, where $\mathbf{w}_k$ is a realisation of the state transition noise (alternatively termed process noise). Due to the state transition noise, the transition can be described by the transition probability density $p(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{a}_k)$.

Objective Function - At each decision step a reward or cost is encountered, which is described by the function $r_x(\mathbf{x}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$.

These common components allow the breadth of stochastic optimization problems considered by the stochastic optimization communities to be described.

B. Partial Observability

A common aspect of cognitive radar problems is that the system state is only *partially observable* through noisy measurements. Therefore, uncertainty is present not only in uncertain state transitions but also through uncertain measurements. Consequently, we extend and adapt the components described Section II-A to the more specific partially observable case, which results in a framework closely resembling a POMDP.

Measurements and Measurement Space - The exogenous information described in Section II-A can now be thought of as a noisy measurement of the system state. Now, the random vector $\boldsymbol{Z}_k$ can be defined more exactly as a measurement with realisation $\mathbf{z}_k \in \mathcal{Z}$ where $\mathcal{Z}$ is the measurement space.

Measurement Likelihood Function - Measurements are related to the system state through the measurement function $\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{a}_k, \mathbf{v}_k)$ where $\mathbf{v}_k$ is a realisation of the measurement noise. Due to the measurement noise, the measurement process can be described by the measurement likelihood function $\mathcal{L}(\mathbf{x}_k|\mathbf{z}_k, \mathbf{a}_k) \equiv p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{a}_k)$.

Information State - As the state of the system is not observable, it is necessary to decide on an action based on the information state. The information state is the set of actions and measurements that have occurred prior to the current decision step. The information state for decision step k is denoted $\mathcal{I}_k = (\mathbf{a}_0, \mathbf{z}_1, ..., \mathbf{a}_{k-1}, \mathbf{z}_k)$. This information state grows with each time step, i.e., $\mathcal{I}_k = \mathcal{I}_{k-1} \cup (\mathbf{a}_{k-1}, \mathbf{z}_k)$.

Belief State - As the cardinality of the information state grows with each time step, it is generally undesirable to be used as the perception upon which actions are decided. Instead, decisions can be based on a belief state. The belief state is a set of parameters with fixed cardinality that are an (ideally sufficient) statistic of the information state. The belief state at decision step k is modelled as a random vector $\boldsymbol{B}_k$ with realisation $\mathbf{b}_k$. For example, under linear Gaussian assumptions a sufficient statistic of the information state is the mean and covariance of the posterior PDF, i.e. $p(\mathbf{x}_k|\mathcal{I}_k) \equiv p(\mathbf{x}_k|\mathbf{b}_k)$. Typical belief states are parameters of a Gaussian, a Gaussian sum, or a set of particles.

Belief State Transition Function - It is necessary to define a transition function for belief states, analog to the system state transition function. This transition function is denoted $\mathbf{b}_{k+1} = f_B(\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$. As the belief state can be thought of as parameters of the posterior PDF $p(\mathbf{x}_k|\mathbf{b}_k)$, the transition function represents the standard Bayesian prediction and update steps. As a cognitive radar is an observer, it is often the case that the system state transition is not influenced by the selected sensing action. However, the belief state transition certainly will be influenced by the selected action.

Objective Function - An objective function is now defined as a function of the belief state, i.e. $r(\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$. This differs from the objective function described in Section II-A, which was a function of the system state. The objective function maps to the reward that is associated with the generation of a measurement $\mathbf{z}_{k+1}$ when the belief state was $\mathbf{b}_k$ and action $\mathbf{a}_k$ was taken.

For the remainder of the paper we will assume a partially observable problem. However, a completely observable problem can be recovered by substituting the belief state with the observable system state, considering the likelihood function as a Dirac delta function, and taking the state transition function instead of the belief state transition function.

III. Objective Functions for Cognitive Radar

The exact form of the objective function $r(\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$ is crucial, as it must accurately represent the physical problem to be solved. Specifying objective functions for cognitive radar can be loosely categorized into task, information, or utility (Quality-of-service) based approaches. However, the separation between the categories is not always distinct and existing approaches form more of a continuum.

A. Task Based Objective Functions

Task based objective functions calculate the cost or reward of an action in terms of a measure that is specific to the task being performed. Relevant task based metrics include radar timeline or spectrum usage, probability of target detection, detection range for an undetected target density, tracking root mean square error (RMSE), track sharpness, track purity, track continuity, and probability of correct target classification, to name a few. Each task based objective function can be regarded as some function $q(\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$ that is combined in some way to produce a scalar objective function that maps into the quality space $\mathcal{Q}$. It is often the case that a desired task based metric is difficult to calculate and is replaced by a surrogate metric such as signal-to-interference plus noise ratio (SINR) or an information theoretic metric.

B. Information Theoretic Objective Functions

A second class of objective functions used in cognitive radar and related fields is based on information theory. Broadly speaking, an information theoretic objective function gauges the relative merit of a sensing action in terms of the information flow it provides. A primary motivation for information-based objective functions is the ability to compare actions which generate different types of knowledge (e.g., knowledge about a target class versus knowledge about target position) using a common measuring stick.

A review of the history of information metrics in this context is provided in [4]. Here, we highlight some of the most commonly used objective functions. The most basic information theoretic objective function is the Posterior Shannon Entropy, given as:

$$\mathcal{H}(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1}) = \int p(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1}) \ln p(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1}) d\boldsymbol{X}_{k+1} \tag{1}$$

A related approach computes the information *gain* between densities rather than just the information contained in the posterior. The most popular approach uses the Kullback-Leibler Divergence (KLD), which is defined using the prior and posterior densities as:

$$\mathcal{D}\Big(p(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})||p(\boldsymbol{X}_{k+1}|\mathbf{b}_k)\Big) = \int p(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1}) \ln \frac{p(\boldsymbol{X}_{k+1}|\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})}{p(\boldsymbol{X}_{k+1}|\mathbf{b}_k)} d\boldsymbol{X}_{k+1} \tag{2}$$

The KLD has several desirable properties [5], including its connection to Mutual Information. There are a number of generalizations of the KLD in the literature, including the Renyi Divergence, the Arimoto α-divergences, and the $f-$divergence [6].

A third approach specific to parameter estimation is the Fisher Information Matrix (FIM) and related Bayesian Information Matrix (BIM) [7], which characterize the amount of information that a distribution contains about individual parameters (such as target position or velocity). The inverse of the BIM is the Bayesian Cramer-Rao Lower Bound (BCRLB), which quantifies the uncertainty in the parameter estimates. The (square root of) the BCRLB has the property that it is in the units of the parameter being estimated and is a lower bound on the RMSE. Thus it is often used as a surrogate for the RMSE and categorized as a task based metric.

C. Utility and QoS Based Objective Functions

Quality-of-Service approaches [2], [8] differ from task or information based objective functions in that they optimize the user or operator satisfaction that is derived from a task. A utility function is defined on the task quality space $\hat{u} : \mathcal{Q} \mapsto [0, 1]$ that should accurately describe the satisfaction that is derived from the different possible task quality levels. Combining the quality and utility functions results in an objective function of the required form $u(\mathbf{b}_k, \mathbf{a}_k, \mathbf{z}_{k+1})$.

This approach is very valuable in the context of radar resource management [2] as it enables a radar with limited resources to optimize multiple tasks based on the task quality levels that are required by the mission. Mapping the quality levels of differing radar tasks into the common utility space enables trade-offs between tasks evaluated using differing quality metrics. The global utility across the multiple tasks is typically formed by taking a weighted sum of task utilities. When considering the resource usage, a resource function $g(\mathbf{b}_k, \mathbf{a}_k)$ can be combined with the utility function to produce the final objective function. This quality-of-service conceptual approach can also be identified under different names [9].

IV. Multi-Step Objective Function

A general objective is to find a policy that determines a feasible action based on the belief state. The policy is a mapping from belief state to action denoted $\mathbf{a}_k = A^\pi(\mathbf{b}_k)$, where π carries information about the type of function and its parameters. As the belief state is a set of parameters describing a perception of the system state, the policy can be thought of as the perception-action cycle for a cognitive radar. The policy is not necessarily an analytical function and may actually represent an optimization problem. This section describes how a multi-step objective function is used to define optimal values and policies that are the basis for the design of perception-action cycles in the following section.

A. Optimal Values and Policies

The objective of a stochastic optimization problem is to maximize rewards or minimize costs over a time horizon comprising H future decision steps. The expected reward achievable over the current and future decision steps that originate from the current belief state is termed the value of the belief state. The value of a belief state when following policy π is the expected value of the summed rewards with respect to the set of future measurements $(\boldsymbol{Z}_{k+1}, ..., \boldsymbol{Z}_{k+H})$, conditioned on the belief state $\mathbf{b}_k$:

$$V_H^\pi(\mathbf{b}_k) = \mathbb{E}\left[\sum_{t=k}^{k+H} r\left(\boldsymbol{B}_t^\pi, A^\pi(\boldsymbol{B}_t^\pi), \boldsymbol{Z}_{t+1}\right) | \boldsymbol{B}_k^\pi = \mathbf{b}_k\right] \tag{3}$$

where the belief state random variables in the summation evolve according to the belief state transition function when following policy π, i.e. $\boldsymbol{B}_{k+1}^\pi = f_B(\boldsymbol{B}_k, A^\pi(\boldsymbol{B}_k), \boldsymbol{Z}_{k+1})$. It is common to rewrite (3) by splitting it into the expected reward for the current time step and the expected reward for subsequent time steps to give:

$$V_H^\pi(\mathbf{b}_k) = R(\mathbf{b}_k, A^\pi(\mathbf{b}_k)) + \mathbb{E}\left[V_{H-1}^\pi(\boldsymbol{B}_{k+1}^\pi) | \boldsymbol{B}_k = \mathbf{b}_k\right] \tag{4}$$

where the expectation is taken with respect to the future measurement $\boldsymbol{Z}_{k+1}$ and the single step reward is the expected reward with respect to the future measurement $\boldsymbol{Z}_{k+1}$:

$$R(\mathbf{b}_k, A^\pi(\mathbf{b}_k)) = \mathbb{E}\left[r\left(\boldsymbol{B}_k, A^\pi(\boldsymbol{B}_k), \boldsymbol{Z}_{k+1}\right) | \boldsymbol{B}_k = \mathbf{b}_k\right] \tag{5}$$

Note that the expectation with respect to the remaining future measurements $(\boldsymbol{Z}_{k+2}, ..., \boldsymbol{Z}_{k+H})$ in (3) is now contained in the future value term $V_{H-1}^\pi(\boldsymbol{B}_{k+1})$ in (4). Equation (4) can be identified as a form of Bellman's equation.

Based on the value of a belief state when following policy π, it is possible to define the optimal value of a belief state as:

$$V_H^*(\mathbf{b}_k) = \max_{\mathbf{a}_k \in \mathcal{A}} \left(R(\mathbf{b}_k, \mathbf{a}_k) + \mathbb{E}\left[V_{H-1}^*(\boldsymbol{B}_{k+1}^{\mathbf{a}_k}) | \boldsymbol{B}_k = \mathbf{b}_k\right]\right) \tag{6}$$

where $\boldsymbol{B}_{k+1}^{\mathbf{a}_k}$ is a random variable representing the belief state in the next decision step that evolves when taking action $\mathbf{a}_k$, i.e. $\boldsymbol{B}_{k+1}^{\mathbf{a}_k} = f_B(\boldsymbol{B}_k, \mathbf{a}_k, \boldsymbol{Z}_{k+1})$. Using the optimal value

function, the optimal policy function can be defined, which is a description of an optimal perception-action cycle:

$$A^*(\mathbf{b}_k) = \arg\max_{\mathbf{a}_k \in \mathcal{A}} \left(R(\mathbf{b}_k, \mathbf{a}_k) + \mathbb{E}\left[V^*_{H-1}(\boldsymbol{B}^{\mathbf{a}_k}_{k+1}) | \boldsymbol{B}_k = \mathbf{b}_k \right] \right) \tag{7}$$

The first term in (7) represents the expected reward associated with the current belief state and the next action, and is relatively easy to calculate. However, the second term that represents the expected reward associated with future belief states in the time horizon is very difficult to calculate. Consequently, solving the optimal policy function is generally intractable. The majority of stochastic optimization approaches focus on approximate solutions to this optimal policy function.

Equation (3) is a multi-step objective function for the case when it is desired to optimize the expected rewards accumulated over the time horizon. Alternatively, the terminal reward may be of interest at the end of the time horizon. This can be accommodated by using an altered objective function that returns zero except for the last decision step in the time horizon. This section has described a problem with finite horizon H. An infinite horizon problem can be described in the same way, but requires the inclusion of a discounting factor.

B. Simplified Multi-Step Objective Functions

Finding policies that solve (7) is very challenging due to the need to evaluate the impact of the current action on expected future rewards, knowing only the current belief state. There are simplifications that are often performed that drastically reduce the complexity of the problem but result in an objective function that does not fully consider the uncertainty present in the problem. These simplifications are often applied in current cognitive radar techniques, as will be shown in Section VI.

1) Myopic Optimization: If the time horizon is taken as a single step, i.e. $H = 1$, then the problem of evaluating the impact of the action on expected future rewards is removed. Hence, the optimal policy function in (7) is significantly simplified to:

$$A^*(\mathbf{b}_k) = \arg\max_{\mathbf{a}_k \in \mathcal{A}} \left(R(\mathbf{b}_k, \mathbf{a}_k) \right) \tag{8}$$

This approach is known as myopic or greedy as it focusses on the immediate expected reward and ignores and the impact of potential future rewards.

2) Deterministic Optimization: A second common simplification is to perform a deterministic optimization based on expected values of the system state and/or future measurements, instead of treating them as random variables and calculating the expected reward. An example of this approach would be to simplify the myopic reward function in (5) as:

$$R(\mathbf{b}_k, A^\pi(\mathbf{b}_k)) \approx r\left(\mathbf{b}_k, A^\pi(\mathbf{b}_k), \mathbb{E}\left[\boldsymbol{Z}_{t+1} | \boldsymbol{B}_k = \mathbf{b}_k\right]\right) \tag{9}$$

Where myopic optimization ignores the propagation of uncertainty into the future, deterministic optimization ignores the uncertainty in the belief state transition and measurement processes. However, by treating the optimization problem as being deterministic, it can be easier to solve.

Stochastic optimization techniques aim to find a policy that closely matches the optimal policy function and therefore perform an action that is optimized considering the uncertainty in the future evolution of the system and the noisy measurement process. However, it should be clear that solving the optimal value and policy functions for realistic problems is intractable. Consequently, existing cognitive radar techniques often simplify the problem by performing myopic or deterministic optimization. However, advances in computational capability combined with the development of new algorithms mean that it is possible to move away from these simplifications and look towards designing perception-action cycles that fully consider the uncertainty in the problem.

V. Policies and Perception-Action Cycles

Solving a stochastic optimization problem involves finding a policy that maps from belief states into actions and hence constitutes a perception-action cycle. This section gives an overview on methods for finding policies that are widely used in stochastic optimization. The purpose of this overview is to show that established algorithmic strategies from the field of stochastic optimization can be valuable tools for designing perception-action cycles in a cognitive radar. More details on these methodologies can be found in [3] and the references therein.

A. Policy Search

The general approach to policy search is to find and tune a policy that matches or approximates the optimal policy function in (7). Generally, the optimal policy is unlikely to be found, instead an approximation to the optimal policy function is sought, in the form of a policy function approximation or a cost function approximation.

1) Policy Function Approximations: Policy function approximations (PFAs) attempt to find and tune a function that approximates the optimal policy function in (7). For example, we can consider a family of functions $\mathcal{F}$, where a function $f \in \mathcal{F}$ is parameterized by $\theta \in \Theta^f$. Our goal is then to find a function and parameterisation f, θ so that the optimal policy function in (7) can be approximated as:

$$A^{f,\theta}_{PFA}(\mathbf{b}_k) = f(\mathbf{b}_k; \theta) \tag{10}$$

The optimal policy will be found if the optimal policy belongs to the family of functions and the corresponding parameter space. The goal of policy function approximations is not to find the optimal policy, but to find the best approximation within a class of function approximations. The function class may be any approach for approximating a function, such as an analytic function or a neural network.

2) Cost Function Approximations: Instead of approximating the entire policy function as with a PFA, a cost function approximation (CFA) finds a functional approximation to only the non-myopic cost function, which is interchangeable with the reward function described in this paper. Consequently, the optimal policy function in (7) is replaced with:

$$A^{\pi,\theta}_{CFA}(\mathbf{b}_k) = \arg\max_{\mathbf{a}_k \in \mathcal{A}^\pi(\theta)} \left[\tilde{r}^\pi(\mathbf{b}_k, \mathbf{a}_k; \theta) \right] \tag{11}$$

which comprises of the approximation to the cost function $\tilde{r}^{\pi}(\mathbf{b}_k, \mathbf{a}_k; \theta)$ as well as a potentially constrained action space $\mathcal{A}^{\pi}(\theta)$.

B. Lookahead Approximations

Lookahead approximations differ from policy search as they attempt to evaluate the influence of an action on future rewards, instead of approximating the policy function. A lookahead approximation can be performed via a value function approximation or by simulating a direct lookahead.

1) Value Function Approximations: A value function approximation uses the optimal policy function in (7), but replaces the true optimal value of future belief states $V^*_{H-1}(\boldsymbol{B}_{k+1})$ with an approximation $\tilde{V}_{H-1}(\boldsymbol{B}_{k+1})$. In some cases the expectation in (7) may be difficult to calculate, in which case a value function approximation can be used to replace $\mathbb{E}\left[V^*_{H-1}(\boldsymbol{B}_{k+1})|\boldsymbol{B}_k = \mathbf{b}_k\right]$.

2) Direct Lookahead: For the cases when it is not possible to find an accurate value function approximation, the expected future value can be evaluated by simulating future system evolutions using available models. As this process is computationally very costly, direct lookahead methods focus on making effective simplifications that still lead to accurate values. Common methods belonging to this class are deterministic lookaheads, Monte Carlo sampling, rollout policies and Monte Carlo tree search.

C. Discussion

General methodologies for finding policies involve finding a function approximation to either the policy function, the cost function or the value function. The difference between these approaches is simply where the functional approximation is made, as illustrated in Figure 1. The effectiveness of these approaches depends on how well a function approximation can capture these respective relationships. All of these methodologies can be implemented with handcrafted models or using machine learning techniques. Although it is typical to perform offline training, these function approximations could be updated online as more data becomes available. Direct lookahead approaches are used when it is not possible to capture the structure of the problem with a function approximation.

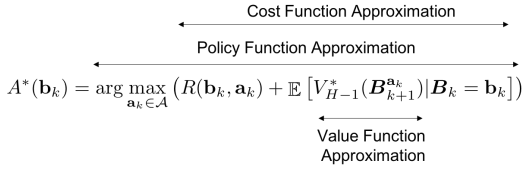

Fig. 1. Different function approximation types for the optimal policy function.

VI. Cognitive Radar Problems

A representative set of cognitive radar problems for different applications can be found in the references. Although it may not always be explicitly stated, these problems can be characterized as stochastic optimization problems that possess the framework components described in Section II. The components are often explicitly stated or can be easily inferred. A variety of solution methodologies have been applied, which can be compared with the strategies described in Section V.

A. Problem Components

In the case of target tracking [2], [9]–[16], the belief state characterizes a posterior probability density function defined on the system state space. Typical belief states are the mean and covariance matrix of the distribution or a set of particles. The belief state transition function incorporates the Bayesian prediction and update processes. The exogenous information is some noisy function of the system state that maps to radar measurements, thus the system state is partially observable. Often, the likelihood function is a Gaussian approximation of the true measurement errors. Adaptive tracking [2], [10], [12] methods select actions in the form of revisit interval times as well as the waveform energy for the next measurement, in order to minimize resource usage while maintaining track. An early approach [10] was to use a function that mapped measurement and track accuracies, and Singer manoeuvre parameters to a revisit interval time. In the context of the methods described in Section V, this can be thought of as an empirically derived policy function approximation. Another strand of work has focussed on waveform selection and adaptation [11], [14], [15], whereby the action space comprised different waveform modulations that were selected in order to minimize track RMSE.

The framework components are easy to identify for tracking problems, because the framework is essentially an extension to the standard Bayesian tracking process. However, other radar functions and applications can also be cast into the framework. For a search problem, the belief state can parameterize an undetected target posterior density. In target detection [17]–[20], the system state is the state of the clutter, interference, and noise environment. Typical belief states include the clutter, interference, and noise covariance matrix or a posterior distribution on a spectrum occupancy state. For imaging and classification [21] the belief state characterizes a posterior probability mass function. Typical belief states are the pairwise likelihood ratios or the posterior probabilities themselves. Some works also consider a combination of radar functions [22]–[24].

Generally, the action space is some set of parameters that characterize the radar transmission and reception, including transmit and receive sensor selection and scheduling, transmit frequency, bandwidth, time, duration, power, and waveform design. The exogenous information is some noisy function of the system state, thus the system state is partially observable. Generally, objective functions differ widely, but can be categorized according to the classes in Section III.

B. Solution Methodologies

The majority of the reference works formulate myopic optimization problems, which represent a simplification with respect to the general non-myopic multi-step objective function.

Depending on the problem, this can be a very valid approach to reduce the complexity of the optimization, especially if it is clear that the current action does not influence future rewards. However, it is worthwhile to explicitly consider how the myopic and non-myopic solutions differ, as there are certainly problems where considering the future rewards associated with the current action can significantly improve performance.

There are also cases in the reference works where an optimization is performed on an expected value of the system state and/or an expected future measurement, instead of treating the system state and future measurements as random variables and calculating the expected reward. This approach has the benefit of enabling deterministic optimization methods to be applied and is a valid approximation if the reward function is not sensitive in the region of significant probability as described by the posterior and expected measurement PDFs. However, this approach ignores or under-utilizes the uncertainty in the future state evolution and corresponding measurements, which could significantly impact performance.

The cognitive radar methodologies in the reference works generally attempt to solve an optimization problem online by performing numerical optimizations or searches over the action space. However, the strategies described in Section V first attempt to identify structure in the policy, cost or value function and attempt to use specific models or machine learning to produce a functional approximation. This is a particularly attractive approach because it can reduce the complexity of the online optimization problem, or remove the need to perform an online optimization, depending on the functional approximation type. This approach is underrepresented in the reference works, but can be identified in [16], where a neural network is used to learn the policy function that an optimizer with more complexity would generate.

VII. Conclusion

Many cognitive radar techniques are emerging that tackle different applications or sub-problems in a radar system. This paper has presented a common framework for describing these cognitive radar problems in terms of a stochastic optimization problem. By doing so, the cognitive radar problem can be addressed using existing algorithmic strategies from the field of stochastic optimization. Specifically, the strategy of finding functional approximations for the optimal policy, cost or value function using machine learning techniques is an attractive approach. Traditionally, cognitive radar and radar management has performed myopic and deterministic optimizations. However, advances in computing and algorithmic capabilities can enable the more general stochastic optimization problem to be tackled, which fully considers uncertain measurements and state transitions as well as the impact of action selection on future rewards.

References

[1] S. Z. Gurbuz, H. D. Griffiths, A. Charlish, M. Rangaswamy, M. S. Greco, and K. Bell, "An Overview of Cognitive Radar: Past, Present, and Future," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 12, pp. 6–18, 2019.

[2] A. Charlish and F. Hoffmann, "Cognitive radar management," in *Novel Radar Techniques and Applications Volume 2: Waveform Diversity and Cognitive Radar, and Target Tracking and Data Fusion*. Institution of Engineering and Technology, 2017, pp. 157–193.

[3] W. B. Powell, "A unified framework for stochastic optimization," *European Journal of Operational Research*, vol. 275, no. 3, pp. 795 – 821, 2019.

[4] A. O. Hero, D. A. Castanon, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*. Springer, 2007.

[5] E. H. Aoki, A. Bagchi, P. Mandal, and Y. Boers, "A theoretical look at information-driven sensor management criteria," in *14th International Conference on Information Fusion*, 2011, pp. 1–8.

[6] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, 2006.

[7] H. L. Van Trees and K. L. Bell, Eds., *Bayesian Bounds for Nonlinear Filtering/Tracking*. Wiley, 2007.

[8] J. P. Hansen, S. Ghosh, R. Rajkumar, and J. Lehoczky, "Resource management of highly configurable tasks," in *18th International Parallel and Distributed Processing Symposium*, Santa Fe, New Mexico, 2004.

[9] A. E. Mitchell, G. E. Smith, K. L. Bell, A. J. Duly, and M. Rangaswamy, "Cost function design for the fully adaptive radar framework," *IET Radar, Sonar, and Navigation*, vol. 12, no. 12, pp. 1380–1389, 2018.

[10] G. van Keuk and S. S. Blackman, "On phased-array radar tracking and parameter control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 186–194, 1993.

[11] D. J. Kershaw and R. J. Evans, "Waveform selective probabilistic data association," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1180–1188, 1997.

[12] T. Kirubarajan, Y. Bar-Shalom, W. D. Blair, and G. A. Watson, "IMM-PDAF for radar management and tracking benchmark with ECM," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1115–1134, 1998.

[13] E. K. P. Chong, C. M. Kreucher, and A. O. Hero, "Monte-Carlo-based partially observable Markov decision process approximations for adaptive sensing," in *9th Intl. Wkshp. on Discrete Event Systems*, 2008, pp. 173–180.

[14] S. P. Sira, Y. Li, A. Papandreou-Suppappola, D. Morrell, D. Cochran, and M. Rangaswamy, "Waveform-agile sensing for tracking," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 53–64, 2009.

[15] S. Haykin, *Cognitive Dynamic Systems: Perception-action Cycle, Radar and Radio*. Cambridge University Press, 2012.

[16] P. John-Baptiste and G. E. Smith, "Utilizing neural networks for fully adaptive radar," in *IEEE Radar Conf.*, 2019.

[17] J. R. Guerci, *Cognitive Radar: The Knowledge-Aided Fully Adaptive Approach*. Artech House, 2010.

[18] A. Aubry, A. De Maio, M. Piezzo, and A. Farina, "Radar waveform design in a spectrally crowded environment via nonconvex quadratic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1138–1152, 2014.

[19] P. Stinco, M. Greco, and F. Gini, "Cognitive radars in spectrally dense environments," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 10, pp. 20–27, 2016.

[20] A. F. Martone, K. I. Ranney, K. Sherbondy, K. A. Gallagher, and S. D. Blunt, "Spectrum allocation for noncooperative radar coexistence," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 90–105, 2018.

[21] N. A. Goodman, P. R. Venkata, and M. A. Neifeld, "Adaptive Waveform Design and Sequential Hypothesis Testing for Target Recognition With Active Sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 105–113, 2007.

[22] C. Kreucher, A. O. Hero, and K. Kastella, "A Comparison of Task Driven and Information Driven Sensor Management for Target Tracking," in *44th IEEE Conference on Decision and Control*, dec 2005, pp. 4004–4009.

[23] K. L. Bell, C. J. Baker, G. E. Smith, J. T. Johnson, and M. Rangaswamy, "Cognitive Radar Framework for Target Detection and Tracking," *IEEE Journal on Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1427–1439, 2015.

[24] A. Charlish and F. Katsilieris, "Array Radar Resource Management," in *Novel Radar Techniques and Applications: Real Aperture Array Radar, Imaging Radar, and Passive and Multistatic Radar*. Institution of Engineering and Technology, 2017, pp. 135–171.