# Adaptive Sampling and Actuation for POMDPs: Application to Precision Agriculture

D.J. Antunes, R.M. Beumer, M.J.G. van de Molengraft, W.P.M.H. Heemels

***Abstract*— Given a partially observable Markov decision process (POMDP) with finite state, input and measurement spaces, and costly measurements and control, we consider the problem of when to sample and actuate. Both sampling and actuation are modeled as control actions in a framework encompassing estimation and intervention problems. The process evolves freely between two consecutive control action times. Control actions are assumed to reset the conditional distribution of the state given the measurements to one of a finite number of distributions. We tackle the problem of deciding when control actions should occur in order to minimize an average cost that penalizes states and the rate of control actions. The problem is first shown to boil down to a stopping time problem. While the latter can be solved optimally, the complexity of the optimal policy is intractable. Thus, we propose two approximate methods. The first is inspired by relaxed dynamic programming, and it is within an additive cost factor of the optimal policy. The second is inspired by consistent event-triggered control and ensures that the cost is smaller than that of periodic control for the same control rate. We conclude that the latter policy can deal with large dimensional problems, as demonstrated in the context of precision agriculture.**

## I. Introduction

There are many control and estimation applications where taking actions or intervening in an otherwise freely evolving process is costly but necessary. Thus, the times of these interventions must be carefully selected, possibly based on available process information. For instance, in remote monitoring, limited sensors can work as a proxy to detect events that need to be confirmed by closer inspection and eventually handled. This is the case in precision farming, where diseases, water stress, nitrogen levels, or weed can be inferred, to some extent, from a few sensors placed in the field. However, costly farmer or (aerial) robot inspections are required for complete situational awareness and handling. For instance, in [1], [2], when to irrigate depends on soil moisture and temperature sensors, and in [3], when to apply nitrogen fertilizers depends on in-situ active-light reflectance measurements. Similar challenges arise in queuing control, predictive maintenance, stock trading, and home surveillance.

These and related problems have been studied in several fields, primarily considering finite (or countable) state, input, and measurement spaces, leading to partially observable Markov decision processes (POMPDs). It is well-known that these problems are intractable; thus finding appropriate (close to optimal) strategies remains challenging. In turn, much research has recently been carried out on event-triggered control (ETC) that concerns choosing the times to sample or actuate in a control loop based on states or events rather than periodically. Since in ETC, state, input, and measurement spaces are typically continuous, the literature that considers finite spaces is scarce. Still, [4], [5], and [6] aim at finding control policies for POMDPs considering finite spaces that depend on events, defined in terms of given state transitions; events are pre-defined, whereas in many ETC papers [7]–[10], as in the present paper, events are to be scheduled. This latter approach is followed in a different research line proposed in [11], [12] and [13], also considering POMDPs with finite spaces. However, [11], [12] propose to decide the next sampling (or control) time based on the information up to the current sampling time and not in between the two; this parallels self-triggered control. Differently, ETC, considered here, continuously monitors the state or output of the process to decide the next sampling or control time.

The present paper considers POMDPs with finite state, input, and measurement spaces. Control actions model both costly sampling through information gathering and costly actuation through process intervention and are triggered at only a subset of possible discrete times. The process evolves freely in between control action times. Control actions are assumed to reset the conditional distribution of the state given the measurements to one of a finite number of distributions. In this sense, the effect of control actions is known, and only *when* control actions should be enforced is to be determined. In fact, we tackle the problem of deciding when these control actions should occur in order to minimize an average cost penalizing state configurations and the number of control actions. We show that the average cost problem can be tackled as a stopping time problem. While an optimal policy can be obtained for this latter problem, its complexity is intractable. Thus, we propose approximate policies.

First, we propose a class of policies inspired by relaxed dynamic programming (RDP) [14]. These policies guarantee a cost within an additive constant factor of the cost of the optimal policy, rather than a multiplicative factor as in original RDP [14]; this is needed for the stopping time problem at hand since the cost can be negative. While the approximate policy's complexity is far smaller than that of the optimal policy, and it provides nearly optimal results when the state dimension is small, it becomes impractical when the state dimension is large (see example in Section VI).

Second, inspired by [9], [10], we propose a class of so-called consistent policies that lead to a strictly smaller

The authors are with the Control Systems Technology Group, Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands. E-mails:{d.antunes, r.m.beumer, w.p.m.h.heemels, m.j.g.v.d.molengraft}@tue.nl. This research is part of the research program SYNERGIA (project number 17626), which is partly financed by the Dutch Research Council (NWO).

cost than that of periodic inspection for the same average inspection rate. The policies proposed here are different from the ones in [9], [10], both in form and derivation, to account for the case that the state probability distribution resets to one of a set of possible distributions rather than a single one; examples of applications where this arises are discussed.

The applicability of the results is highlighted by a numerical case study in the context of precision farming. The proposed policies rely on limited remote sensors indicating weed presence to decide the timings of weed removal. Due to space discretization the state dimension is rather large. Unlike RDP, the policy inspired by consistent ETC can handle problems with a large state dimension. This shows that ideas inspired by the ETC literature can help determine when to sample and actuate a POMDP.

The remainder of the paper is organized as follows. Section II provides the problem formulation and some applications. Section III provides the optimal policy and explains the intractability issue. Sections IV and V provide the approximate policies and main results based on RDP and the consistent policies, respectively. Simulation results are discussed in Section VI and concluding remarks in Section VII. The proofs of the results are omitted.

## II. Problem formulation

Consider a dynamical system

$$\begin{aligned} x_{t+1} &= \underline{f}(x_t, u_t, w_t) \\ y_t &= \underline{h}(x_t, u_t, v_t) \end{aligned} \tag{1}$$

where $x_t \in \{1, 2, \dots, n\}$, $u_t \in \{1, 2, \dots, n_u\}$, $y_t \in \{1, 2, \dots, n_y\}$, $w_t \in \{1, 2, \dots, n_w\}$, $v_t \in \{1, 2, \dots, n_v\}$ are the state, control input, measurement, process disturbance input, and measurement noise input at time $t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, respectively. The disturbance sequences $\{w_t | t \in \mathbb{N}_0\}$ and $\{v_t | t \in \mathbb{N}_0\}$ are assumed to be independent and identically distributed disturbance sequences (i.i.d.), which are also mutually independent. The fact that the output in (1) depends on $u_t$ allows for tackling sensor management problems [15]. Consider also the average cost

$$J_s = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\underline{g}(x_t, u_t)], \tag{2}$$

which exists under assumptions provided shortly. The control rate is assumed to be costly, and thus $u_t$ can only be decided upon at so-called control times $s_\ell \in \mathbb{N}_0$, $s_{\ell+1} = s_\ell + \tau_\ell$, with $\tau_\ell \in \mathbb{N}$, $\forall \ell \in \mathbb{N}_0$. For convenience, the intervals between control times are assumed to be bounded: $\tau_\ell \le \bar{h}$, $\forall \ell \in \mathbb{N}_0$. Since $\bar{h} \in \mathbb{N}$ can be arbitrarily large, this assumption is not restrictive. Let $\sigma_t = 1$ if $s_\ell = t$ for some $\ell$ and $\sigma_t = 0$ otherwise. We assume that $s_0 = 0$ and, thus, $\sigma_0 = 1$. Note that $\{t \in \mathbb{N}_0 | \sigma_t = 1\} = \{s_\ell | \ell \in \mathbb{N}_0\}$. At times $t = s_\ell$, a sequence of current and future inputs $U_{0|t}$, $U_{1|t}$, …, $U_{\bar{h}-1|t}$ is computed to be applied between control times $s_\ell$, i.e.,

$$u_{t+j} = U_{j|t} \text{ for } j \in \{0, \dots, \tau_\ell - 1\} \text{ when } t = s_\ell \text{ for some } \ell.$$

Often there is one decision in the set $u_t = \underline{u} \in \{1, \dots, n_u\}$ that corresponds to a free (not controlled) mode of the system and $U_{j|s_\ell} = \underline{u}$ for $j \in \{0, \dots, \tau_\ell - 1\}$. However, more general cases can be considered.

The initial state probability distribution is denoted by $\tilde{p}_0 = [\tilde{p}_{0,1} \dots \tilde{p}_{0,n}]^\intercal$ with $\tilde{p}_{0,i} = \text{Prob}[x_0 = i]$ and $\tilde{p}_0 \in \mathcal{P}_n := \{p = [p_1 \dots p_n]^\intercal | 1_n^\intercal p = 1, p_i \ge 0, \forall i\}$, where $1_n$ denotes a column vector with $n$ entries equal to one. Let also $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{y_t, \sigma_{t-1}, u_{t-1}\}$ for $t \in \mathbb{N}$ with $\mathcal{I}_0 = \{\tilde{p}_0\} \cup \{y_0\}$ denote the information available for decisions up to time $t$ and $p_{t|t} = [p_{t|t,1} \dots p_{t|t,n}]^\intercal$ denote the probability distribution of the state $x_t$ given the information set $\mathcal{I}_t$, i.e., $p_{t|t,i} = \text{Prob}[x_t = i | \mathcal{I}_t]$, with $p_{t|t} \in \mathcal{P}_n$. Let $p_{t+1|t}$ be defined similarly but with $p_{t+1|t,i} = \text{Prob}[x_{t+1} = i | \mathcal{I}_t]$. A crucial assumption is that either $p_{t|t}$ or $p_{t+1|t}$ belongs to a known set of $b$ possible distributions, denoted by $\rho_1, \dots, \rho_b$, when actuation is computed (at control times). Note that $b$ does not depend on $t$. Formally:

*Assumption 1:* One of the following conditions holds

$$\text{(i) } p_{s_\ell | s_\ell} \in \{\rho_1, \dots, \rho_b\}, \text{ for every } \ell \in \mathbb{N}_0. \tag{3a}$$

$$\text{(ii) } p_{s_\ell + 1 | s_\ell} \in \{\rho_1, \dots, \rho_b\}, \text{ for every } \ell \in \mathbb{N}_0. \tag{3b}$$

Assumption 1(ii) captures applications where the state becomes either known or has a known probability distribution at time $t+1$ (through map $\underline{f}$) when a control intervention on the process is carried out at time $t$, while Assumption 1(i) captures applications where the control actions at time $t$ influence measurements at time $t$ through map $\underline{h}$.

Let $\phi_\ell \in \{1, \dots, b\}$ be such that $p_{s_\ell | s_\ell} = \rho_{\phi_\ell}$ or $p_{s_\ell + 1 | s_\ell} = \rho_{\phi_\ell}$, when Assumptions 1(i), 1(ii) hold respectively (if both hold either choice for $\phi_\ell$ can be picked). We can define a Markov chain with $b$ states and transition probability matrix $Q$ with entries $Q_{ij} = \text{Prob}[\phi_{t+1} = i | \phi_t = j]$. It is assumed to be ergodic, i.e., aperiodic and irreducible. Thus, it has a stationary probability distribution.

*Assumption 2:* There exists a unique $a \in \mathcal{P}_b$ such that $a = Qa$.

Due to this assumption, and since we are interested in an average cost (2), we can assume that the initial distribution of $\phi_0$ is $a$, i.e., $\text{Prob}[\phi_0 = i] = a_i$, $i \in \{1, \dots, b\}$.

A third assumption imposes that the control sequence $U_{j|s_\ell}$ only depends on $\phi_\ell$.

*Assumption 3:* $U_{j|s_\ell} = \theta(j, \phi_\ell)$ for every $j \in \{0, \dots, \bar{h} - 1\}$, every $\ell \in \mathbb{N}_0$, and for a given function $\theta$.

These assumptions are motivated by and met in the applications discussed below. Due to these assumptions, between control times, the system evolves freely as

$$x_{t+1} = f(x_t, \zeta_t, \phi_\ell, w_t), \quad s_\ell \le t \le s_{\ell+1} - 1,$$

where $\zeta_t = t - s_{\bar{\ell}(t)}$, $\bar{\ell}(t) = \max\{\ell | s_\ell \le t\}$, and $f(x_t, \zeta_t, \phi_\ell, w_t) = \underline{f}(x_t, \theta(\zeta_t, \phi_\ell), w_t)$. Likewise, letting $L(T) = \max\{\ell | s_\ell \le T - 1\}$, the average cost is rewritten as

$$J_s = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[ \sum_{\ell=0}^{L(T)-1} \sum_{t=s_\ell}^{s_{\ell+1}-1} g(x_t, \zeta_t, \phi_\ell) + \sum_{t=s_{L(T)}}^{T-1} g(x_t, \zeta_t, \phi_{L(T)}) \right], \tag{4}$$

with $g(x_t, \zeta_t, \phi_\ell) = \underline{g}(x_t, \theta(\zeta_t, \phi_\ell))$, and the output as $y_t = h(x_t, \zeta_t, \phi_\ell, v_t)$, $s_\ell \le t \le s_{\ell+1} - 1$, with $h(x_t, \zeta_t, \phi_\ell, v_t) =$

$\underline{h}(x_t, \theta(\zeta_t, \phi_\ell), v_t)$. Costly actuation is captured by defining a cost that penalizes the average rate of control actions $J_c := \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\sigma_t]$. Then, minimizing

$$J_{av} = J_s + \delta J_c \tag{5}$$

over different values of $\delta$ is equivalent to finding Pareto optimal policies. Another interpretation for $J_{av}$ is that the actuation is actually costly, and noticing that the running cost for $J_{av}$ is $\underline{g}(x_t, u_t) + \delta\sigma_t$. The goal is to find a policy

$$\sigma_t = \mu_t(\mathcal{I}_t), \quad t \in \mathbb{N}_0, \tag{6}$$

that minimizes $J_{av}$. Two applications are discussed next.

*1) Remote estimation with costly transmission:* Consider a process described by a special case of (1)

$$x_{t+1} = f(x_t, w_t), \quad y_t = x_t, \tag{7}$$

which can equivalently be described by a transition matrix $P$ with entries $P_{ij} = \text{Prob}[x_{t+1} = i | x_t = j]$, assumed to be ergodic. Model (7) results from quantizing a process $\xi_{t+1} = \underline{a}(\xi_t, \omega_t)$ with $\xi_t \in \mathbb{R}^{\underline{n}}$ and $\omega_t$ i.i.d. disturbances, and state $x_t$ labels one of $n$ representative values $\underline{\xi}^i$ of the quantized state variable $\xi_t$. Thus, there exists a labeling map $\pi$ such that $\underline{\xi}^i = \pi(i)$. The state is known to an agent who wishes to send it to a remote estimator. Control times $s_\ell$ are understood here in a broad sense as the times at which information is sent to the remote estimator; $\sigma_t \in \{0, 1\}$ determines when information is sent ($\sigma_t = 1$) or not ($\sigma_t = 0$). Transmissions are assumed to be expensive, e.g., due to battery limitations on the sensor side. On the remote side, an estimator obtains $q_t = \begin{bmatrix} q_{t,1} & \cdots & q_{t,n} \end{bmatrix}^\intercal \in \mathcal{P}_n$, $q_{t,i} = \text{Prob}[x_t = i | \mathcal{Z}_t]$, with $\mathcal{Z}_t := \{y_\ell | 0 \le \ell \le t, \sigma_\ell = 1\}$, and computes $\hat{\xi}_t := \mathbb{E}[\pi(x_t)|\mathcal{Z}_t] = \sum_{i=1}^n \pi(i) q_{t,i}$ Note that, assuming $\sigma_0 = 1$, $q_t = \delta_{x_t}$, if $\sigma_t = 1$, $q_t = P^{\zeta_t} \delta_{x_{t-\zeta_t}}$, if $\sigma_t = 0$, where $\delta_i$ is the column vector of zeros except at position $i$, where it equals 1. The running cost of an average cost penalizes the difference between the state and the estimated state $\|\pi(x_t) - \hat{\xi}_t\|^2$. While (7) does not depend on the control input $u_t$ we can use $u_t$ as a modeling variable to ensure we can write (2) with this running cost. In fact, we can define

$$u_{1,t} = U_{\zeta_t | s_\ell} = x_{s_\ell}, \quad u_{2,t} = \zeta_t, \text{ for } s_\ell \le t \le s_{\ell+1} - 1,$$

and $u_t \in \{1, \dots, n_u\}$, with $n_u = n\bar{h}$ to assign a unique label to the pair $(u_{1,t}, u_{2,t}) \in \{1, \dots, n\} \times \{0, \dots, \bar{h} - 1\}$. Then $\|\pi(x_t) - \hat{\xi}_t\|^2$ can be written as $\underline{g}(x_t, u_t)$ as $\hat{\xi}_t$ are functions of the state and control input since, when $\sigma_t = 0$, $q_t = P^{u_{2,t}} \delta_{u_{1,t}}$ and when $\sigma_t = 1$, $q_t = \delta_{x_t}$. Moreover, $p_{s_\ell | s_\ell} = \delta_{x_{s_\ell}} \in \{\delta_1, \dots, \delta_n\}$ belongs to a finite set, $U_{j|s_\ell}$ are functions of $\phi_\ell = x_{s_\ell}$ and $j$ and ergodicity of $Q$ follows from ergodicity of $P$. The goal is to find a policy (6) for when to apply control actions (send remote data) in order to minimize (5) with the proposed running cost in $J_s$.

*2) Costly interventions based on limited data:* Consider a set of $N$ discrete interdependent states $x_i \in \{1, \dots, n_i\}$ evolving in a free, or non-controlled, fashion according to

$$x_{i,t+1} = f_i(x_{1,t}, \dots, x_{N,t}, w_{i,t}), \quad i \in \{1, \dots, N\}, \tag{8}$$

when there are no control interventions ($\sigma_t = 0$), where the disturbance inputs $w_{i,t}$ live in a finite set. At control or intervention times ($\sigma_t = 1$) the state is reset to

$$x_{1,t+1} = \alpha_1, \ \dots, \ x_{N,t+1} = \alpha_N \tag{9}$$

and cost $\delta$ is paid per intervention. Only a subset of states is measured by $M$ sensors. Each sensor $j \in \{1, \dots, M\}$, depends on a subset of $n_j$ states $\mathcal{L}_j = \{l_1, \dots, l_{n_j}\}$ with $l_j \in \{1, \dots, N\}$, $y_j = h_j(x_{l_1}, \dots, x_{l_{n_j}}, v_j)$, and where the measurements $y_j \in \{1, \dots, n_{y,j}\}$ might be corrupted by noise $v_j$, also living in a finite set. The running cost of an average cost is

$$g_A(x_{1,t}, \dots, x_{N,t}) = \sum_{i=1}^N g_i(x_{i,t}) \tag{10}$$

In precision agriculture (see Section VI), each $x_{i,t}$ is a binary variable representing if there is weed ($x_{i,t} = 2$) in a given subarea of a field at time $t$ or not ($x_{i,t} = 1$); $x_{i,t}$ depends on neighboring states and only a subset of subareas can be measured. At intervention times $s_\ell$ the weed is completely removed, setting all the states to $x_{i,s_\ell+1} = 1$, $\forall i \in \{1, \dots, N\}$. For some constant $d$, representing the cost of having weed between $t$ and $t+1$, $g_i(2) = d$ and $g_i(1) = 0 \quad \forall i \in \{1, \dots, N\}$.

For the general setting, we can find single variables $x_t \in \{1, \dots, n\}$, $y_t \in \{1, \dots, n_y\}$, $n = n_1 \times \cdots \times n_N$, $n_y = n_{y,1} \times \cdots \times n_{y,M}$ to label all possible state and output combinations and write the problem in the canonical formula described above, see special case in Section VI. Here $p_{t+1|t} = \delta_\alpha$, when $\sigma_t = 1$, corresponding to $x_{t+1} = \alpha$, where $\alpha \in \{1, \dots, n\}$ is the label corresponding to (9), and Assumption 1 is trivially met since $Q$ corresponds to a Markov chain with just one state. The control input $u_t$ can be used to model the process evolution and in particular the state reset (9), but it can also be omitted. The goal is to find a policy for $\sigma_t$ as a function of the information set $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{y_t, \sigma_{t-1}\}$ for $t \in \mathbb{N}$, with $\mathcal{I}_0 = \{y_0\}$ to minimize (5) with the proposed running cost.

## III. Optimal policy

We start by providing a result that converts the average cost problem to the following stopping time problem. Consider (1) for $t \in \{0, 1, \dots, \bar{h}\}$. The information available to make a decision at time $t \in \{1, \dots, \bar{h}\}$ on either $s_1 = \tau_0 \in \{1, \dots, \bar{h}\}$ is equal to $t$ or larger, is summarized in $p_{0|0} = \rho_{\phi_0}$, if (3a) holds, or in $p_{1|0} = \rho_{\phi_0}$, if (3b) holds, and in the measurements $y_1$, $y_2$, …, $y_t$. Thus, we define the information set $\mathcal{H}_t^0 = \{\phi_0\} \cup \{y_\kappa | \kappa \in \{1, \dots, t\}\}$. Consider the optimal stopping time problem for $\ell = 0$ and $s_0 = 0$:

$$J_{\text{stop}} = \min_{\tau_\ell} \frac{1}{\mathbb{E}[\tau_\ell]} \left( \mathbb{E}[\sum_{t=0}^{\tau_\ell - 1} g(x_{s_\ell + t}, \zeta_{s_\ell + t}, \phi_{s_\ell})] + \delta \right) \tag{11}$$

where $\tau_0$ is a stopping time with respect to the filtration corresponding to the information set $\mathcal{H}_t^0$. In other words, the event $[\tau_0 = m]$ is a function of $\mathcal{H}_m^0$. Similarly, we can define stopping times $\tau_\ell$ with respect to the filtration corresponding

to the information set $\mathcal{H}_r^\ell = \{\phi_\ell\} \cup \{y_{s_\ell+1}, y_{s_\ell+2}, \ldots, y_{s_\ell+r}\}$ and consider an identical stopping time problem to (11) for a general $\ell$. These problems are identical due to Assumptions 1, 2, 3 as stated next. These stopping times $\tau_\ell$ define the control times according to $s_{\ell+1} = s_\ell + \tau_\ell$.

*Lemma 1:* Suppose that Assumptions 1, 2, 3 hold. Then the optimal stopping time policies for problems (11) for $\ell \in \mathbb{N}_0$ are identical in the sense that they take the form

$$\begin{aligned} \tau_\ell &= \min\{r \in \{1, \ldots, \bar{h}\} | \sigma_{s_\ell+r} = 1\} \\ \sigma_{s_\ell+r} &= \xi_r(H_r^\ell), \text{ for every } \ell \in \mathbb{N}_0 \end{aligned} \tag{12}$$

for the same functions $\xi_r$, $r \in \{1, \ldots, \bar{h}\}$. Moreover, $\sigma_t = \sigma_{s_{\bar{\ell}(t)}+\zeta_t}$ with $\sigma_{s_\ell+r}$ given by (12) is also an optimal policy for the average cost problem of minimizing (5) and $J_{\text{av}} = J_{\text{stop}}$. Furthermore, the optimal policy for problem (11) when $\ell = 0$ can be obtained by solving the stopping time problem

$$\min_{\tau_0} \mathbb{E}[\sum_{t=0}^{\tau_0 - 1} (g(x_t, \zeta_t, \phi_0) - \beta)] + \delta \tag{13}$$

where $\beta \in \mathbb{R}_{\geq 0}$ is the largest value for which the optimal solution to (13) results in a zero cost and is given by $\beta = J_{\text{av}}$. □

The optimal policy and cost for problem (13) for a given $\beta$ can be obtained by the dynamic programming algorithm. Let $q_t = [q_{t,1} \ldots q_{t,n}]^\intercal$ with $q_{t,i} = \text{Prob}[x_t = i | \mathcal{H}_t^0]$ and $\bar{g}_t = [g(1, t, \phi_0) \ldots g(n, t, \phi_0)]^\intercal$, $t \in \{0, \ldots, \bar{h} - 1\}$, where the dependence of $\bar{g}_t$ on the given $\phi_0$ is omitted. Then, one should iterate for $t \in \{\bar{h} - 1, \ldots, 0\}$

$$\begin{aligned} J_{\bar{h}}(q_{\bar{h}}) &= \delta, \\ J_t(q_t) &= \min\{\delta, q_t^\intercal \bar{g}_t - \beta + \mathbb{E}[J_{t+1}(q_{t+1}) | \mathcal{H}_t^0]\} \end{aligned}$$

and the optimal policy is

$$\begin{aligned} \tau_0 &= \min\{t \in \{1, \ldots, \bar{h}\} | \sigma_t = 1\} \\ \sigma_t &= \begin{cases} 1 \text{ if } J_t(q_t) = \delta \text{ or if } t = \bar{h}, \\ 0 \text{ otherwise.} \end{cases} \end{aligned}$$

Note that $q_t$ can be iterated with the Bayes' filter

$$q_{t+1} = \frac{1}{\alpha(y_{t+1})} D(y_{t+1}) P q_t$$

with $P_{ij} = \text{Prob}[x_{t+1} = i | x_t = j]$ and $D(y_{t+1}) = \text{diag}([R_{y_{t+1},1} \ldots R_{y_{t+1},n}])$ with $R_{\ell,j} = \text{Prob}[y_t = \ell | x_t = j]$, for $\ell \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$ and $\alpha(\ell) = \text{Prob}[y_{t+1} = \ell | \mathcal{I}_t] = \sum_{j=1}^n R_{\ell,j} \bar{r}_{t,j}$. with $\bar{r}_t = [\bar{r}_{t,1} \ \ldots \ \bar{r}_{t,n}]^\intercal$, $\bar{r}_t = P q_t$. The filter is initialized with $q_0 = \rho_{\phi_0}$ if (3a) holds and $\bar{r}_0 = \rho_{\phi_0}$ if (3b) holds. It is well known that (see, e.g., [14]) $J_t(q_t) = \min_{c \in \mathcal{J}_t} c^\intercal q_t$. One can obtain that, for $t \in \{\bar{h} - 1, \ldots, 0\}$,

$$\begin{aligned} \mathcal{J}_t = \{ & d_{1,t} + \sum_{\ell=1}^{n_y} P^\intercal D(\ell)^\intercal \bar{c}_{j_\ell} | \bar{c}_{j_\ell} \in \mathcal{J}_{t+1}, \\ & j_1, \ldots, j_{n_y} \in \{1, \ldots, |\mathcal{J}_{t+1}|\}\} \cup \{d_2\} \end{aligned} \tag{14}$$

with $d_{1,t} = \bar{g}_t - \beta 1_n$, $d_2 = \delta 1_n$, and $\mathcal{J}_{\bar{h}} = d_2$. However, the size of set $\mathcal{J}_t$, denoted by $|\mathcal{J}_t|$, grows as $|\mathcal{J}_t| = 1 + |\mathcal{J}_{t+1}|^{n_y}$. Thus, this is a computationally intractable method.

When $\bar{g}_t$ depends on $\phi_0 = i$ so will the cost-to-go and now we stress this by denoting the cost-to-go at time $t = 0$ by $J_0^{i,\beta}(q_0)$ where the dependence on $\beta$ is also added. Cost (13) is then equal to $\sum_{i=1}^b a_i J_0^{i,\beta}(\rho_i)$ when (3a) holds. Thus, one needs to find $\beta$ for which this cost is zero. To this end we can simply run a bisection algorithm as it can be shown that the cost is a strictly monotone and continuous function of $\beta$.

## IV. Relaxed dynamic programming policy

The idea of relaxed dynamic programming [14] is to find simpler functions to approximate $J_t(q_t)$. Here we consider the following functions $V_{\bar{h}}(q_{\bar{h}}) = \delta$ and

$$V_t(q_t) = \min_{c \in \mathcal{V}_t} c^\intercal q_t, \quad t \in \{0, 1, \ldots, \bar{h} - 1\}, \tag{15}$$

where $\mathcal{V}_t \subseteq \mathcal{J}_t$ can be seen as a pruned version of $\mathcal{J}_t$. We will provide a procedure to obtain this pruned set in such a way that

$$J_t(q_t) \leq V_t(q_t) \leq J_t(q_t) + \epsilon(\bar{h} - t), \ t \in \{0, 1, \ldots, \bar{h} - 1\}, \tag{16}$$

so that $V_t(q_t)$ is always within an additive factor $\epsilon(\bar{h} - t)$ of the optimal policy and the resulting policy as well. Although the original idea of relaxed dynamic programming considered a multiplicative factor for the guarantees, i.e., $J_t(q_t) \leq V_t(q_t) \leq (1 + \epsilon) J_t(q_t)$, here an additive factor is chosen for two reasons. First, here $J_t$ and $V_t$ take in general negative values (due to subtracting $\beta$ from the running cost (13)), thus the multiplicative bound makes no sense. Second, here $J_t(q_t) \leq \max_{x,\phi} g(x, 0, \phi) + \delta$ for every $t$ and $q_t$ so that scaling issues can be avoided.

Towards this, let us define

$$J_t^\epsilon(q_t) = \min\{\delta + \epsilon, (q_t^\intercal \bar{g} - \beta + \epsilon) + \mathbb{E}[V_{t+1}(q_{t+1}) | \mathcal{H}_t^0]\}$$

Using similar steps to the ones that lead to (14) (see, e.g., [14]) one can conclude that $J_t^\epsilon(q_t) = \min_{c \in \mathcal{J}_t^\epsilon} c^\intercal q_t$,

$$\begin{aligned} \mathcal{J}_t^\epsilon = \{ & d_{1,t} + \epsilon 1 + \sum_{\ell=1}^{n_y} P^\intercal D(\ell)^\intercal \bar{c}_{j_\ell} | \bar{c}_{j_\ell} \in \mathcal{V}_{t+1}, \\ & j_1, \ldots, j_{n_y} \in \{1, \ldots, |\mathcal{V}_{t+1}|\}\} \cup \{d_2 + \epsilon 1\} \end{aligned}$$

Function $J_t^\epsilon$ coincides with $J_t$ when $\epsilon = 0$. However, this function is defined with an extra cost term for the running cost when $\epsilon > 0$. Given a $\tilde{c}^\epsilon \in \mathcal{J}_t^\epsilon$, consider a corresponding $\tilde{c}^0$ from the set $\mathcal{J}_t = \mathcal{J}_t^\epsilon|_{\epsilon=0}$ defined in (14).

At each time $t$, the set $\mathcal{V}_t$ is a pruned version of the set $\mathcal{J}_t$. To facilitate the pruning operation, which is carried out iteratively, it is wise to define a heuristic function $H(c)$ which assigns a score to the elements $c$ of a given set (say $J_t^\epsilon$). The higher $H(c)$ the larger the belief that $c$ can be pruned.

*Relaxed Dynamic Programming procedure:*

1) Initialize $\mathcal{V}_t$ as empty.
2) Take the element $\tilde{c}^\epsilon$ in $\mathcal{J}_t^\epsilon \setminus \mathcal{V}_t$ with the smallest $H$ and check if it satisfies

$$\min_{c \in \mathcal{V}_t} c^\intercal q \leq \tilde{c}^{\epsilon\intercal} q \quad \forall q \in \mathcal{P}_n. \tag{17}$$

3) If (17) is not satisfied (or if $\mathcal{V}_t$ is empty), then add $\tilde{c}^0$ (corresponding to $\tilde{c}^\epsilon$) to $\mathcal{V}_t$. If there are no more elements in $\mathcal{J}_t^\epsilon$, then stop, otherwise go to step 2.

The heuristic used evaluates $c^\intercal \xi_i$ for each of $n_c$ members $c$, and for $s$ fixed values $\xi_i$, $i \in \{1,\dots,s\}$ and assigns a reward $r_i \in \{1,\dots,n_c\}$ to each member corresponding to the ranking in the $i$th ordered set according to $c^\intercal \xi_i$. Then $H(c) = \sum_{i=1}^s r_i$.

*Lemma 2:* Let $V_t$ be defined by (15) with the set $\mathcal{V}_t$ obtained from the relaxed dynamic programming procedure. Then (16) holds.

□

The following test provides a sufficient condition to test if (17) holds and if (17) is replaced by this test the same guarantees can also be given: if there exists $\alpha_i \geq 0$ with $\sum_{i=1}^{|\mathcal{V}_t|} \alpha_i = 1$ such that, with $c_i \in \mathcal{V}_t$, $\sum_{i=1}^{|\mathcal{V}_t|} \alpha_i c_i \leq \tilde{c}$ then (17) holds. To test this latter condition we can simply test the feasibility of the simplex $A\alpha \leq b$ with $A = [C^\intercal \; -I \; 1 \; -1]^\intercal$, $b = [\tilde{c}^\intercal \; 0 \; 1 \; -1]^\intercal$, $C = [c_1 \; \dots \; c_{|\mathcal{V}_t|}]$.

## V. Consistent policy

We start by providing a result stating the performance of periodic control.

*Lemma 3:* Suppose that Assumptions 1, 2, 3 hold and that $\tau_\ell = h$, $\forall \ell \in \mathbb{N}_0$, are constant, corresponding to periodic control. Then $J_{\text{av}} := J_s + \delta J_p$ with

$$J_s = \eta_h := \sum_{i=1}^{b} \nu_{h,i} a_i, \quad J_p = \frac{1}{h} \tag{18}$$

where

$$\nu_{h,i} := \begin{cases} \frac{1}{h}(\sum_{\ell=0}^{h-1} \bar{g}_\ell^\intercal P^\ell)\rho_i, \text{ if (3a) holds} \\ \frac{1}{h}(g_0^\intercal P^{\ell-1} + \sum_{\ell=0}^{h-2} \bar{g}_{\ell+1}^\intercal P^\ell)\rho_i, \text{ if (3b) holds.} \end{cases}$$

□

The proposed policy yields a better trade-off between average number of actions and average cost in the following sense. Suppose that we define the curve $(s, J_{\text{per}}(s))$ with

$$J_{\text{per}}(s) = \eta_r + (\eta_{r+1} - \eta_r)(s - r) \text{ if } s \in [r, r+1).$$

Then $(\bar{\tau}_\pi, J_\pi)$, with $\bar{\tau}_\pi = \mathbb{E}[\tau_0] = 1/J_c$ the average inter-control time and $J_\pi = J_{\text{av}}$ the policy's cost, is below this curve. We call this a *consistent* policy. We propose such a consistent policy inspired by [9], [10] but different both in form and derivation. It is defined as follows:

$$\tau_\ell = \min\{r \in \{1,\dots,\bar{h}\} | \sum_{t=s_\ell}^{s_\ell+r} \bar{g}_t^\intercal p_{t|t} > -\delta + \omega_{c,\phi_\ell} r\} \tag{19}$$

for $\omega_{c,i} < \omega_{m,i}$ where $\omega_{m,i} = \min\{\nu_{r,i} + \delta\frac{1}{r} | r \in \{1,\dots,\bar{h}\}\}$. Intuitively, assuming for simplicity $b = 1$ and $\ell = 0$, if we knew the state and if we could make sure that $\sum_{t=0}^{\tau_0-1} g(x_t,t,1) + \delta - \omega_{c,1}\tau_0 \leq 0$ we would obtain also that $\mathbb{E}[\sum_{t=0}^{\tau_0-1} g(x_t,t,1) + \delta - \omega_{c,1}\tau_0] \leq 0$, which would imply that $\frac{1}{\mathbb{E}[\tau_0]}\mathbb{E}[\sum_{t=0}^{\tau_0-1} g(x_t,t,1)] + \delta\frac{1}{\mathbb{E}[\tau_0]} \leq \omega_{c,1} < \omega_{m,1}$ meaning that we could ensure that such a policy would yield a better combined cost than that of the periodic policy for any $h$. Although the state is not available, this policy still ensures this property by replacing the term $g(x_t,t,\phi_0)$ by its expected value given the information up to time $t$, $\bar{g}_t^\intercal q_t$, and by taking into account the initial condition.

A limitation is that if $\delta$ is large and $w_{m,i}$ is small the policy might lead to many control actions. However, the consistency property holds for any choice of $\delta \geq 0$ and $\omega_{c,i} < \omega_{m,i}$. The only restriction then comes from $\omega_{m,i}$.

*Theorem 1:* Suppose that Assumptions 1, 2, 3 hold. Let $J_\pi$ be the cost $J_{\text{av}}$ of the proposed policy (19) and let $\bar{\tau}_\pi = \mathbb{E}[\tau_0] = 1/J_c$ be the average inter-control time, for given $\delta \geq 0$ and $\omega_{c,i} < \omega_{m,i}$, $i \in \{1,\dots,b\}$. Then,

$$J_\pi \leq J_{\text{per}}(\bar{\tau}_\pi). \tag{20}$$

## VI. Application in precision agriculture

Consider the precision farming setting of Section II.2) and assume that the field is a strip of potato crops divided into $N$ subareas. The state $x_t \in \{1,\dots,2^N\}$ labels all possible states of the $N$ weed indicator variables $x_{t,i} \in \{1,2\}$. At initialization and directly after the interventions at times $s_\ell$, the whole field has no weed. Weed can simply appear in a given subarea or spread from one of the neighboring subareas, which is summarized by, for $a, b \in \{1,2\}$,

$$\text{Prob}[x_{t+1,l} = 2 | x_{t,l-1} = a, x_{t,l} = 1, x_{t,l+1} = b] = r_{ab}$$

when $l \notin \{1, N\}$, with $0 < r_{11} < r_{21} = r_{12} < r_{22} < 1$ and $\text{Prob}[x_{t+1,1} = 2 | x_{t,2} = i, x_{t,1} = 1] = \text{Prob}[x_{t+1,N} = 2 | x_{t,N-1} = i, x_{t,N} = 1] = r_i$, $i \in \{1,2\}$, with $r_1 = r_{11}$, $r_2 = r_{21}$. When a subarea has weed, it remains until an intervention, $\text{Prob}[x_{t+1,i} = 2 | x_{t,i} = 2] = 1$ for every $i$. From these assumptions (8) can be computed or equivalently the transition probability matrix $P_{ij} = \text{Prob}[x_{t+1} = i | x_t = j]$ can be computed, as follows. First, let $\pi_j(i) = x_{t,j}$ extract the value of the indicator variable $x_{t,j}$ $j \in \{1,\dots,N\}$ for a given state $x_t = i \in \{1,\dots,2^N\}$ and let $\mathcal{N}(i,t) = (x_{t,i-1}, x_{t,i+1})$, $\mathcal{N}(1,t) = x_{t,2}$, $\mathcal{N}(N,t) = x_{t,N-1}$ be the neighboring indicator states. Then $P_{ij} = 0$ if there exists $\ell$ such that $\pi_\ell(i) = 1$ and $\pi_\ell(j) = 2$ and, otherwise, $P_{ij} = \left(\prod_{\ell \in \mathcal{M}_{i,j,2}} r_{\mathcal{N}(\ell,t)}\right)\left(\prod_{\ell \in \mathcal{M}_{i,j,1}} (1 - r_{\mathcal{N}(\ell,t)})\right)$ with $\mathcal{M}_{i,j,\kappa} = \{\ell \in \{1,\dots,N\} | \pi_\ell(j) = 1, \pi_\ell(i) = \kappa\}$, $\kappa \in \{1,2\}$. While each sensor $i \in \{1,\dots,M\}$ can, in general, measure several subareas, here only one subarea per sensor is assumed $\ell_i \in \{1,\dots,N\}$. Thus $y_{t,i} \in \{1,2\}$ and an error probability is assumed $\text{Prob}[y_{t,i} = 2 | x_{t,\ell_i} = 1] = \text{Prob}[y_{t,i} = 1 | x_{t,\ell_i} = 2] = e_p$. From these assumptions, the output map can be computed or equivalently $R_{ij} = \text{Prob}[y_t = i | x_t = j]$. Let $\chi_\kappa(i) = y_{t,\kappa}$ extract the value of the subarea measurement $y_{t,\kappa}$, $\kappa \in \{1,\dots,M\}$ for a given measurement $y_t = i \in \{1,\dots,2^M\}$ and let $s_{ij}$ be the number of subarea measurements associated with $y_t = i$ that provide a correct estimate for the corresponding subarea state associated with $x_t = j$, i.e., $s_{ij} = |\{\kappa \in \{1,\dots,M\} | \chi_\kappa(i) =$

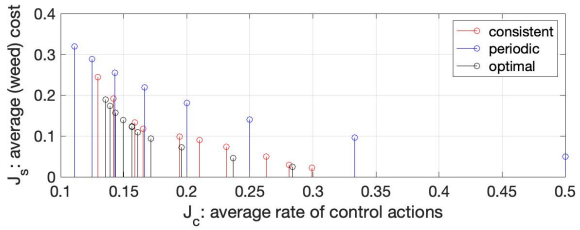


Fig. 1: Average (weed) cost $J_s$ versus average rate of control actions $J_c$ for three policies: periodic, optimal and consistent for a simple example with $N = M = 1$

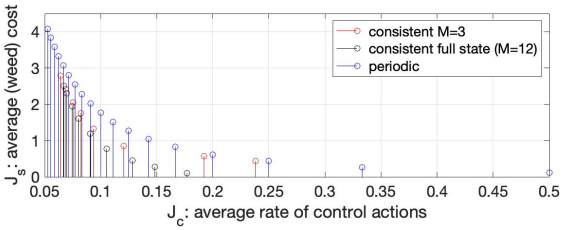


Fig. 2: Average (weed) cost $J_s$ versus average rate of control actions $J_c$ when $N = 12$ for periodic control, consistent with $M = 3$ sensors and full state feedback $M = 12$

$\pi_{\ell_\kappa}(j)\}|$. Then $R_{ij} = (1-e_p)^{s_{ij}} e_p^{M-s_{ij}}$. Running cost (10) can be written as the running cost in (11) as

$$g(i,t,0) = d\left[\pi_1(i) - 1 \quad \dots \quad \pi_N(i) - 1\right] 1_N. \qquad (21)$$

We start by considering a very special case with just one field $N = 1$, $x_t \in \{1, 2\}$, one measurement $y_t \in \{1, 2\}$ and

$$P = \begin{bmatrix} 1 - r_{11} & 0 \\ r_{11} & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 - e_p & e_p \\ e_p & 1 - e_p \end{bmatrix},$$

This simple case allows one to still compute the optimal policy cost and understand how close is the cost of the consistent policy, for this example. The parameters considered are $\bar{h} = 10$, $d = 1$, $r_{11} = 0.1$, $e_p = 0.2$. Figure 1 shows the results of periodic control, optimal policy (or relaxed dynamic programming with $\epsilon = 0$) considering $\delta \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5\}$ and of the consistent policy with $\delta = 0$ and $\omega_{m,1} \in \{0.01 + 0.02i | i \in \{0, 1, \dots, 8\}\} \cup \{0.2, 0.3, 0.4\}$. As it can be seen, the optimal policy yields a significant reduction of cost $J_s$, when the running cost is (21), for the same average intervention interval with respect to periodic control. The consistent policy provides results close to optimal.

We now consider a more realistic example with $N = 12$ subareas, $M = 3$ sensors $l_1 = 3$, $l_2 = 6$, $l_3 = 9$ and parameters $d = 1$, $r_{11} = 0.02$, $r_{21} = 0.2$, $r_{22} = 0.4$, $e_p = 0.02$, $\bar{h} = 20$. The number of states is $n = 2^{12} = 4096$. We can no longer apply relaxed dynamic programming since solving the corresponding linear programs with $c \in \mathbb{R}^{4096}$ is computationally hard. However, we can still compute the consistent policy. The results are depicted in Figure 2 considering parameters $\omega_{m,1} \in \{0.2, 0.4, 0.6, 1, 1.5, 2, 2.5, 3, 3.5\}$ and $\delta = 0$. Note that in fact the policy provides a significant reduction of cost for the same average rate of control actions with respect to periodic control. To access the value of information, the case $M = 12$ and $e_p = 0$ is also shown, leading to a reduction of cost.

## VII. Conclusion

We have proposed a new framework to determine when a partially observable Markov decision process with finite state, input, and measurement spaces should be sampled or/and intervened, assuming these operations are costly. We proposed two approaches to find a policy for the time intervals between interventions. The approach based on relaxed dynamic programming can provide nearly optimal cost when the state dimension is very small. Still, it is not suitable for larger dimensional problems, as this involves solving large linear programs. In turn, the approach based on consistent control provides a simple and effective solution.

## References


[1] G. Vellidis, M. Tucker, C. Perry, C. Kvien, and C. Bednarz, "A real-time wireless smart sensor array for scheduling irrigation," *Comput. Electron. Agric.*, vol. 61, no. 1, p. 44–50, apr 2008.

[2] J. A. Cabrera, J. R. Pedrasa, A. M. Radanielson, and A. Aswani, "Mechanistic crop growth model predictive control for precision irrigation in rice," in *2021 European Control Conference (ECC)*, 2021, pp. 1243–1248.

[3] J. Shanahan, N. Kitchen, W. Raun, and J. Schepers, "Responsive in-season nitrogen management for cereals," *Computers and Electronics in Agriculture*, vol. 61, no. 1, pp. 51–62, 2008, emerging Technologies For Real-time and Integrated Agriculture Decisions.

[4] D.-X. Wang and X.-R. Cao, "Event-based optimization for POMDPs and its application in portfolio management," in *18th IFAC World Congress Milano (Italy)*, 2011, pp. 3228–3233.

[5] L. Xia, Q. Jia, and X. Cao, "A tutorial on event-based optimization—a new optimization framework," *Discrete Event Dyn. Syst.*, vol. 24, pp. 103–132, 2014.

[6] J. Messias, M. Spaan, and P. Lima, "Multiagent POMDPs with asynchronous execution," in *Proceedings of the 12th International Confer- ence on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, 2013, pp. 1273–1274.

[7] A. Molin and S. Hirche, "Structural characterization of optimal event-based controllers for linear stochastic systems," in *Proceedings of the IEEE International Conference on Decision and Control (CDC 2010)*, 2010, pp. 3227–3233.

[8] Y. Xu and J. Hespanha, "Optimal communication logics in networked control systems," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 4, Dec 2004, pp. 3527–3532 Vol.4.

[9] D. J. Antunes and M. H. Balaghi I., "Consistent event-triggered control for discrete-time linear systems with partial state information," *IEEE Control Systems Letters*, vol. 4, no. 1, pp. 181–186, Jan 2020.

[10] D. J. Antunes and B. A. Khashooei, "Consistent dynamic event-triggered policies for linear quadratic control," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1386–1398, 2018.

[11] V. Krishnamurthy and D. V. Djonin, "Structured threshold policies for dynamic sensor scheduling—a partially observed Markov decision process approach," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4938–4957, 2007.

[12] V. Krishnamurthy, "How to schedule measurements of a noisy Markov chain in decision making?" *IEEE Trans. Inf. Theor.*, vol. 59, no. 7, p. 4440–4461, jul 2013.

[13] M. Rezaeian, "Sensor scheduling for optimal observability using estimation entropy," in *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07)*, 2007, pp. 307–312.

[14] B. Lincoln and A. Rantzer, "Relaxing dynamic programming," *IEEE Trans. on Automatic Control*, vol. 51, no. 8, pp. 1249–1260, Aug 2006.

[15] A. O. Hero, D. A. Castanon, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*, 2008.