Optimal stopping for updating controls

Maben Rabi[1] and Karl H. Johansson[1]

Automatic Control lab
Royal Institute of Technology (KTH)
Osquldas väg 10, plan 6
100 44 Stockholm, Sweden
firstname.lastname@ee.kth.se

1 Control under limited communications

We study a problem of minimum variance control of a diffusion process with a restriction on the class of admissible control policies. This restriction is motivated by networked control systems where, the sensor, controller and actuator are remotely located to each other and hence connected via links with severe limits on communication rates. In practical situations such as in wireless industrial control [Antsaklis & Baillieul(2007)], frequently, it is the link from sensors to controllers that have restricted rates.

Assume that the sensor obtains perfect and continuous measurements of the state process. Since the sensor is located remotely from the controller, continuous transmission of the sensor observations to the controller is impossible; only a sampled version of the observations can be forwarded to the controller. The control process is hence generated with some compressed information about the state process. It is generated based on the timed sequence of samples received, and, is forced to be piecewise deterministic. The traditional solution is to sample sensor observations periodically choosing a sampling rate equalling the average communication rate permitted by the channel. Here, we pursue the more efficient strategy of letting the sensor choose sampling times to be stopping times w.r.t. its observations. The rate of arrival of the stopping times then should be no greater than the rate permitted by the channel.

This pursuit boils down to designing two elements namely, an *Event detector* at the sensor end which applies a set of stopping rules to generate samples, and, a *Control generator* inside the controller node which produces piece-wise deterministic waveforms derived from the sequence of received samples. In mathematical terms, we get a sequential decision problem for a two person team where, the control signal, and, sampling times are decision variables. Moreover, this decision problem has a non-standard information pattern with the control values and the stopping times being chosen based on different information.

We study the finite horizon version of the problem. The communication constraint is a hard limit on the number of times inside the finite horizon when the control waveform is modified based upon feedback from the sensor. On the horizon $[0, T]$, consider the controlled scalar Diffusion process x_t which obeys the SDE:

$$dx_t = \mu(x_t)\, dt + \sigma(x_t)\, dB_t + \kappa(x_t)\, u_t dt, \tag{1}$$

where, B_t is a standard scalar Brownian motion process, the functions μ, σ, and, κ satisfy the usual Lipschitz and regularity conditions ensuring existence and uniqueness of the state process x_t. The control signal is as usual required to be measurable w.r.t. the filtration $\mathcal{F}_t^x$. Because the control is actually generated based on compressed information regarding the state, it has a further restriction in the manner of its generation. This is explained subsequently.

Based on its observations, the sensor chooses up to N sampling times $\{\tau_1, \tau_2, \ldots \tau_N\}$. The controller observes its own output continuously and the discrete sequence of sampling times and sample values passed on from the sensor. Based on its observations, the controller updates the control waveform at the sampling times generated by the sensor. Thus an admissible control process must be measurable w.r.t. the filtration generated by the output of the sensor.

The goal of the control system designer is to keep the state as close to the origin as possible while using no more than N samples. We adopt the standard squared error as the control performance measure.

Hence, given the initial value x_0, the objective is to pick the sequence of N sampling times and the control values so that the following aggregate control error is minimized:

$$J_N\left(x_0, \mathcal{U}, \{\tau\}_{i=1}^N\right) = \mathbb{E}\left[\int_0^T x_s^2 ds \middle| x_0\right]. \tag{2}$$

1.1 Relation to previous work

Kushner [Kushner(1964)] has considered a variation of the finite horizon LQG control problem where the measurement sampling instants are also decision variables which are to be chosen offline.

Åström and Bernhardsson [Åström & Bernhardsson(2002)] treat a problem of minimum variance stabilization of a scalar linear plant using state-resetting impulse controls. By explicit calculations, they show that for the same average rate of impulse invocation, the level-triggered scheme provides lower average stabilization error. This article provides a similar comparison when the control is not of impulse type, but is piecewise continuous. In fact, we will later restrict the control signals to be is piecewise constant, these being natural candidates for situations in networked control. The work reported in this paper extend earlier work [Rabi *et al.*(2008)Rabi, Johansson, & Johansson] by the authors.

We focus on the single sampling problem where the sample budget N is exactly one. The multiple sampling problem is not much harder and it can be solved by recursively solving N problems of the single sampling type.

2 Optimal choice of stopping times and control updates

For the single stopping problem, the control waveform has the form:

$$u_t = \begin{cases} U_0 \;=\; U_0\left(x_0, T\right) & \text{if } \; 0 \le t < \tau, \\ U_1 \;=\; U_1\left(x_\tau, \tau, T\right) & \text{if } \; \tau \le t \le T, \end{cases}$$

where, the switch time τ is a stopping time w.r.t. the x-process, one which is restricted to fall inside the interval $[0, T]$. Thus the set of decision variables is the triple: (U_0, U_1, τ). Because of the non-classical information pattern, we cannot use the existing results on combined stopping and control. We can however describe some properties of the optimal choice of policies.

We will first decompose the aggregate quadratic cost. Let $\Phi_U(t_2, t_1, x)$ denote the solution to the SDE 1 at time t_2 withe initial condition at time $t_1 \le t_2$ being x, and, with the constant control input U over the interval $[t_1, t_2)$. Then,

$$\begin{aligned} J_N\left(x_0, \{U_0, U_1\}, \tau\right) &= \mathbb{E}\left[\int_0^\tau x_s^2 ds + \int_\tau^T x_s^2 ds\right] \;=\; \mathbb{E}\left[\int_0^\tau \Phi_{U_0}^2(s, 0, x_0)\, ds + \int_\tau^T \Phi_{U_1}^2(s, \tau, x_\tau)\, ds\right], \\ &= \mathbb{E}\left[\int_0^T \Phi_{U_0}^2(s, 0, x_0)\, ds - \int_\tau^T \Phi_{U_0}^2(s, 0, x_0)\, ds + \int_\tau^T \Phi_{U_1}^2(s, \tau, x_\tau)\, ds\right], \\ &= \int_0^T \mathbb{E}\left[\Phi_{U_0}^2(s, 0, x_0)\right] ds - \mathbb{E}\left[\int_\tau^T \left\{\Phi_{U_0}^2(s, 0, x_0) - \Phi_{U_1}^2(s, \tau, x_\tau)\right\} ds\right], \\ &= \int_0^T p_{U_0}(s, x_0)\, ds - \mathbb{E}\left[\int_\tau^T \left\{\Phi_{U_0}^2(s, 0, x_0) - \Phi_{U_1}^2(s, \tau, x_\tau)\right\} ds\right], \end{aligned}$$

where, $p_{U_0}(s, x_0)$ is the second moment of the state at time t, under the constant control value U_0. Denote the first intergral by $\alpha\left(x_0, U_0, T\right)$. Now, using iterated expectations, we get:

$$\begin{aligned} J_N\left(x_0, \{U_0, U_1\}, \tau\right) &= \alpha\left(x_0, U_0, T\right) - \mathbb{E}\left[\mathbb{E}\left[\int_\tau^T \left\{\Phi_{U_0}^2(s, 0, x_0) - \Phi_{U_1}^2(s, \tau, x_\tau)\right\} ds \middle| \tau, x_\tau\right]\right], \\ &= \alpha\left(x_0, U_0, T\right) - \mathbb{E}\left[\int_\tau^T \mathbb{E}\left[\Phi_{U_0}^2(s, 0, x_0) - \Phi_{U_1}^2(s, \tau, x_\tau) \middle| \tau, x_\tau\right] ds\right], \\ &= \alpha\left(x_0, U_0, T\right) - \mathbb{E}\left[\int_\tau^T \mathbb{E}\left[\Phi_{U_0}^2(s, \tau, x_\tau) - \Phi_{U_1}^2(s, \tau, x_\tau)\right] ds\right], \end{aligned}$$

where, for a given, U_0 and τ, the integral $\int_\tau^T \mathbb{E}\left[\Phi^2_{U_0}(s,\tau,x_\tau) - \Phi^2_{U_1}(s,\tau,x_\tau)\right] ds$ can be minimized by a straightforward choice of U_1 via a standard optimal control problem for the integral cost. Let $U_1^*(x_\tau,\tau,T)$ denote an optimal choice obtained this way. Then the aggregate control distortion can be described as:

$$J_N(x_0,\{U_0,U_1\},\tau) \stackrel{\Delta}{=} \alpha(x_0,T) - \mathbb{E}\left[\beta(x_0,U_0,\tau,T)\right], \tag{3}$$

where the function: $\beta(x_0,U_0,\tau,T) = \int_\tau^T \mathbb{E}\left[\Phi^2_{U_0}(s,\tau,x_\tau) - \Phi^2_{U_1^*(x_\tau,\tau,T)}(s,\tau,x_\tau)\right]$. Notice that we can first minimize $\mathbb{E}[\beta]$ over all possible τ and use it to find the optimal choice of U_0. Minimizing the expected value of the function β is an optimal stopping problem with a reward collected at the stopping time. Although the reward is time-varying, the problem in is standard form and can be converted into a time-homogeneous problem by adding time as a state variable. Its solution is the first time the state hits time-varying and double sided barriers, or, if the current time exceeds the length of the horizon [Øksendal(2003)].

2.1 Person-by-person optimality

A standard result from Team theory states that the optimal set of action policies should satisfy an equilibrium condition. Essentially, any optimal set of policies $\mathcal{SP}$ is such that, each person/agent employs a policy which is the best possible as long as the others employ the fixed policies as per $\mathcal{SP}$. This provides us with the following necessary conditions for optimality:

$$\begin{cases} \tau^*(x_0) & = \operatorname{ess\,sup}_\tau \ \mathbb{E}\left[\beta(x_0,U_0^*(x_0),\tau,T)\right], \\ U_0^*(x_0) & = \inf_U \left\{\alpha(x_0,U,T) - \mathbb{E}\left[\beta(x_0,U,\tau^*(x_0),T)\right]\right\}. \end{cases} \tag{4}$$

3 An iterative search procedure

The iterative scheme we provide is inspired by the necessary conditions for optimality described earlier (eqn. 4). In this scheme, we find the optimal combination of policies for sampling and control through a possibly infinite sequence of policy iteration steps. In each round of the iteration, we execute two steps. In step one, we seek the best sampling policy for the previously derived control policy. This is an optimal stopping problem in standard form. In the second step, we utilize the newly computed sampling policy and seek the best control update policy that goes with it. The second step is thus an optimal control problem. This solution scheme leads to the optimal policies because of the team nature of the problem and because the system being controlled is linear. However, each round of the solution scheme is computationally intensive.

Definition 1 (Iterative search algorithm). For a given initial condition x_0, the iterative search algorithm to minimize the aggregate distortion starts with the decomposition of eqn. 3 and an admissible pair of a control level and a stopping time: (ν_0,θ_0). In each round of the algorithm, the pair (ν_k,θ_k) is transformed into a less costly pair (ν_{k+1},θ_{k+1}) via a sequence of two steps:
Step 1: The improved stopping time is given by:

$$\theta_{k+1} = \operatorname{ess\,sup}_\tau \ \mathbb{E}\left[\beta(x_0,\nu_k,\tau,T)\right].$$

Step 2: The improved control level is given by:

$$\nu_{k+1} = \inf_U \left\{\alpha(x_0,U,T) - \mathbb{E}\left[\beta(x_0,U,\theta_{k+1},T)\right]\right\}.$$

Notice that in each of the two steps, the cost may or may not be lowered but never increased. In practical implementation, the order of execution of the steps has no influence on the outcome of the algorithm. The order of steps we have given however, allows an interpretation of the procedure as policy iteration.

3.1 Equivalence to policy iteration

Theorem 1 (Global convergence of iterative search). *The procedure described in defn. 1 converges for any feasible pair of initial policies and the limit of the policy iterates is unique.*

Proof. Convergence of the iterations is easy to establish because at each step, the cost can only decrease or stay the same. Since the cost is lower bounded by zero, the sequence of iterates is necessarily convergent.

We discretize time and establish convergence and the uniqueness of the limit. We attack the continuous time problem using a continuity argument by making the time-discretizations infinitesimally fine. We use two key properties of the optimization problem to show that the iterates converge . The first fact is that the optimal stopping problem of step 1 of our algorithm has as its solution, the first hitting time of barriers subject to a time-out namely, the length of the horizon. This means that the search for a stopping time is the search for double-sided time-varying barriers. The latter search can be viewed as a problem of control up to an absorption time [Kushner(1971),Bertsekas & Tsitsiklis(1991)] also called a stochastic shortest path problem. Such a conversion is in general not possible for stopping time problems. But here, we use the finiteness of the horizon and the threshold crossing nature of the optimal stopping time to enable this conversion.

Consider the Markov process $\xi_n = \left(x_{[n]}, u_{[n]}\right)$. We stop this process when it is absorbed by a double sided barrier (of the sort that produce the optimal stopping time). We mandate that the upper and lower barriers are both equal to zero at the end time of the horizon T. This makes sure that all admissible trajectories are stopped at least by T.

We now specify the control sets. At time zero, $u_{[0]}$ can be chosen to be any real number. We also choose two levels (barriers Upp_n, Low_n with $Upp_n \geq Low_n$) to be applied to the state at the next discrete time step. This is a device which allows us to "control" the probability of getting absorbed at the next time tick. At all other times, the control set is such that $u_{[n]}$ cannot be influenced. But we are allowed to choose two real-valued levels to determine absorption at the next time-tick. In the penultimate and ultimate time-ticks, the control set is really the Null set. The dynamics of the discrete process ξ_k inherit the dynamics of the x and u processes. Notice that in our setup, u can be chosen arbitrarily at the zeroth time-tick but not influenced at all subsequently.

Why is this characterization useful to us ? Because, it casts the mapping: $(\nu_k, \{Upp_n(k), Low_n(k)\}) \rightarrow (\nu_{k+1}, \{Upp_n(k), Low_n(k)\})$ as policy iteration. In fact, the procedure of defn. 1 carries out policy iteration by backwards dynamic programming. If we start with finite levels for the barriers at each time tick, the transition kernel of the Markov process ξ_k is always strictly contractive. From this, the existence of a limit and its uniqueness follow [Kushner(1971),Bertsekas & Tsitsiklis(1991)].

Because the controlled process is continuous and the integrand of the cost is a C^2 function, as the time-discretizations get infinitesimally finer, the limiting policy and cost converge to the optimal one for the continuous time problem.

4 Single stopping for controlling Brownian motion

Here, we utilize the results of the previous section for a controlled Brownian motion process:

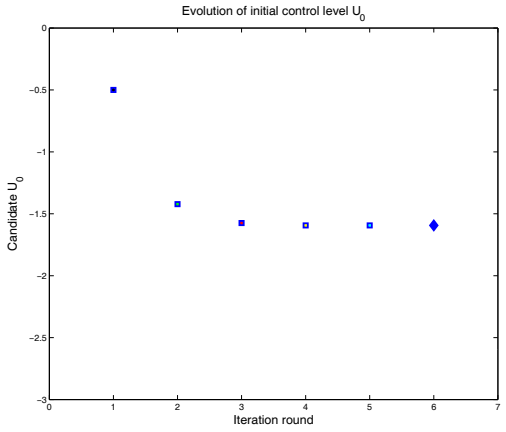
$$dx_t = u_t dt + dB_t.$$

The optimal terminal control level U_1^* is the linear feedback law [Rabi *et al.*(2008)Rabi, Johansson, & Johansson]:

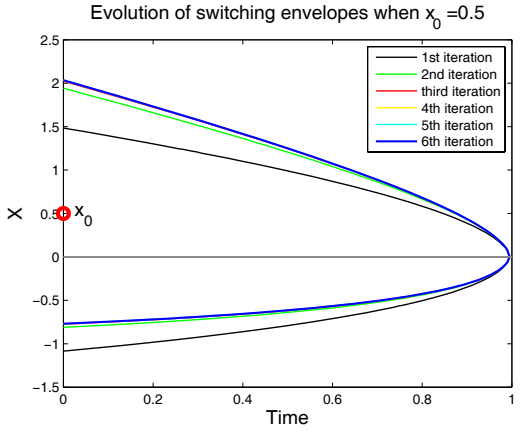
$$U_1^*\left(x_\tau, T-\tau\right) = -\frac{3x_\tau}{2(T-\tau)}.$$

Then, the variables U_0 and τ can be chosen by minimizing the cost:

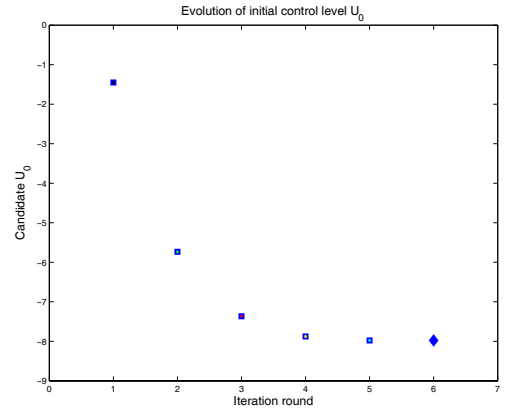
$$\tfrac{1}{T^2} J\left(x_0, \mathcal{U}, \tau\right) = \left(\tfrac{x_0}{\sqrt{T}}\right)^2 + \tfrac{1}{2} + \left(\frac{\frac{x_0}{\sqrt{T}}\sqrt{3}}{2} + \frac{U_0\sqrt{T}}{\sqrt{3}}\right)^2 - \mathbb{E}\left[\left(\frac{\frac{x_\tau}{\sqrt{T}}\sqrt{3}}{2} + \frac{U_0\sqrt{T}\left(1-\frac{\tau}{T}\right)}{\sqrt{3}}\right)^2 \left(1-\tfrac{\tau}{T}\right)\right].$$



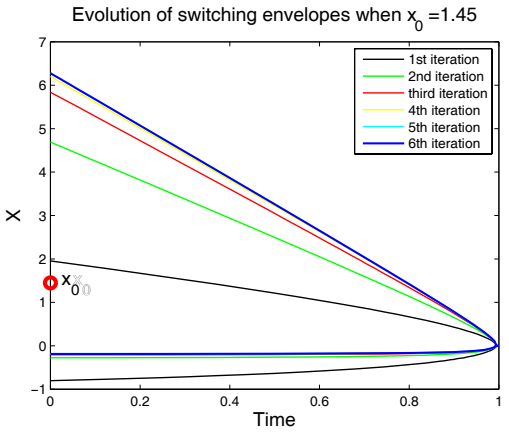
(a) Control level ν_k when $x_0 = 0.5$



(b) Envelopes for sampling time θ_k when $x_0 = 0.5$



(c) Control level ν_k when $x_0 = 1.45$



(d) Envelopes for sampling time θ_k when $x_0 = 1.45$

Fig. 1. Subfigs (a,b) show the results of the iterative search procedure for a relatively modest value of 0.5 for x_0. The convergence to the optimal pair happens by the third round and the minimum aggregate distortion $J\left(0.5, \mathcal{U}^*\left(0.5\right), \tau^*\left(0.5\right)\right) = 0.29$. Subfigs (c,d) show the results for a value of $x_0 = 1.45$. This puts the optimal initial control effort "on the edge" of becoming an impulse. Here, the optimal pair is obtained at round five and the minimum aggregate distortion $J\left(1.45, \mathcal{U}^*\left(1.45\right), \tau^*\left(1.45\right)\right) = 0.47$.

This suggests that the length of the horizon does not really matter for the optimization problem. We can use a new time scale with time $s = \frac{t}{T}$ and also use the change of variables: $\bar{x} = \frac{x_0}{\sqrt{T}}$ and $\bar{u} = U_0\sqrt{T}$. In this time-scale, the length of the horizon is exactly one. Hence we will simply assume that $T = 1$.

4.1 Case of zero initial condition

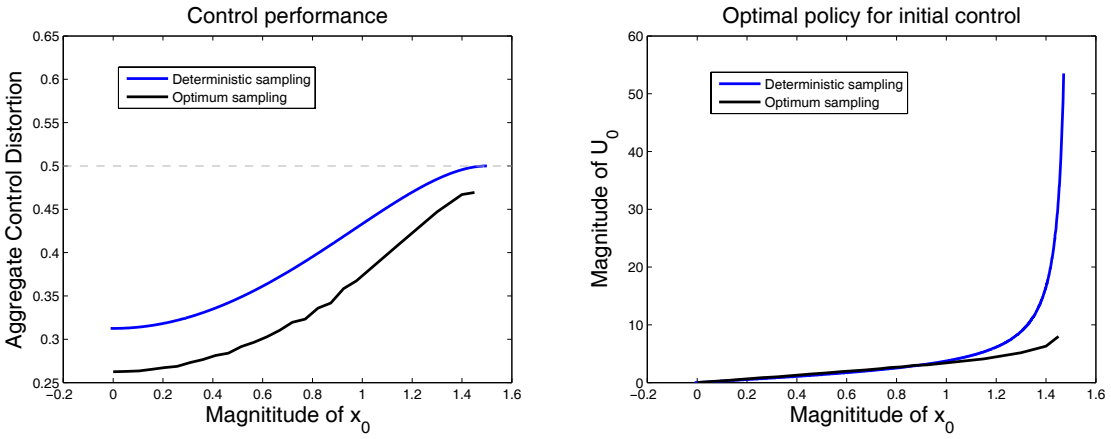
If the initial value is zero, then, the optimal control and its performance can be explicitly computed in closed form [Rabi *et al.*(2008)Rabi, Johansson, & Johansson]. Because of symmetry, U_0. The optimal stopping rule is the symmetric quadratic envelope:

$$\tau^* = \inf\left\{ t \middle| x_t^2 \geq \sqrt{3}\left(T - t\right) \right\}.$$

The expected control performance cost incurred by the optimal switching scheme is then $\frac{3\sqrt{3}-1}{16}T^2$, while the cost of using deterministic switching is $\frac{5}{16}T^2$.

4.2 Starting away from zero

One admissible strategy is to have a very large magnitude initial control level U_0 to drive the state quickly towards zero and to switch off the control level when the state does reach zero. Since we do not penalize



(a) Relative performance (b) Initial controls

Fig. 2. Subfig (a) shows the minimum aggregate control error $J(x_0, \mathcal{U}, \tau)$ for Brownian Motion with one allowed sample due to deterministic switching and due to optimal switching. When the initial value x_0 approaches 1.5, the cost approaches the critical value of 0.5. Subfig (b) shows the initial control policy mapping the initial value x_0 to the initial control level U_0. Notice that the optimal switching scheme is less aggressive than the deterministic one when x_0 approaches 1.5.

control effort directly, the cost of this strategy if not impacted by the magnitude of the initial condition. This cost is actually equal to 0.5. This means that for a given initial condition, if no admissible policy can incur a cost lesser than 0.5, the one we described will be optimal. This situation does occur for a range of initial values (See fig. 2).

Optimal sampling We determine the optimal initial control level and the optimal envelope (double-side barrier) by the iterative algorithm of defn. 1. Notice that for x_0 larger than 1.5, immediate resetting to zero is the best option.

5 Conclusions and Further remarks

It is important to extend the results of this paper to the general piece-wise deterministic class of control waveforms. Extension to the vector case and the partially observed case would also be useful. Finally, we need to examine the impact of different control performance measures.

References

[Antsaklis & Baillieul(2007)]Antsaklis, P. & Baillieul, J. (2007). Control and communication challenges in networked real-time systems. *Proceedings of the IEEE*, *95*(1), 9–28. Special issue on Technology of Networked Control Systems.

[Åström & Bernhardsson(2002)]Åström, K. J. & Bernhardsson, B. (2002). Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *Proceedings of the 41st IEEE conference on Decision and Control (Las Vegas NV, 2002)*. IEEE Control Systems Society, 2011–2016.

[Bertsekas & Tsitsiklis(1991)]Bertsekas, D. P. & Tsitsiklis, J. N. (1991). An analysis of stochastic shortest path problems. *Math. Oper. Res.*, *16*(3), 580–595.

[Kushner(1971)]Kushner, H. (1971). *Introduction to stochastic control*. Rinehart and Winston, Inc., New York: Holt.

[Kushner(1964)]Kushner, H. J. (1964). On the optimum timing of observations for linear control systems with unknown initial state. *IEEE Trans. Automatic Control*, *AC-9*, 144–150.

[Øksendal(2003)]Øksendal, B. (2003). *Stochastic differential equations*. Universitext. Berlin: Springer-Verlag, sixth ed.

[Rabi *et al.*(2008)Rabi, Johansson, & Johansson]Rabi, M., Johansson, K. H., & Johansson, M. (2008). Optimal stopping for event-triggered sensing and actuation. In *Proceedings of the 47th IEEE Conference on Decision and Control*.