Optimal Policies Search for Sensor Management : Application to the AESA Radar

Thomas Bréhard, Pierre-Arnaud Coquelin, Emmanuel Duflos

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Optimal Policies Search for Sensor Management : Application to the AESA Radar

Thomas Bréhard — Pierre-Arnaud Coquelin — Emmanuel Duflos

N° 6361

11-16-2007

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6361--FR+ENG

Optimal Policies Search for Sensor Management : Application to the AESA Radar

Thomas Bréhard , Pierre-Arnaud Coquelin , Emmanuel Duflos *

Thème COG — Systèmes cognitifs
Projet SequeL

Rapport de recherche n° 6361 — 11-16-2007 — 21 pages

Abstract: This report introduces a new approach to solve sensor management problems. Classically sensor management problems are formalized as Partially-Observed Markov Decision Process (POMPD). Our original approach consists in deriving the optimal parameterized policy based on stochastic gradient estimation. Two differents techniques nammed Infinitesimal Approximation (IPA) and Likelihood Ratio (LR) can be used to adress such a problem. This report discusses how these methods can be used for gradient estimation in the context of sensor management . The effectiveness of this general framework is illustrated by the managing of an Active Electronically Scanned Array Radar (AESA Radar).

Key-words: Sensor Management, AESA Radar, Stochastic Gradient, Partially Observable arkov Decision Process, Particle Filtering

* Emmanuel Duflos is Professor at the Ecole Centrale de Lille and is also with the Laboratoire d'Automatique, Génie Informatique et Signal (LAGIS UMR CNRS 8146)

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 — Télécopie : +33 1 60 19 66 08

Recherche de Politiques Optimale en Gestion de Capteur : Application à un Radar AESA

Résumé : Ce rapport introduit une nouvelle approche pour développer des méthodes de gestions optimales de capteurs. De tels problèmes peuvent classiquement être modélisés par des POMDP (Partially-Observed Markov Decision Process). L'approche originale développée dans ce rapport consiste à rechercher des politiques optimales paramétrées et de mettre en œuvre des méthodes telles que IPA (Infinitesimal Approximation) et LR (Likelihood Ratio) pour déterminer les paramètres. Nous expliquons comment ces deux méthodes peuvent être mise en œuvre dans notre contexte par le biais de méthodes d'estimation de gradients stochastiques. La méthode générale développée dans la première partie est illustrée dans le cas particulier du Radar AESA.

Mots-clés : Gestion de capteurs, Radar Balayage Electronique AESA, Gradient Stochastique,Partially Observable arkov Decision Process, Filtrage Particulaire

Notations

t_n: instant time of the n-th observation,

n_t: number of the last observation before instant time t i.e. $n_t = \max_j\{t_j < t\}$.

1 Introduction

Let us consider a Partially Observable Markovian Decision Process (POMDP) where $(X_t)_{t\geq 0}$ is the state process. The latter is observed via a sequence of actions $(A_n)_{n\in N}$ such that the observation process $(Y_n)_{n\in\mathbb{N}}$ is linked to the state process by the conditional probability measure :

$$\mathbb{P}(Y_n \in dy_n | X_{t_n} = x_{t_n}, A_n) \tag{1}$$

where t_n is the instant time of the n-th observation. Using a judicious sequence of actions, one can expect an accurate estimate of the state process. This problem is known in the literature as a sensor management problem. From a general point of view, sensor management deals with ressource allocation, scheduling and adaptive deployment of multiple sensors for detection, tracking and identification of targets, this term being used here in its more general meaning.

Input A_n can be any tunable parameter of one or several sensors. In [1], A_n refers to the mode of the sensor of an airborn platform (radar or Infra-Red). As a matter of facts, the choice of the mode is critical when considering ”smart” targets. When such targets detects it is under analysis by an active sensor, it reacts to make surveillance more difficult. Alternatively, in the optimal measurement scheduling problem [2], A_n is directly related to the accuracy of the measurement. The problem consists in determining the time-distribution of measurements under some specific constraints. Otherwise, in multi-sensor applications [3], A_n denotes the activate sensor at time t. In this case, sensor management aims at trading off tracking error with sensor usage cost. Thus in the domain of antisubmarine warefare [4], only a limited number of sensor can provide measurements to the tracker due to bandwidth constraints. In the optimal observer trajectory problem [5], A_n denotes the position of the observer at time t. Finally, a major application concerns the Active Electronically Scanned Array (AESA) radar [6]. The AESA radar is an agile beam radar which means that it is able to point its beam in any direction of environnement. The goal is to minimize the use of the radar ressources while maintening targets under track and detect new ones. Different parameters of this sensor are tunable. In [7], the authors consider the optimization of the direction of the beam of the radar. In [8], A_n is the waveform. It is worth being noticed that different waveforms can be used to achieve good performance, good Doppler and good range resolution but not simultaneously.

When the POMDP is Gaussian linear with a quadratic cost function, Meier et al [9] derived a closed-form solution. Nevertheless, one can not expect closed-form solutions in the non-linear non-Gaussian cases. Thus, a first approach consists to combine a Q-value approximation with a particle filtering [3]. Particle filtering [10] is a Monte-Carlo method for estimation in Hidden Markov Model. The Q-value approximation estimate the Q-value i.e the expected cumulative cost associated to each condidate action.

The main contributions of this report are the following:

- A general framework to find a parameterized optimal policy for sensor management problems.
- Derivation of a parameterized optimal policy based on stochastic gradient estimation.
- A general approach to use IPA and LR methods for gradient estimation.
- An application to the mangement of an Active Electronically Scanned Array radar.

In Sec.2, we derive two general algorithms to solve a POMDP based on Infinitesimal Perturbation Analysis and Likelihood Ratio methods. We discuss in Sec.3 how gradient estimation can be used to solve the management of an Active Electronically Scanned Array Radar.

2 Gradient estimation for Partially-Observable Markov Decision Process

2.1 Partially-Observable Markov Decision Process

Let us consider a probability space denoted by $(\Omega, \sigma(\Omega), \mathbb{P})$. A Partially-Observable Decision Proceess is defined by a *state process* $(X_t)_{t\geq 0}$, an *observation process* $(Y_n)_{n\in\mathbb{N}}$ and a set of action $(A_n)_{n\in\mathbb{N}}$.

The state process is an homogeneous Markov chain taking its values in a continuous state space denoted by $(\mathcal{X}, \sigma(\mathcal{X}))$ and with initial probability measure $\mu(dx_0)$ and Markov transition kernel $K(dx_{t+1}|x_t)$, i.e., $\forall t \geq 0,\ X_{t+1} \sim K(\cdot|X_t)$ and $X_0 \sim \mu$ ([11]). In the following we assume that there exists two generative functions $F_\mu : U \rightarrow \mathcal{X}$ and $F : \mathcal{X} \times U \rightarrow \mathcal{X}$, where $(U, \sigma(U), \nu)$ is a probability space, such that for any measurable *test function* f on $\mathcal{X}$

$$\int_{\mathcal{X}} f(x_t) K(dx_t|x_{t-1}) \quad = \quad \int f(F(x_{t-1}, u))\nu(du) \tag{2}$$

and

$$\int_{\mathcal{X}} f(x_0)\mu(dx_0) \quad = \int f(F_\mu(u))\nu(du). \tag{3}$$

In many practical situations, $U = [0,1]^{n_U}$, and u is a n_U-uple of pseudo random numbers generated by a computer. For sake of simplicity, we adopt the notations $K(dx_0|x_{-1}) \triangleq \mu(dx_0)$ and $F(x_{-1}, u) \triangleq F_\mu(u)$. Under this framework, the Markov Chain $(X_t)_{t\geq 0}$ is fully specified by the following dynamical equation $X_{t+1} = F(X_t, U_t),\ U_t \overset{i.i.d.}{\sim} \nu.$. The observation process $(Y_n)_{n\in\mathbb{N}}$ is defined on the measurable space $(\mathcal{Y}, \sigma(Y))$ and is linked with the state process by the conditional probability measure $\mathbb{P}(Y_n \in dy_n | X_{t_n} = x_{t_n}, A_n) = g(y_n, x_{t_n}, A_n)\lambda(dy_t)$, where $A_n \in \mathcal{A}$ is an n-th action variable where $(\mathcal{A}, \sigma(\mathcal{A}))$ is the action space. Term t_n is the instant time of the n-th observation, λ is a fixed probability measure on $\mathcal{Y}$ and $g : \mathcal{Y} \times \mathcal{X} \rightarrow [0,1]$ a positive function. We assume that observations are conditionally independent given the state process, i.e.:

$$\forall 1 \leq i,j \leq t,\ i \neq j, \qquad \mathbb{P}(Y_i \in dy_i, Y_j \in dy_j | X_{0:t}, A_i, A_j) = \mathbb{P}(Y_i \in dy_i | X_{0:t}, A_i)\mathbb{P}(Y_j \in dy_j | X_{0:t}, A_j) \tag{4}$$

where we have adopted the usual notation $z_{i:j} = (z_k)_{i \leq k \leq j}$.

2.2 Filtering distribution in a Partially-Observable Markov Decision Process

Given a sequence of action $A_{1:n}$ and a sample trajectory of the observation process $y_{1:n}$ and indices $\{n_1, n_2, t_1, t_2\}$ such that $1 \leq n_1 \leq n_2 \leq n$ and $0 \leq t_1 \leq t_{n_1} \leq t_{n_2} \leq t_2 \leq t_n$, we define the posterior probability distribution by ([12])

$$\begin{aligned} M_{t_1:t_2|n_1:n_2}(dx_{t_1:t_2}) &\triangleq \mathbb{P}(X_{t_1:t_2} \in dx_{t_1:t_2} | Y_{n_1:n_2} = y_{n_1:n_2}, A_{n_1:n_2}) && (5) \\ &= \frac{\prod_{t=t_1}^{t_2} K(dx_t|x_{t-1}) \prod_{j=n_1}^{n_2} G_{t_j}(x_{t_j})}{\int_{\mathcal{X}^{t_2-t_1}} \prod_{t=t_1}^{t_2} K(dx_t|x_{t-1}) \prod_{j=n_1}^{n_2} G_{t_j}(x_{t_j})}, && (6) \end{aligned}$$

where for simplicity $G_{t_n}(x_{t_n}) \triangleq g(y_n, x_{t_n}, A_n)$ and $G_0(x_0) \triangleq 0$. One of the main interest here is to recover the state at time t from noisy observations $y_{1:n_t}$. From a bayesian point of view this information is completely contained in the *filtering distribution* $M_{t:t|1:n_t}$. In the following, the index t and the observations $y_{1:n_t}$ are fixed, and the filtering distribution is simply denoted by M_t.

2.3 Numerical methods for estimating the filtering distribution

Given a measurable test function $f : \mathcal{X} \to \mathbb{R}$, we want to evaluate

$$M_t(f) = \mathbb{E}[f(X_t) | Y_{1:n_t} = y_{1:n_t}, A_{1:n_t}] = \frac{\mathbb{E}[f(X_t) \prod_{j=1}^{n_t} G_{t_j}(X_{t_j})]}{\mathbb{E}[\prod_{j=1}^{n_t} G_{t_j}(X_{t_j})]}. \tag{7}$$

In general, it is impossible to find $M_t(f)$ exactly except for simple cases such as linear/gaussian (using Kalman filter) or for finite state space Hidden Markov Models. In the general dynamics, continuous space case considered here, possible numerical methods for computing $M_t(f)$ include the Extended Kalman filter, quantization methods, Markov Chain Monte Carlo methods and Sequential Monte Carlo methods (SMC).

The basic SMC method, called Bootstrap Filter (see [10] for details), approximates $M_t(f)$ by an empirical distribution $M_t^N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_i^N)$ made of N particles. The reader can find some convergence results of $M_t^N(f)$ to $M_t(f)$ (e.g. Law of Large Numbers or Central Limit Theorems) in [12], but for our purpose we note that under weak conditions on the test function and on the HMM dynamics, we have the asymptotic consistency property in probability, i.e. $\lim_{N\to\infty} M_t^N(f) \overset{\mathbb{P}}{=} M_t(f)$.

2.4 Optimal Parameterized Policy for Partially-Observable Markov Decision Process

Let R_t be a real value reward function

$$R_t \triangleq R(X_t, M_t(f)) \; . \tag{8}$$

The goal is to find at each new iteration a policy $\pi : \mathcal{A}^n \times \mathcal{Y}^n \to \mathcal{A}$ that maximizes the criterion performance i.e.

$$J_\pi = \int_0^T \mathbb{E}[R_t] dt \tag{9}$$

where T is the duration of the scenario. In practice, designing a sequence of policies in which each policy depend on the whole trajectory of the past observations/actions is unrealistic. It has been proved that the class of stationary policies that depend on the filtering distribution conditionally to past observations/actions M_t contains the optimal policy. In general the filtering distribution is an infinite dimensional object, and it cannot be represented in a computer. We propose to look for the optimal policy in a class of parameterized policies $(\pi_\alpha)_{\alpha\in\Gamma}$ that depend on a statistic of the filtering distribution

$$A_{n+1} = \pi_\alpha(M_{t_n}(f)) \tag{10}$$

where f is a test function. As the policy π is parameterized by α, the performance criterion depends only on α, thus we can maximize it by achieving a stochastic gradient ascent with respect to α.

$$\alpha_{k+1} = \alpha_k + \eta_k \nabla J_{\alpha_k}, \quad k \geq 0 \tag{11}$$

where ∇J_{α_k} denotes the gradient of J_{α_k} w.r.t α_k. By convention ∇J_{α_k} is column vector whose i-th component is the partial derivative with respect to α_i. $(\eta_k)_{k\geq 0}$ is a non-increasing positive sequence tending to zero. We present in the two following subsection two approaches to estimate ∇J_{α_k}: Infinitesimal Perturbation Analysis (IPA) and Likelihood Ratio (LR).

2.5 Infinitesimal Perturbation Analysis for gradient estimation

Notice first that under appropriate assumptions, $\nabla J_\alpha = \int_0^T \nabla_\alpha \mathbb{E}[R_t]dt$ (for simplicity suscribe k has been avoided). We have the following decomposition of the gradient

$$\nabla_\alpha \mathbb{E}[R_t] = \mathbb{E}[M_t(fS_t)\nabla_{M_t(f)}R_t] - \mathbb{E}[M_t(f)M_t(S_t)\nabla_{M_t(f)}R_t] + \mathbb{E}[R_t S_t] \tag{12}$$

where

$$S_t = \sum_{j=1}^{p_t} \frac{\nabla_\alpha G_{t_j}(X_{t_j})}{G_{t_j}(X_{t_j})} \tag{13}$$

Eq.(12) is proved in Appendix A. We deduce directly Algorithm 1 from (12).

2.6 Likelihood Ratio for gradient estimation

The method below is an application of the work described in [11]. The aim is to find an approximation of the gradient using a finite difference method.

$$\nabla_\alpha \mathbb{E}[R_{t,\alpha}] = \nabla_\alpha \int \mathbf{r}_{t,\alpha}\mathcal{Y}_{t,\alpha}\mathcal{Z}_{t,\alpha} \tag{14}$$

where

$$\mathcal{Y}_{t,\alpha} = \prod_{j=1}^{J_t^\alpha} p(d\tilde{Y}_{t_j^\alpha}|X_{t_j^\alpha}, \pi_\alpha(\mathbf{m}_{t_{j-1}^\alpha,\alpha}), \mathbf{z}_{t_j^\alpha}) \tag{15}$$

$$\mathcal{Z}_{t,\alpha} = \prod_{j=1}^{J_t^\alpha} p(d\mathbf{z}_{t_j^\alpha}|X_{t_j^\alpha}, \pi_\alpha(\mathbf{m}_{t_{j-1}^\alpha,\alpha})) \tag{16}$$

and

$$\mathbf{m}_{t,\alpha} = \mathbb{E}\left\{f(X_t)|Y_{t_1^\alpha:t_{J_t}^\alpha}\right\} \tag{17}$$

$$\mathbf{m}_{t,\alpha+h} = \frac{\mathbb{E}\left\{f(X_t)\frac{\mathcal{Y}_{t,\alpha+h}}{\mathcal{Y}_{t,\alpha}}|Y_{t_1^\alpha:t_{J_t}^\alpha}\right\}}{\mathbb{E}\left\{\frac{\mathcal{Y}_{t,\alpha+h}}{\mathcal{Y}_{t,\alpha}}|Y_{t_1^\alpha:t_{J_t}^\alpha}\right\}} \tag{18}$$

The proof of the expression of $\mathbf{m}_{t,\alpha+h}$ is given in Appendix B. The gradient may then be approximated as :

$$\nabla_\alpha\left\{\mathbb{E}\mathbf{r}_{t,\alpha}\right\} \approx \frac{\int \mathbf{r}_{t,\alpha+h}\mathcal{Y}_{t,\alpha+h}\mathcal{Z}_{t,\alpha+h} - \int \mathbf{r}_{t,\alpha}\mathcal{Y}_{t,\alpha}\mathcal{Z}_{t,\alpha}}{h} \tag{19}$$

$$\approx \frac{\int \mathbf{r}_{t,\alpha+h}\mathcal{Y}_{t,\alpha+h}\frac{\mathcal{Z}_{t,\alpha+h}}{\mathcal{Z}_{t,\alpha}}\mathcal{Z}_{t,\alpha} - \int \mathbf{r}_{t,\alpha}\mathcal{Y}_{t,\alpha}\mathcal{Z}_{t,\alpha}}{h} \tag{20}$$

The corresponding algorithm is Algorithm 2.

3 Application to Active Electronically Scanned Array Radars

The AESA is an agile beam radar which means that it is able to point its beam in any direction of the environnement instantaneously without inertia. However, the targets in the environement are detected w.r.t a probability of detection which depends on the direction of the beam and the time of observation in this direction (see Appendix C and Apprendix D). We precise first the nature of action, the influence of the action the probability of detection and finally the nature of the observations.

Definition of the action The main property of an AESA is that it can point its beam without mechanically adjusting the antenna. An AESA radar provides measurements in a direction θ. We note δ, the time of observation in this direction. The n-th action is noted:

$$A_n = \begin{bmatrix}\theta_n & \delta_n\end{bmatrix}^T \tag{21}$$

with

$$\begin{cases} \theta_n \in [-\frac{\pi}{2}, \frac{\pi}{2}] \ , \\ \delta_n \in \mathbb{R}^+ \end{cases} \qquad \forall n \geq 0 \ . \tag{22}$$

The action does not influence directly the observation produced by the AESA but the probability of detection of a target.

The probability of detection P_d refers to the probability to obtain an estimation of the state of a target p at time t_n denoted $X_{t_n,p}$ with action a_n. $X_{t_n,p}$ is composed of the localisation and velocity components of the target p at time t_n in the x-y plane:

$$X_{t_n,p} = \begin{bmatrix} rx_{t_n,p} & ry_{t_n,p} & vx_{t_n,p} & ry_{t_n,p} \end{bmatrix}^T \tag{23}$$

The terms $rx_{t_n,p}$ and $ry_{t_n,p}$ refers here to the position and $vx_{t_n,p}$ and $vy_{t_n,p}$ the velocity of target p at time t_n. We also denote $D_{n,p}$ the random variable which takes values 1 if the radar produces a detection (and therefore an estimation) for target p and 0 else :

$$D_n = \begin{bmatrix} D_{n,1} & \dots & D_{n,P} \end{bmatrix}^T \ . \tag{24}$$

This probability also depends on the time of observation δ_n. If the reflectivity of a target can be modelled using a Swerling I model [13] then we have the following relation between the probability of detection and the probability of false alarm [6]:

$$P_d(x_{t_n,p}, A_n) = P_{fa}^{\frac{1}{1+\rho(x_{t_n,p},A_n)}} \tag{25}$$

where P_{fa} is the probability of false alarm (the probability to obtain a measurement knowing that there is no target) and $\rho(x_{t_n,p}, A_n)$ the target signal-to-noise ratio. The equation (25) is derived in Appendix A. The signal-to-noise ratio for an AESA radar, $\rho(x_{t_n,p}, A_n)$, is defined as :

$$\rho(x_{t_n,p}, A_n) = \alpha\delta_n \frac{cos^2\theta_n}{r^4_{t_n,p}} e^{-\frac{(\beta_{t_n,p}-\theta_n)^2}{2B^2}} \tag{26}$$

where $r_{t_n,p}$ is the target range and $\beta_{t_n,p}$ the azimuth associated to target p at instant time t_n. α is a coefficient which includes all the parameters of the sensor and B is the beamwidth of the radar. This radar equation (26) is derived in Appendix B. If we make the assumption that all the detections are independant, we can write :

$$\mathbb{P}(D_n = d_n | X_{t_n} = x_{t_n}, A_n) = \prod_p^P \mathbb{P}(D_{n,p} = d_{n,p} | X_{t_n,p} = x_{t_n,p}, A_n) \tag{27}$$

where

$$\mathbb{P}(D_{n,p} = d_{n,p} | X_{t_n,p} = x_{t_n,p}, A_n) = P_d(x_{t_n,p}, A_n)\delta_{d_{n,p}=1} + (1 - P_d(x_{t_n,p}, A_n))\delta_{d_{n,p}=0} \tag{28}$$

Observation equation At instant time t_n, the radar produces a raw observation Y_n composed of P measurements :

$$Y_n = \begin{bmatrix} Y_{n,1} & \ldots & Y_{n,P} \end{bmatrix}^T . \tag{29}$$

where $Y_{n,p}$ is the observation related to target $x_{t_n,p}$ obtained with action A_n. Remark that we do not consider here the problem of measurement-target association. Moveover, we assume that the number of targets P is known. Each of these measurements has the following formulation :

$$Y_{n,p} = \begin{bmatrix} r_{n,p} & \beta_{n,p} & \dot{r}_{n,p} \end{bmatrix}^T \tag{30}$$

where $r_{n,p}$, $\beta_{n,p}$, $\dot{r}_{n,p}$ are range, azimuth and range rate. The equation observation can be written

$$\mathbb{P}(Y_n \in dy_n | X_{t_n} = x_{t_n}, A_n) = \prod_p^P \mathbb{P}(Y_{n,p} \in dy_{n,p} | X_{t_n,p} = x_{t_n,p}, A_n) \tag{31}$$

where

$$\mathbb{P}(Y_{n,p} \in dy_{n,p} | X_{t_n,p} = x_{t_n,p}, A_n) = g(y_{n,p}, x_{t_n,p}, A_n)\lambda(dy_{n,p}) \tag{32}$$

$$g(y_{n,p}, x_{t_n,p}, A_n) = \begin{bmatrix} \mathcal{N}\left(h_t(x_{t_n,p}), \Sigma_y\right) P_d(x_{t_n,p}, A_n) & 1 - P_d(x_{t_n,p}, A_n) \end{bmatrix} \tag{33}$$

and

$$\lambda(dy_{n,p}) = \lambda_{cont}(dy_{n,p}) + \lambda_{disc}(dy_{n,p}) \tag{34}$$

The relation between the state and the raw observations is given by :

$$Y_{n,p} = h_{t_n}(X_{t_n,p}) + W_{n,p} \tag{35}$$

with

$$h_{t_n}(x_{t_n,p}) = \begin{pmatrix} \sqrt{(rx_{t_n,p} - rx^{obs}_{t_n})^2 + (ry_{t_n,p} - ry^{obs}_{t_n})^2} \\ \operatorname{atan}\left\{\frac{ry_{t_n,p} - ry^{obs}_{t_n}}{rx_{t_n,p} - rx^{obs}_{t_n}}\right\} \\ \frac{(rx_{t_n,p} - rx^{obs}_{t_n})(vx_{t_n,p} - vx^{obs}_{t_n}) + (ry_{t_n,p} - ry^{obs}_{t_n})(vy_{t_n,p} - vy^{obs}_{t_n})}{\sqrt{(rx_{t_n,p} - rx^{obs}_{t_n})^2 + (ry_{t_n,p} - ry^{obs}_{t_n})^2}} \end{pmatrix} \tag{36}$$

and $W_{n,p}$ a gaussian noise the covariance matrix of which is given by :

$$\Sigma_y = diag(\sigma_r^2, \sigma_\beta^2, \sigma_{\dot{r}}^2) \ . \tag{37}$$

State equation First let us introduce the definition of the unknown state X_t at time t and its evolution through time. $X_{t,p}$ is the state of the target p. It has been defined above. Let P be the known number of targets in the space under analysis at time t. X_t has the following form: .

$$X_t = \begin{bmatrix} X_{t,1} & \dots & X_{t,P} \end{bmatrix}^T \tag{38}$$

Based on [14] works, we classically assume that all the targets follow a nearly constant velocity model. We use a discretized version of this model ([15]) :

$$X_{t,p} = F(X_{t-1,p}, U_t) \text{ where } U_t \sim \mathcal{N}\left(0, \sigma^2 Q\right) \tag{39}$$

where

$$F = \begin{bmatrix} 1 & 0 & \beta & 0 \\ 0 & 1 & 0 & \beta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} \frac{\beta^3}{3} & 0 & \frac{\beta^2}{2} & 0 \\ 0 & \frac{\beta^3}{3} & 0 & \frac{\beta^2}{2} \\ \frac{\beta^2}{2} & 0 & \beta & 0 \\ 0 & \frac{\beta^2}{2} & 0 & \beta \end{bmatrix} \ . \tag{40}$$

4 Conclusion

This report shows how to combine the POMDP modelling of a sensor management problem and optimal parametrized policies search using stochactic gradient estimation to derive optimal sensor management strategies. This work is based upon recent developments in gradient estimation. Both techniques Infinitesimal Perturbation Analysis and Likelihood Ratio are analysed and two policy search algorithms are derived. We then show how the proposed methods can be applied to the specific case of an AESA Radar.

Appendix A: Proof of (12)

First let us rewrite $\nabla_\alpha \mathbb{E}[R_t]$ as following:

$$\nabla_\alpha \mathbb{E}[R_t] = \nabla_\alpha \int_{\mathcal{X}^t \times \mathcal{Y}^{n_t}} R_t U_t V_t \prod_{j=n_1}^{n_t} \lambda(dy_j) \quad \text{where} \left\{ \begin{array}{ll} U_t & = \prod_{i=0}^{t} K(dx_i|x_{i-1}) \ , \\ V_t & = \prod_{j=1}^{n_t} G_{t_j}(x_{t_j}) \end{array} \right. . \quad (41)$$

Remark that only R_t and V_t depends on α so that we obtain

$$\left\{ \begin{array}{ll} \nabla_\alpha V_t & = S_t V_t \ , \\ \nabla_\alpha R_t & = \nabla_\alpha M_t(f) \nabla_{M_t(f)} R_t \end{array} \right. \quad (42)$$

where S_t is given by eq.(13). Incorporating (41) in (42), we obtain

$$\nabla_\alpha \mathbb{E}[R_t] = \quad \mathbb{E}[\nabla_\alpha M_t(f) \nabla_{M_t(f)} R_t] \ . + \mathbb{E}[R_t S_t] \ . \quad (43)$$

Now using one more time (42), we have

$$\begin{aligned} \nabla_\alpha M_t(f) &= \nabla_\alpha \mathbb{E}\Big[f(X_t) \frac{V_t}{\mathbb{E}[V_t]}\Big] \\ &= \mathbb{E}\Big[f(X_t) \frac{\nabla_\alpha V_t}{\mathbb{E}[V_t]}\Big] - \mathbb{E}\Big[f(X_t) \frac{V_t \mathbb{E}[\nabla_\alpha V_t]}{\mathbb{E}[V_t]^2}\Big] \\ &= \mathbb{E}\Big[f(X_t) S_t \frac{V_t}{\mathbb{E}[V_t]}\Big] - M_t S_t \mathbb{E}\Big[\frac{V_t}{\mathbb{E}[V_t]}\Big] \\ &= M_t(f S_t) - M_t(f) M_t(S_t) \end{aligned} \quad (44)$$

so that we obtain (12) by incorporating (44) in (43).

Appendix B: proposition 2's proof

$$\begin{aligned} \mathbf{m}_{t,\alpha+h} &= \mathbb{E}\left\{ f(X_t) | Y_{t_1^{\alpha+h}:t_{J_t}^{\alpha+h}} \right\} && (45) \\ &= \mathbb{E}\left\{ f(X_t) \frac{\mathcal{Y}_{t,\alpha+h}}{\mathbb{E}\left\{\mathcal{Y}_{t,\alpha+h}\right\}} \right\} && (46) \\ &= \frac{\mathbb{E}\left\{ f(X_t) \frac{\mathcal{Y}_{t,\alpha+h}}{\mathcal{Y}_{t,\alpha}} \frac{\mathcal{Y}_{t,\alpha}}{\mathbb{E}\{\mathcal{Y}_{t,\alpha}\}} \right\}}{\mathbb{E}\left\{ \frac{\mathcal{Y}_{t,\alpha+h}}{\mathcal{Y}_{t,\alpha}} \frac{\mathcal{Y}_{t,\alpha}}{\mathbb{E}\{\mathcal{Y}_{t,\alpha}\}} \right\}} && (47) \end{aligned}$$

Appendix C: Probability of detection

We show in this Appendix how the probability of detection is derived. First, the radar transmits a pulse expressed as follows

$$\begin{aligned} s(t) &= \alpha(t)\cos(w_c t) && (48) \\ &= \text{Re}\{\alpha(t) e^{j w_c t}\} && (49) \end{aligned}$$

where $\alpha(t)$ is the envelope also called the transmitted pulse and w_c the carrier frequency. This pulse is modified by the process of reflection. A target is modelled as a set of elementary reflectors, each reflecting: time delayed, Doppler shift, Phase shift and attenuated version of the transmitted signal. We usually assume that the reflection process is linear and frequency independent within the bandwidth of the transmitted pulse. The return signal has the following formulation:

$$s_r(t) = G \sum_i \alpha(t-\tau_i) g_i e^{j(w_c(t-\tau_i+\frac{2\dot{r}_i}{c}t)+\theta_i)} + n(t) \qquad (50)$$

where

- g_i is the radar cross section associated to reflector i,
- θ_i is the phase shift associated to reflector i,
- $\dot{r}_i$ is the radial velocity between the antenna and the object (Doppler frequency shift),
- G: others losses heavily range dependent due to spatial spreading of energy,

Algorithm 1 Policy Gradient in POMDP via IPA

Initialize $\alpha_0 \in \Gamma$

for $k = 1$ **to** ∞ **do**

for $t = 1$ **to** T **do**

Sample $u_t \sim \nu$

Set $x_t = F(x_{t-1}, u_t)$,

If $t = t_n$, sample $y_n \sim g(., x_t, a_n)\lambda(.)$

Set $s_t = \begin{cases} s_{t-1} + \frac{\frac{\partial g}{\partial \alpha}(x_t, y_n, a_n)}{g(x_t, y_n, a_n)} & \text{if } t = t_n \\ s_{t-1} & \text{else} \end{cases}$

Set $\forall i \in \{1 \ldots, I\}$

$\tilde{x}_t^{(i)} = F(x_{t-1}^{(i)}, a_{t-1}, u_t^{(i)})$ where $u^{(i)} \overset{iid}{\sim} \nu$

$\tilde{s}_t^{(i)} = \begin{cases} s_{t-1}^i + \frac{\frac{\partial g}{\partial \alpha}(x_t^i, y_n, a_n)}{g(x_t^{(i)}, y_n, a_n)} & \text{if } t = t_n \\ s_{t-1}^i & \text{else} \end{cases}$

$\tilde{w}_t^{(i)} = \begin{cases} \frac{g(x_t^{(i)}, y_n, a_n)\tilde{w}_{t-1}^{(i)}}{\sum_j g(x_t^{(j)}, y_n, a_n)\tilde{w}_{t-1}^{(j)}} & \text{if } t = t_n \\ \tilde{w}_{t-1}^{(i)} & \text{else} \end{cases}$

Set $(x_t^{(i)}, s_t^{(i)})_{i \in \{1,\ldots,I\}} = (\tilde{x}_t^{(i)}, \tilde{s}_t^{(i)})_{i \in \{k_1,\ldots,k_I\}}$, $k_{1:I}$ are selection indices associated to $(\tilde{w}^{(i)})_{i \in \{1,\ldots,I\}}$,

$m_t(f) = \frac{1}{I}\sum_i f(x_t^{(i)})$, $m_t(s_t) = \frac{1}{I}\sum_i s^{(i)}$, $m_t(fs_t) = \frac{1}{I}\sum_i f(x_t^{(i)}) s_t^{(i)}$,

$a_{n+1} = \pi_{\alpha_k}(m_t)$ if $t = t_n$

$r_t = R(x_t, m_t(f))$

$\nabla r_t = (m_t(fs_t) - m_t(f)m_t(s_t))\frac{\partial R}{\partial m_t(f)}(x_t, m_t(f)) + r_t s_t$

$\nabla J_{\alpha_k} = \nabla J_{\alpha_k} + \nabla r_t$

end for

$\alpha_{k+1} = \alpha_k + \eta_k \nabla J_{\alpha_k}$

end for

Algorithm 2 Finite Difference Stochastic Approximation for Sensor Management

Initialize α_0,

For $l = 1, \ldots, L$

- Generate a trajectory $X^k_{\tilde{t}_1:\tilde{t}_I}$
- Initialize the set of particles: $X^{(n)}_{\tilde{t}_1} \sim p(X_{\tilde{t}_1})$
- Initialize the weights of particles: $w^{(n)}_{\tilde{t}_1} = \frac{1}{N}$
- $\mathbf{m}^k_{t_1} \approx \sum_{n=1}^N f(X^{(n)}_{t_1}) w^{(n)}_{t_1}$
- First action $a_{t_1^{\alpha_l}} = \pi_{\alpha_l}(\mathbf{m}^k_{\tilde{t}_1})$
- $a_{t_1^{\alpha_l+h}} = \pi_{\alpha_l+h}(\mathbf{m}^k_{\tilde{t}_1})$
- Initialize : $\mathcal{Y}^{(n)}_{\tilde{t}_1} = 1$
- Initialize : $\mathcal{Z}^{(n)}_{\tilde{t}_1} = 1$
- For $i = 2 : I$
  –
- Estimation
  – $\mathbf{m}^k_{t_1} \approx \sum_{n=1}^N f(X^{(n)}_{t_1}) w^{(n)}_{t_1}$
  – $\nabla_{\alpha_l} \mathbf{m}^k_{t_1} \approx \sum_{n=1}^N f(X^{(n)}_{t_i}) \mathbf{s}^{(n)}_{t_i} w^{(n)}_{t_i} - \hat{\mathbf{m}}^k_{t_i} \sum_{n=1}^N \mathbf{s}^{(n)}_{t_i} w^{(n)}_{t_i}$
- Compute initial time of action t_1^{α}
- Compute initial time of action $t_1^{\alpha+h}$

- $n(t)$ is a thermal noise of the receiver such that $\text{Re}\{n(t)\}, \text{Im}\{n(t)\} \sim \mathcal{N}(0, \sigma_n^2)$.

We make the following approximations:

$$\begin{cases} \dot{r}_i \approx \dot{r} \\ \alpha(t-\tau_i) \approx \alpha(t-\tau) \end{cases} \tag{51}$$

where $\dot{r}$ is the mean radial velocity of the target τ is the mean time delay of the target. Using these approximations, the return signal can be rewritten as follows:

$$s_r(t) = \alpha(t-\tau) G e^{j w_D t} b + n(t) \tag{52}$$

where

$$\begin{cases} w_D & = w_c(1 + \frac{2\dot{r}_i}{c}) \\ b & = \sum_i g_i e^{j(-w_c \tau_i + \theta_i)} \end{cases} . \tag{53}$$

The fluctuations of b are known and modelled using Swerling 1 model [13]. There are differents models availables (Swerling 1, 2, 3,...) corresponding to different types of targets. Swerling 1 given below is convenient for aircrafts. We can then write :

$$\text{Re}\{b\}, \text{Im}\{b\} \sim \mathcal{N}(0, \sigma_{RCS}^2) . \tag{54}$$

This modelling of b assumes that the phase shifts θ_i are independent and uniformly distributed and the magnitudes g_i are identically distributed. If the number of reflector is large, the central limit theorem gives that b is a complex-valued Gaussian random variable centered at the origin. Now, a matching filter is applied to our return signal

$$s_m(t) = \int_{-\infty}^{+\infty} s_r(t) h(s) ds \tag{55}$$

where $h(t)$ is a shifted, scaled and reversed copy of $s_r(t)$

$$h(s) = \alpha(\delta - t) e^{-j w_D (\delta - t)} . \tag{56}$$

We choose $t = \delta + \tau$ which yields the best signal to noise ratio where δ is the length of the transmitted pulse. The probability of detection is based on quantity $|s_m(\delta+\tau)|^2$. We can show that

$$s_m(\delta+\tau) = G e^{j w_D \tau} b + \int_{-\infty}^{+\infty} n(\delta + \tau - s) h(s) ds . \tag{57}$$

One can remark that $s_m(\delta+\tau)$ is the sum of two complex-value Gaussian variables. We look at the following statistic

$$\Lambda = \frac{|s_m(\delta+\tau)|^2}{2\sigma_n^2} \tag{58}$$

and we introduce the following notation

$$\sigma_s^2 = G^2 \sigma_{RCS}^2 \tag{59}$$

Now we construct the test

$$\begin{cases} \mathcal{H}_1 : \text{data generated by signal + noise} \\ \mathcal{H}_0 : \text{data generated by noise} \end{cases} \tag{60}$$

$$\begin{cases} \mathcal{H}_1 : p_\Lambda(x) = \frac{1}{\frac{\sigma_s^2}{\sigma_n^2}+1} e^{-\frac{x}{\frac{\sigma_s^2}{\sigma_n^2}+1}} \\ \mathcal{H}_0 : p_\Lambda(x) = e^{-x} \end{cases} \tag{61}$$

Then, we derive the probability of detection and false alarm.

$$\begin{cases} P_d = \int_\gamma^{+\infty} p_\Lambda(x|\mathcal{H}_1 \text{ is true}) = e^{-\frac{\gamma}{\frac{\sigma_s^2}{\sigma_n^2}+1}} \\ P_{fa} = \int_\gamma^{+\infty} p_\Lambda(x|\mathcal{H}_0 \text{ is true }) = e^{-\gamma} \end{cases} \tag{62}$$

Consequently

$$P_d = P_{fa}^{\frac{1}{\frac{\sigma_s^2}{\sigma_n^2}+1}} \tag{63}$$

The ratio $\frac{\sigma_s^2}{\sigma_n^2}$ is called the Signal-to-Noise Ration noted ρ. This SNR is related to the parameters of the system and the target.

Appendix D: Radar equation

We show in this section the link between the SNR and the parameters of the system and the target. It seems that there are different possible equations. The one used by [6] do not introduce the length of the transmitted pulse which is an important parameter.

However, it introduces reduction of gain related to the deviation of the beam which will be also an important factor in our analysis.

We show here that ρ is a function of the target x_t, the time of illumination δ_t and the direction of the beam θ_t. The classical radar equation is given by the following formula:

$$\rho = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 r^4} \tag{64}$$

where P_t is the energy of the transmitted pulse, G_t is the gain of the transmitted antenna, G_r is the gain of the received antenna, σ is the radar cross section (for an aircraft between 0.1 and 1 m^2), r is the target range, γ is the system noise temperature and L is a general loss term. However, the above formula does not take into account for the sake of simplicity the losses due to atmospheric attenuation and to the imperfection of the radar. Thus , extra terms must be added [16]

$$\rho = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 k b L \gamma r^4} \tag{65}$$

where b is the receiver noise bandwith (generally consider equal to the signal bandwidth so that $b = \frac{1}{\delta_t}$), k is Boltzmann's constant, γ is the temperature of the system and L some losses. Moreover, the gain reduces with the deviation of the beam from the antenna normal in an array antenna.

$$G_t \quad = G_0 cos^{\alpha}(\theta_t) \ , \tag{66}$$

$$G_r \quad = G_0 cos^{\alpha}(\theta_t) \tag{67}$$

where G_0 is the gain of the antenna. In [16], $\alpha = 2$, in [6], $\alpha = 2.7$. According [17], there is also a beam loss because the radar beam is not pointing directly so that the radar equation is:

$$\rho = \frac{P_t G_0^2 \lambda^2 \sigma \delta_t \cos^2(\theta_t)}{(4\pi)^3 k L \gamma r^4} e^{-\frac{(\theta_t - \beta_t)^2}{2B^2}} \tag{68}$$

where is B is the beamwidth.

References

[1] C. Kreucher, D. Blatt, A. Hero, and K. Kastella. Adaptive multi-modality sensor scheduling for detection and tracking of smart targets. *Digital Signal Processing*, 16(5):546–567, September 2005.

[2] M. Shakeri, K.R. Pattipati, and D.L. Kleinman. Optimal Measurements Scheduling for State Estimation. IEEE *Trans. on Aerospace and Electronic Systems*, 31(2):716–729, april 1995.

[3] Y. Li, L.W. Krakow, E.K.P. Chong, and K.N. Groom. Dynamic sensor management for multisensor multitarget tracking. In *Proc. of the 40th Annual Conference on Information Sciences and Systems*, 2006.

[4] M.L. Hernandez, T. Kirubarajan, and Y. Bar-Shalom. Multi-sensor resource deployment using posterior cram-rao bounds. IEEE *Trans. on Aerospace and Electronic Systems*, 2(40):399–416, 2004.

[5] S.S. Singh, N. Kantas, B-.N. Vo, A. Doucet, and R.J. Evans. Simulation-based optimal sensor scheduling with application to observer trajectory planning. *Automatica*, 43(5):817–830, 2007.

[6] J. Wintenby. *Resource Allocation in Airborn Surveillance Radar*. PhD thesis, Chalmers University of Technology, 2003.

[7] P.D. Blair and W.D. Blair. Optimal Phased Array Radar Beam Pointing for MTT. In IEEE *Aerospace Conference Proceedings*, July 2004.

[8] B. La Scala, W. Moran, and R.J. Evans. Optimal Adaptive Waveform Selection for Target Detection. In *Radar 2003, Adelaide, Australia*, September 2003.

[9] L. Meier, J. Peschon, and R.M Dressler. Optimal control of measurement subsystems. IEEE *Trans. on Aut. Contr.*, 12(5):528–536, 1967.

[10] A. Doucet, S. Godsill, and C. Andrieu. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Technical report, Cambridge University Engineering Department, 2000.

[11] P.A. Coquelin, R. Deguest, and R. Munos. Numerical methods for sensitivity analysis of feynman-kac models. Technical report, INRIA-Futurs, 2007.

[12] P. Del Moral. *Feynman-Kac Formulae Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[13] G.sR. Curry. *Radar System Performance Modeling, Second Edition*. Artech House, 2005.

[14] X. Rong Li and V. Jilkov. A Survey of Maneuvering Target Tracking Part I: Dynamics Models. IEEE *Trans. on Aerospace and Electronic Systems*, 39(4):1333–1364, October 2003.

[15] J-.P. Le Cadre and O. Tremois. Bearings-only tracking for maneuvering sources. IEEE *Trans. on Aerospace and Electronic Systems*, 34(1):179–193, January 1998.

[16] M. De Vilmorin. *Contribution à la grstion optimale de capteurs: application à la tenue de situations aériennes*. PhD thesis, Ecole Centrale de Lille et Université des Sciences et Technologie de Lille, 2002.

[17] G. Van Keuk and S.S. Blackman. On Phased-Array Radar Tracking and Parameter Control. IEEE *Trans. on Aerospace and Electronic Systems*, 1(29):186–194, January 1993.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399