Optimal Policy for Scheduling of Gauss-Markov Systems

Stephen Howard
DSTO,EWRD
PO Box 1500
Edinburgh, 5111
Australia
stephen.howard@defence.dsto.gov.au

Sofia Suvorova
DSTO, EWRD
PO Box 1500
Edinburgh, 5111
Australia
sofia.suvorova@defence.dsto.gov.au

Bill Moran
Dept. of EEE
University of Melbourne
Parkville, 3010
Australia
b.moran@ee.mu.oz.au

Abstract – *This paper considers the problem of multi-step scheduling the measurement of several Gauss-Markov systems. Such a problem occurs, for example, in the allocation of the time of a single beam phased-array radar system among several targets. The problem is put into a mathematical framework and is solved for the case of two one-dimensional systems for which the cost function is the sum of the variances of the estimates of the states of the two systems.*

Keywords: multi-step schedule, Gauss-Markov, dynamic programming, adaptive phased-array radar

1 Introduction

Modern radar systems with phased array antennas have the capacity to switch beam patterns and direction, waveforms, pulse repetition intervals (PRIs) and other modes on a PRI by PRI basis. When these modes affect the performance on the transmit side of the system, it is important to select the most appropriate mode of the radar to optimally perform its function within its present environment.

For such scheduling problems, it is necessary to have a cost function of the mode of the radar (PRI, waveform, beam pattern and direction, etc) that reflects the function that the radar has to perform, as well as some measure of the current environment. The simplest method of scheduling such a system is the so-called "greedy" approach in which the mode is chosen for the next PRI so as to optimize this cost function at that PRI. If the system is attempting to track a target (or targets) such a cost function might be expressible in terms of the track error covariance matrix. This technique is explored by Evans and Kershaw in [1, 2]. There the waveforms are scheduled to optimize the track error covariance. Their approach is a *greedy* one, in that the schedule seeks only to choose appropriate waveform to optimize the value of the cost function at the next PRI.

This greedy approach is not necessarily the best that can be done if it is possible to perform scheduling over multiple PRIs. Various approaches to this problem have been attempted, mostly based on a hidden markov model (HMM) structure for the dynamics of the system (see [3, 4, 5, 6]). While partial solutions are available in this context, none is entirely satisfactory. All known methods that seek optimality or near-optimality are computationally intensive. It is clear that much remains to be understood here.

Here we attempt to address the problem of steering a beam (as do [4, 3], though as we have said they take a quite different approach). Our aim is to discuss optimal scheduling algorithms for sensor tracking systems for multiple targets. The motivation for this work came from a radar context, where several targets are being tracked by a phased array system. The radar is capable of switching attention among the targets. The goal is to devise a schedule (policy) to do this in such a way as to optimize a measure of track error. While this would seem considerably simpler than the problem of switching the many other radar parameters in addition to beam direction, it is relatively easy to see using the concept of a "virtual target" that the problems are essentially the same. As we have already stated, various methodologies for this are considered in the literature.

In our approach the targets are assumed to have a simple dynamical model. We write $\mathbf{y}_n^{(i)}$ for the state of the ith system (target) at epoch n, and $\mathbf{z}_n$ for one measurement of the system at that epoch. We note here that only one system is observed at any one epoch, so that there may be no measurement for a specific system at a given epoch. The states $\mathbf{y}^{(n)}$ for each system satisfy a non-interacting dynamical model, namely the Gauss-Markov system:

$$\begin{aligned} \mathbf{y}_{n+1}^{(i)} &= F\mathbf{y}_n^{(i)} + \mathbf{w}_n^{(i)} \\ \mathbf{z}_n &= \sum_i \delta(u_n, i)\big(H\mathbf{y}_n^{(i)} + \mathbf{v}_n^{(i)}\big). \end{aligned} \tag{1}$$

Here $\mathbf{y}_n^{(i)}$ is a member of $\mathbf{R}^N$, $\mathbf{z}_n$ is a member of $\mathbf{R}^M$, F and H are $N \times N$ and $M \times N$ matrices, representing the dynamics and the measurement process, and $\mathbf{w}_n^{(i)}$ and $\mathbf{v}_n^{(i)}$ are independent Gaussian processes. It is assumed that there are T targets. A *policy* is a sequence $\mathbf{u} = (u_n)$ of integers with values in the range $[1, T]$. If $u_n = i$, then the ith target is observed on the nth PRI. This is reflected in Eq. (1) by the function δ which has the meaning:

$$\delta(u, i) = \begin{cases} 1 & \text{if u=i} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

When the policy does not select to measure a target the result is zero. From the perspective of the tracker, noise is added as a result of the noisy dynamics, but not compensated by a measurement.

An appropriate cost function is expressed in terms of the track error covariance matrix $P_n^{(i)}$ as produced by the Kalman filter for the ith target. Specifically,

$$J_N(\mathbf{u}) = \sum_{n=1}^{N} \sum_i h(P_n^{(i)}), \tag{3}$$

where h is some appropriate function on the set of positive definite matrices and taking positive real values. Two possibilities for h are the trace and the determinant. The problem is then to choose an optimal policy $(u_1, u_2, \ldots, u_N)$ as a function only of the initial state $\mathbf{y}_0$ that minimizes the cost function $J_N(\mathbf{u})$.

This problem appears very difficult, and here we restrict attention just to the one-dimensional case. In this case there is no loss of generality in making F and H the identity matrix. Thus the system becomes

$$y_{n+1}^{(i)} = y_n^{(i)} + w_n^{(i)} \tag{4}$$

$$z_n = \sum_i \delta(u_n, i)\big(y_n^{(i)} + v_n^{(i)}\big). \tag{5}$$

Such a case might exist, for example, where range only is under consideration. However, our main interest is in understanding the mathematics of this simple case to guide research on the more general problem.

For simplicity too, for the most part we consider only the case when there are just two systems under observation. For two systems it is possible to visualize the dynamics of the systems. The restriction to two systems does not appear to be serious and it seems to be relatively easy to extend the results to more than two.

The key theorem is that for the systems we are interested in in this paper the optimal solution is the greedy algorithm. In other words if, at the nth epoch, we choose to observe the system which optimizes the component of the cost $\sum_i h(P_n^{(i)})$, then this will optimize the total cost over any finite horizon starting at time n. This somewhat surprising result has important consequences for practical scheduling if it persists to more complex and realistic systems. It would say that there is no need to develop complicated and computationally intensive algorithms for scheduling such systems. We do not believe, however, that the result does persist. Nonetheless it is important to understand at what point the greedy approach ceases to be optimal and this paper represents a contribution to our understanding of that.

2 Problem Formulation

As already stated, we assume two state space models

$$y_{n+1}^{(i)} = y_n^{(i)} + w_n^{(i)} \tag{6}$$

$$z_n = \sum_i \delta(u_n, i)\big(y_n^{(i)} + v_n^{(i)}\big), \tag{7}$$

where the noise processes $(w_n^{(i)})_n$ and $(v_n^{(i)})_n$ $(i = 1, 2)$ are independent, white and Gaussian with

$$E[w^{(i)} w^{(i)T}] = \Sigma_i^2$$

and

$$E[v^{(i)} v^{(i)T}] = R_i^2.$$

We assume that these processes are tracked using a Kalman filter to update the state estimates whenever a measurement becomes available. The evolution of the tracking process can be completely expressed, of course, in terms of the state estimate and the filtered state error covariance. Since the cost function J only involves the latter, it is its dynamics that are important for our problem.

For each i, the filtered state error covariance evolves as (dropping the superscript i for the moment),

$$\sigma_n^2 = \sigma_{n-1}^2 + \Sigma^2 \tag{8}$$

if no measurement is taken of the ith system at time n; that is, if $u_n \neq i$. If a measurement is taken the Kalman update equation yields

$$\sigma_n^2 = (\sigma_{n-1}^2 + \Sigma^2)\Big(1 + \frac{\Sigma^2}{R^2} + \frac{\sigma_{n-1}^2}{R^2}\Big)^{-1} \tag{9}$$

for the evolution of the error covariance. Note that Eq. (8) is just Eq. (9) with $R = \infty$.

Our goal is to minimize the sum of the track error variances for each of the two systems

$$\sum_{n=1}^{N} (\sigma_n^{2(1)} + \sigma_n^{2(2)}) \tag{10}$$

over some finite horizon of length N, and over the set of all policies $\mathbf{u} = (u_1, u_2, \ldots, u_{N-1})$, where each u_i takes the value 1 or 2. This denotes the choice of system being measured at each epoch.

Now we can consider the problem as that of determining the optimal control of a controlled deterministic dynamical system. It is convenient to change the notation (if only to eliminate the exponent 2):

$$x_n^{(i)} = \begin{cases} x_{n-1}^{(i)} + \Sigma^2, & \text{if } u_n \neq i \\ (x_{n-1}^{(i)} + \Sigma^2)(1 + \frac{\Sigma^2}{R^2} + \frac{x_{n-1}^{(i)}}{R^2})^{-1} & \text{if } u_n = i, \end{cases} \tag{11}$$

where we have defined

$$x_n^{(i)} = {\sigma_n^{(i)}}^2. \tag{12}$$

Further simplification of the mathematics can be achieved by replacing x by $x\Sigma^2$, noting that $\Sigma^2 = 0$ would be a degenerate case in which there would be no point in observing the system since its dynamics are then completely determined. The equations now simplify to

$$x_n^{(i)} = \begin{cases} x_{n-1}^{(i)} + 1, & \text{if } u_n \neq i \\ (x_{n-1}^{(i)} + 1)(1 + c + c x_{n-1}^{(i)})^{-1} & \text{if } u_n = i, \end{cases} \tag{13}$$

where $c = \frac{\Sigma^2}{R^2}$. Observe that the non-measurement case now corresponds to $c = 0$.

It is convenient to express this dynamical system in terms of maps of $\mathbf{R}^+$ to itself. We define

$$\begin{aligned} S(x) &= x + 1 \\ T(x) &= \frac{x+1}{1 + c + cx}. \end{aligned} \tag{14}$$

These two transformations are *linear fractional transformations*; that is, transformations of the form $V(x) = \frac{ax+b}{cx+d}$. More particularly they each have determinant $(ad - bc)$ equal to 1. It follows from the linear fractional transformation property that composition of the transformations can be calculated in terms of the corresponding matrices:

$$\begin{aligned} S &\sim \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ T &\sim \begin{pmatrix} 1 & 1 \\ c & c+1 \end{pmatrix} \end{aligned} \tag{15}$$

We shall be interested in multiple compositions of the two maps S and T of the form $S^{w_1}T^{w_2}S^{w_3}\ldots T^{w_m}$ where w_k are non-negative integers. The resulting map on $\mathbf{R}^+$ is also a linear fractional transformation whose matrix is the corresponding product of matrices.

It should be observed, too, that T maps any point x in $\mathbf{R}^+$ to the (half-open) interval $[\frac{1}{c+1}, \frac{1}{c})$ whereas, of course, S shifts x to the right by 1. In fact T is a contraction on $\mathbf{R}^+$ (with contraction coefficient $1/(c+1)^2$) and therefore has a unique fixed point

$$x_T = \frac{1}{2}\left(\sqrt{1+\frac{4}{c}} - 1\right) \tag{16}$$

in $\mathbf{R}^+$. In other words, continued observation of one system will result in a variance that approaches this value.

We note too that

$$ST(x) = \frac{(1+c)x + 2 + c}{cx + c + 1}$$

and

$$TS = \frac{x+2}{cx+2c+1}$$

are also contractions with fixed points

$$\begin{aligned} x_{ST} &= \sqrt{\frac{2}{c}+1} \\ x_{TS} &= \sqrt{\frac{2}{c}+1} - 1. \end{aligned} \tag{17}$$

Now we define two maps $U^{(1)}, U^{(2)} : \mathbf{R}^2 \to \mathbf{R}^2$ by

$$\begin{aligned} U^{(1)}(x^{(1)}, x^{(2)}) &= (T(x^{(1)}), S(x^{(2)})) \\ U^{(2)}(x^{(1)}, x^{(2)}) &= (S(x^{(1)}), T(x^{(2)})) \end{aligned} \tag{18}$$

Writing $\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)})$, we see that the dynamics are given by

$$\mathbf{x}_{n+1} = U^{(i)}(\mathbf{x}_n) \tag{19}$$

if system $i = 1, 2$ is observed at time n. The cost for a given policy $\mathbf{u} = (u_1, u_2, \ldots, u_N)$ and initial state $\mathbf{x}_0$ is now given by

$$J(\mathbf{x}_0, \mathbf{u}) = \sum_{n=1}^{N} f(U^{(u_n)}U^{(u_{n-1})}\cdots U^{(u_1)}\mathbf{x}_0), \tag{20}$$

where $f(\mathbf{x}) = x^{(1)} + x^{(2)}$.

An interesting generalization of the problem would replace S by T with a different value of c (other than 0). This would correspond to a situation of two different measurement processes with different noise on each process and with the capability of switching their use between the two targets.

3 Dynamic Programming Solution

We describe here the solution to this problem. No proofs of the major results will be given; these will be published in a forthcoming paper. We note first that a state of the controlled deterministic dynamical system is describable in terms of the pair $\mathbf{x} = (x^{(1)}, x^{(2)})$ and that, as we have already stated, the dynamics is given in terms of the transformations $U^{(1)}$ and $U^{(2)}$.

We begin by stating the theorem. To do so we define

$$\text{ind}(\mathbf{x}) = \begin{cases} 1 & \text{if } x^{(1)} > x^{(2)} \\ 2 & \text{otherwise.} \end{cases} \tag{21}$$

Theorem 1 *For the problem described in Section 2 the optimal control is*

$$u_n = \text{ind}(\mathbf{x}_\text{n}) \tag{22}$$

for $n = 1 \cdots N-1$.

It may be claimed that this does not produce the solution we said we needed, since the choice of policy at epoch n is determined by the value of $\mathbf{x}_n$. However, the dynamics of the system are entirely known, so we know in advance where $\mathbf{x}_n$ is, and so the policy can be determined in advance. In fact, in our problem the dynamics permit us to describe the optimal solution in advance in very simple terms.

To do this, we need to define certain regions of the plane. First we describe a collection of hyperbolae

$$\begin{aligned} \mathcal{H}_k^{(1)} &= \{\mathbf{x} : S^k(x^{(2)}) = T^k(x^{(1)})\} \qquad (k = 0, 1, 2, \ldots) \\ \mathcal{H}_k^{(2)} &= \{\mathbf{x} : T^k(x^{(2)}) = S^k(x^{(1)})\} \qquad (k = 0, 1, 2, \ldots). \end{aligned} \tag{23}$$

Note that these are repeated inverse images of the line $x^{(1)} = x^{(2)}$ under the maps $U^{(1)}$ and $U^{(2)}$, respectively. Note that they are rectangular hyperbolae with asymptotes parallel to the coordinate axes.

The region between the hyperbolae $\mathcal{H}(i)_{k+1}$ and $\mathcal{H}_k^{(i)}$, we write as $A_k^{(i)}$, $(k = 0, 1, 2, \ldots, i = 1, 2)$. Fig. 1 illustrates these regions for $c = 0.1$ and $k = 0, \cdots, 3$. These regions tile the positive quadrant of the plane, as can be seen. In fact, for some values of c some of the hyperbolae will not even intersect the positive quadrant and then the tiling is degenerate. Note that the structure of the regions is, of course, symmetric about the diagonal line $\mathcal{H}_0^{(1)} = \mathcal{H}_0^{(2)}$.

We need to make one further observation before we can describe the dynamics associated with the optimal solution. This is that the image of $A_0^{(i)}$ under $U^{(i)}$ is contained in $A_0^{(j)}$ (here we use the convention that $j \neq i$). It should be noted that in a more general context this may not be the case. To

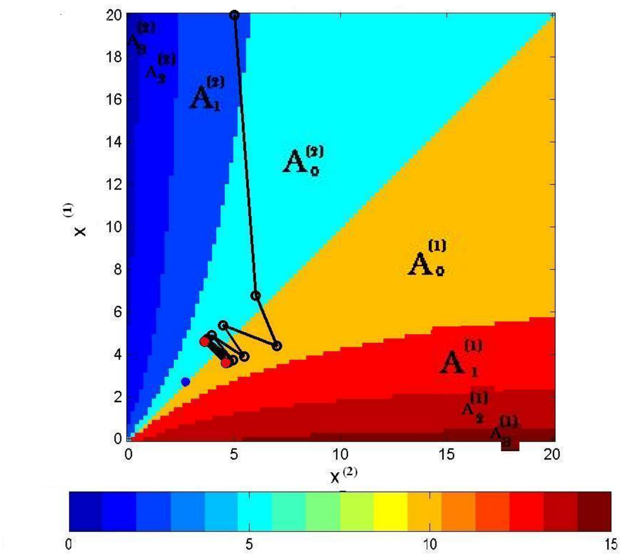

Fig. 1: The Regions $A_k^{1,2}, k = 0, \cdots, 3$ for the Case $c = 0.1$

see that this is true here, assume that $(x_0, y_0) \in A_0^{(1)}$; that is,

$$\begin{aligned} x_0 &> y_0, \\ T(x_0) &< S(y_0). \end{aligned} \tag{24}$$

If we apply $U^{(1)}$ to it we obtain $(x_1, y_1) = (T(x_0), S(y_0))$ In order to prove $(x_1, y_1) \in A_0^{(2)}$, we need to show that

$$\begin{aligned} y_1 &> x_1, \\ T(y_1) &< S(x_1). \end{aligned} \tag{25}$$

We note that $S - T$ is positive (*it is always better to measure*), as is $ST - TS$, since

$$(ST{-}TS)(x) = \frac{c(2x + 3 + cx^2 + 3cx + 2c)}{(cx + c + 1)(cx + 2c + 1)} > 0. \tag{26}$$

(*It is always better to measure later.*) It follows that

$$y_1 - x_1 = S(y_0) - T(x_0) > 0 \tag{27}$$

and

$$S(x_1){-}T(y_1) = ST(x_0){-}TS(y_0) > (ST{-}TS)(y_0) > 0. \tag{28}$$

This gives the conclusion we need. By symmetry, it follows that $U^{(2)}(A_0^{(2)}) \subset A_0^{(1)}$.

Now we are in a position to describe the optimum policy. Suppose that we start in region $A_k^{(i)}$ ($k = 0, 1, \ldots$). In this case we apply $U^{(i)}$ $k + 1$ times. Each application moves the system one region closer to the diagonal until finally it crosses the diagonal, where it must go into $A_0^{(j)}$ in the light of what we have just proved. At this point we switch to applying $U^{(j)}$, mapping the system into $A_0^{(i)}$ again. From this point on the policy alternates between applying $U^{(i)}$ and $U^{(j)}$. Thus the optimal is to observe the system with larger variance until its variance drops below that of the other system. From that point the systems are viewed alternately.

We note too that the result of this optimal policy is that the state of the system converges (exponentially fast) to a situation where it flips between the points

$$(x_{ST}, x_{TS}) = \Big(\sqrt{\frac{2}{c} + 1}, \sqrt{\frac{2}{c} + 1} - 1\Big) \tag{29}$$

and

$$(x_{TS}, x_{ST}) = \Big(\sqrt{\frac{2}{c} + 1} - 1, \sqrt{\frac{2}{c} + 1}\Big). \tag{30}$$

In terms of the original system (equations (8) and (9)), these points become

$$\Big(\frac{1}{\Sigma}\sqrt{2R^2 + 1}, \frac{1}{\Sigma}\big(\sqrt{2R^2 + 1} - 1\big)\Big) \tag{31}$$

and

$$\Big(\frac{1}{\Sigma}\big(\sqrt{2R^2 + 1} - 1\big), \frac{1}{\Sigma}\sqrt{2R^2 + 1}\Big), \tag{32}$$

respectively.

4 Scalar Kalman Filter, M Systems

Theorem 1 can be extended to M similar systems, where we assume that only one can be measured at any one epoch.

We have

$$(x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(i)}, \cdots x_n^{(M)}) = \begin{cases} T(x_{n-1}^{(i)}), & \text{if } i\text{-th system is measured, i.e } u_n = i \\ S(x_{n-1}^{(j)}) & \text{for all systems } j \neq i, \end{cases} \tag{33}$$

where $i, j = 1, \cdots, M$ and $T(x)$, $S(x)$ are as defined in Section 2. The problem is to minimize

$$J = \sum_{n=1}^{N} \sum_{i=1}^{M} x_n^{(i)}. \tag{34}$$

In this case the optimal control at each n is

$$u_n = \text{ind}(\mathbf{x}) = \arg\min_i(x^{(i)}). \tag{35}$$

5 Conclusion

We have described an optimal policy for the scheduling of the measurement of several one-dimensional Gauss-Markov models subject to the cost function that is the sum of the variances of the individual systems. This optimal policy is exactly the greedy policy in the sense that it chooses, at each epoch, the action that reduces the component of the cost for that epoch. Moreover, it is also the policy that chooses at each epoch to measure the system with the largest variance. For two systems this results in a policy which is ultimately alternating in that it switches attention at each epoch between the two systems.

This work attacks the simplest problem in the scheduling of observation of multiple Gauss-Markov systems. More work is needed to handle higher dimensional and more realistic systems.

Acknowledgements

The authors would like to thank Professor Rob Evans and Dr Barbara La Scala for many useful discussions on this problem.

This work was supported in part by the Defense Advanced Research Projects Agency of the US Department of Defense and was monitored by the Office of Naval Research under Contracts No. N000014-02-1-0802 to the University of Melbourne and No. N00014-02-C-0323 to Prometheus Inc. of Rhode Island. The United States government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

The work was also supported by the Electronic Warfare and Radar Division of the Australian Defence Science and Technology Organization.

References

[1] D.J. Kershaw and R.J. Evans. Optimal waveform selection for tracking systems. *IEEE Transactions on Information Theory*, 40(5):1536–50, September 1994.

[2] D.J. Kershaw and R.J. Evans. Waveform selective probabilistic data association. *IEEE Aerospace and Electronic Systems*, 33(4):1180–88, October 1997.

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

[4] V. Krishnamurthy and R. J. Evans. Correction to "Hidden Markov model multiarm bandits: A methodology for beam scheduling in multitarget tracking. *IEEE Trans. on Signal Processing*, 51(6):1662–1663, 2003.

[5] V. Krishnamurthy. A value iteration algorithm for partially observed Markov decision process multi-armed bandits. Submitted to IEEE Trans. on Automatic Control, 2001.

[6] B. F. La Scala, W. Moran, and R. J. Evans. Optimal scheduling for target detection with agile beam radars. In *Proc. of the NATO SET-059 Symposium on Target Tracking and Sensor Data Fusion for Military Observation Systems*, Budapest, Hungary, 2003.