Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

Multiple Sampling for Estimation on a Finite Horizon*

Maben Rabi, George V. Moustakides, and John S. Baras

***Abstract*— We discuss some multiple sampling problems that arise in finite horizon real-time estimation when there is an upper limit on the number of allowable samples. Measuring estimation quality by the aggregate squared error, we compare the performances of the best deterministic, level-triggered and the optimal sampling schemes. We restrict the signal to be either a Wiener or an Ornstein-Uhlenbeck process. For the Wiener process, we provide closed form expressions and series expansions, whereas for the Ornstein Uhlenbeck process, procedures for numerical computation. Our results indicate that the best level-triggered sampling is almost optimal when the signal is stable.**

I. Event-triggered sampling

In many decision and control problems, we can impose a certainty-equivalence type separation into a signal estimation part and a control or decision design part. For example, an optimal control or a signal detection problem can be solved by certainty-equivalence policies which treat a least-squares estimate of the signal waveform as the true state. In these situations, the processing of available measurements should be geared towards obtaining the best quality signal estimate. In Networked Control Systems [1] where the sensors have only a limited number of packets (samples) to transmit to the supervisor, the sampling design affects the quality of the signal estimate.

Kushner [2] has treated a problem of picking a fixed number of deterministic sampling times for a finite horizon linear optimal control problem. He establishes the validity of the separation principle and obtains closed form expressions for the minimum cost in the scalar case. The collection [3] treats some randomized as well as deterministic but irregular sampling schemes for smoothing and control. Cambanis and Masry [4] have treated the problem of picking the best deterministic and random sampling schedules for hypothesis testing based on a smoothed estimate. Their random sampling schedules are however not adapted to the signal observed by the sensor.

For the problems treated in this paper, we seek to characterize the performance gains provided by event-trigerred sampling policies. Event-triggered sampling has been referred to as 'Lebesgue-type' sampling in the control literature [5]. We solve a sampling design problem within three classes of sampling strategies with the sampling design tailored to a signal filtering task. The design objective is to minimize, over a finite horizon, the distortion of a filter (real-time estimator) of the signal based upon the stream of samples. This minimization is performed with a fixed constraint on the maximum number of samples used. In [6] a related problem in discrete-time is treated.

For the signal to be estimated (the state process) $x_t,\ t \in [0,T]$, we will assume that a sensor has perfect observation of the state and transmits at times it chooses, current samples of the state process. The sensor is allowed to generate *at most* N samples to be transmitted to a supervisor. The sampling times $S = \{\tau_1, \dots, \tau_N\}$ have to be an increasing sequence of stopping times with respect to the x-process. They also have to lie within the interval $[0,T]$. Based on the samples and the sampling times, the least-squares estimate for the supervisor $\hat{x}_t$ is given by [7]:

$$\hat{x}_t = \begin{cases} \mathbb{E}[x_t|\mathscr{F}_0] & \text{if } 0 \le t < \tau_1, \\ \mathbb{E}[x_t|\mathscr{F}_{\tau_i}] & \text{if } \tau_i \le t < \tau_{i+1} \le \tau_N, \\ \mathbb{E}[x_t|\mathscr{F}_{\tau_N}] & \text{if } \tau_N \le t \le T. \end{cases} \tag{1}$$

The quality of this estimate is measured by the aggregate squared error distortion:

$$\begin{aligned} J(S) &= \mathbb{E}\left[\int_0^T (x_s - \hat{x}_s)^2 ds\right] \\ &= \mathbb{E}\Bigg[\int_0^{\tau_1} (x_s - \hat{x}_s)^2 ds + \sum_{i=2}^{N} \int_{\tau_{i-1}}^{\tau_i} (x_s - \hat{x}_s)^2 ds \\ &\quad + \int_{\tau_N}^{T} (x_s - \hat{x}_s)^2 ds\Bigg] \end{aligned} \tag{2}$$

We will consider three sampling strategies and characterize their performance. The strategies are:

Deterministic sampling: The sampling sequence S is chosen apriori and hence independent of the signal trajectory. It is chosen to minimize the expected distortion J. In this scheme, the supervisor too knows in advance when the sensor will generate and transmit samples.

Level-triggered sampling: The sensor chooses the sampling times based on times the error signal $x_t - \hat{x}_t$ crosses chosen thresholds. The actual sampling times are the lesser of these threshold times and the end time T. The sampling times are dependent on the actual sensor observations.

Optimal sampling: To choose sampling times, the sensor solves a sequence of optimal stopping time problems and applies the resulting stopping rule. Here too, these times are dependent on the actual sensor observations.

Let us now start with a detail on the single sampling instance, i.e. $N = 1$, since this is going to serve as a basis for solving the general case.

*This research was supported by the Army Research Office under the ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Networked Communicating Control Systems (through Boston University), and under ARO Grant No. DAAD190210319.

Maben Rabi and John S. Baras are at the Institute for Systems Research and Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA. George Moustakides is at the Department of Electrical and Computer Engineering, University of Patras, 26500 Rio, Greece

A. Single Sample case

We can express the distortion J as follows:

$$\begin{aligned} J(\tau_1) &= \mathbb{E}\left[\int_0^{\tau_1} x_s^2 + \int_{\tau_1}^T (x_s - \hat{x}_s)^2\, ds\right] \\ &= \mathbb{E}\left[\int_0^T x_s^2 - 2\int_{\tau_1}^T x_s\hat{x}_s\, ds + \int_{\tau_1}^T (\hat{x}_s)^2\, ds\right]. \end{aligned}$$

Now notice that the second term can be written as follows

$$\mathbb{E}\left[\int_{\tau_1}^T x_s\hat{x}_s\, ds\right] = \mathbb{E}\left[\int_{\tau_1}^T \mathbb{E}[x_s|\mathscr{F}_{\tau_1}]\hat{x}_s\, ds\right] = \mathbb{E}\left[\int_{\tau_1}^T (\hat{x}_s)^2\, ds\right],$$

where we have used the fact that $\hat{x}_s$ is $\mathscr{F}_{\tau_1}$-measurable. Because of this observation the performance measure $J(\tau_1)$ takes the form

$$J(\tau_1) = \int_0^T \mathbb{E}\left[x_s^2\right] ds - \mathbb{E}\left[\int_{\tau_1}^T (\hat{x}_s)^2\, ds\right]. \tag{3}$$

The above expression is valid for all stopping times τ_1 satisfying $0 \le \tau_1 \le T$. We also observe that the first term is constant not depending on the sampling strategy. Next we are going to focus on sampling the Wiener and the Ornstein-Uhlenbeck process and attempt to quantify the performance of the three sampling strategies in the case of single and multiple samples.

II. Sampling the Wiener Process

Here our signal is defined with the help of the following SDE

$$dx_t = dW_t, \quad t \in [0, T],$$

with $x_0 = 0$ and W_t a standard Wiener process.

A. Single sample case

The least-squares estimate in this case takes the simple form [7]:

$$\hat{x}_t = \mathbb{E}[x_t|\mathscr{F}_{\tau_i}] = x_{\tau_i},$$

which when substituted in (3) yields the following single sample performance

$$J(\tau_1) = \frac{T^2}{2} - \mathbb{E}\left[x_{\tau_1}^2 (T - \tau_1)\right].$$

1) Optimum deterministic sampling: When τ_1 is deterministic, the distortion becomes:

$$J(\tau_1) = \frac{T^2}{2} - (T - \tau_1)\mathbb{E}\left[x_{\tau_1}^2\right] = \frac{T^2}{2} - (T - \tau_1)\,\tau_1, \tag{4}$$

which is straightforwardly minimized for $\tau_1 = 0.5T$ resulting in $\min_{\tau_1} J(\tau_1) = \frac{T^2}{4}$.

2) Optimum level-triggered sampling: For a given $\eta \ge 0$, let τ_η denote the following level-crossing time:

$$\tau_\eta = \inf_{t \ge 0}\{t : |x_t| \ge \eta\}. \tag{5}$$

The actual sampling time is given by $\tau_1 = \tau_\eta \wedge T$, and the distortion by

$$J(\tau_1) = \frac{T^2}{2} - \eta^2 \mathbb{E}\left[(T - \tau_\eta)^+\right].$$

We do not have a closed form expression for the pdf of the stopping time τ_η. There is a series expansion provided in page 99 of [8] which is not directly useful to our calculations. Instead, we compute a version of the generating function of τ_η.

Lemma 2.1: Given that $x_0 = 0$ and the threshold η,

$$\mathbb{E}\left[e^{-s\tau_\eta}\right] = \frac{1}{\cosh(\eta\sqrt{2s})} = F_\eta(s).$$

Proof: Apply the Itô formula on the function $h(x,t) = e^{-st}[1 - \frac{\cosh(x\sqrt{2s})}{\cosh(\eta\sqrt{2s})}]$ to obtain:

$$\begin{aligned} &\mathbb{E}\left[h\left(x_{\tau_\eta}, \tau_\eta\right) - h(0,0)\right] \\ &\quad = \mathbb{E}\left[\int_0^{\tau_\eta}\left\{h_t(x_t,t) + \frac{1}{2}h_{xx}(x_t,t)\right\}dt\right], \\ &\quad = \mathbb{E}\left[e^{-s\tau_\eta}\right] - 1. \end{aligned}$$

Since, $h(x_{\tau_\eta}, \tau_\eta) = 0$, the result holds. ■

We know that the pdf $f_\eta(t)$ of τ_η is the inverse Laplace transform of $F_\eta(s)$, therefore we can write

$$\begin{aligned} \mathbb{E}\left[(T - \tau_\eta)^+\right] &= \int_0^T (T - t) f_\eta(t) dt, \\ &= \int_0^T (T - t)\left[\frac{1}{2\pi j}\oint F_\eta(s) e^{st} ds\right] dt, \\ &= \frac{1}{2\pi j}\oint F_\eta(s)\left[\int_0^T (T - t) e^{st} dt\right] ds, \\ &= \frac{1}{2\pi j}\oint \frac{e^{sT} - 1 - sT}{s^2\cosh(\eta\sqrt{2s})} ds. \end{aligned}$$

The last integral is on a contour which encompasses the entire left half plane. We compute it through an application of the residue theorem. Firstly, $s = 0$ is not a pole of the integrand as it is a double zero for its numerator. The only poles of the integrand come from zeroes of the function $\cosh(\eta\sqrt{2s})$. These are $s_k = -(2k+1)^2\frac{\pi^2}{8\eta^2},\ k = 0, 1, 2, \ldots,$ and they all lie inside the left half plane. The contour integral we need to compute is given through a sum of the residues at all of these poles:

$$\begin{aligned} \mathbb{E}\left[(T - \tau_\eta)^+\right] &= \frac{1}{2\pi j}\oint \frac{e^{sT} - 1 - sT}{s^2\cosh(\eta\sqrt{2s})} ds, \\ &= \sum_{k=0}^{\infty} \frac{e^{s_k T} - 1 - s_k T}{s_k^2} \lim_{s \to s_k} \frac{s - s_k}{\cosh(\eta\sqrt{2s})}, \\ &= \sum_{k=0}^{\infty} \frac{e^{s_k T} - 1 - s_k T}{s_k^2} \times (-1)^{(k+1)} \frac{4 s_k}{\pi(2k+1)}. \end{aligned}$$

Using the above result, the distortion can be written as:

$$J(\tau_\eta) = \frac{T^2}{2}\varphi(\lambda),$$

where $\lambda = \frac{\pi^2 T}{8\eta^2}$, and

$$\varphi(\lambda) = 1 - \frac{\pi}{\lambda^2}\sum_{k=0}^{\infty}(-1)^k \frac{e^{-(2k+1)^2\lambda} - 1 + (2k+1)^2\lambda}{(2k+1)^3}. \quad (6)$$

Because of the above relations, we can see that minimizing the distortion J with respect to the threshold η is the same as minimizing it with respect to λ:

$$\inf_{\eta} J(\tau_\eta) = \inf_{\lambda} \frac{T^2}{2}\varphi(\lambda).$$

If λ_0 minimizes $\varphi(\cdot)$, then the optimum threshold η_0 is given by:

$$\eta_0 = \frac{\pi}{2\sqrt{2\lambda_0}}\sqrt{T}.$$

Numerically evaluating λ_0, we obtain $\eta_0 = 0.9391\sqrt{T}$ and the corresponding minimum distortion to be $0.3952(T^2/2)$. This is about a 20% improvement over the minimum distortion of the deterministic sampling scheme: $0.5(T^2/2)$.

3) Optimum single sampling: We seek the sampling strategy that minimizes the distortion. We seek a stopping time τ_1 satisfying $0 \leq \tau_1 \leq T$ and minimizing:

$$J(\tau_1) = \frac{T^2}{2} - \mathbb{E}\left[x_{\tau_1}^2 (T - \tau_1)\right].$$

Only the second term depends on the stopping time τ_1. Furthermore, relaxing the constraint $\tau_1 \leq T$ does not change the optimum or the optimizing policy.

So, we are interested in finding a stopping time τ_1 that maximizes the following expected reward:

$$\mathbb{E}\left[x_{\tau_1}^2 (T - \tau_1)\right].$$

Consider the following candidate *maximum expected reward function* (*the Snell envelope* [9]):

$$g(x,t) = A\left\{\frac{1}{2}(T-t)^2 + x^2(T-t) + \frac{x^4}{6}\right\},$$

where A is a constant to be specified subsequently. Since $x_t = W_t$, using Itô calculus it is straightforward to prove

$$dg(x_t,t) = A\left\{2W_t(T-t) + \frac{2}{3}W_t^3\right\}dW_t.$$

If τ is any stopping time, we have:

$$\mathbb{E}\left[g(x_\tau,\tau) - g(x_0,0)\right] = \mathbb{E}\left[\int_0^\tau dg(x_t,t)\right] = 0,$$

which leads us to:

$$\mathbb{E}\left[g(x_\tau,\tau)|x_0\right] = g(x_0,0) = A\left\{\frac{1}{2}T^2 + x_0^2(T) + \frac{x_0^4}{6}\right\}.$$

Let us now pick A such that $g(x,t) \geq x^2(T-t)$ but also with equality for some family of pairs (x,t). Then we observe that for

$$A = \frac{\sqrt{3}}{1+\sqrt{3}}$$

the difference:

$$g(x,t) - x^2(T-t) = A\left(\frac{x^2}{\sqrt{6}} - \frac{T-t}{\sqrt{2}}\right)^2 \geq 0,$$

becomes a perfect square. And we do have equality for pairs (x,t) such that $x^2 = \sqrt{3}(T-t)$. Thus the optimal stopping rule is given by:

$$\tau_1^* = \inf_t\left\{t : x_t^2 \geq \sqrt{3}(T-t)\right\}.$$

The corresponding minimum distortion is given by:

$$J(\tau_1^*) = \frac{T^2}{2} - \frac{\sqrt{3}}{1+\sqrt{3}}\frac{T^2}{2} = \frac{1}{1+\sqrt{3}}\frac{T^2}{2} = 0.366\frac{T^2}{2},$$

which is smaller than the corresponding optimum level triggering scheme.

B. N-Sample case

We will now use the results of section II-A to characterize the performance of the three sampling strategies when the allowed number of samples is more than one.

1) Deterministic sampling: We will show through induction that uniform sampling on the interval $[0,T]$ is the optimal deterministic choice for N samples $0 \leq \tau_1 \leq \tau_2, \leq \cdots \leq \tau_N \leq T$, given that the initial value of the signal is zero. For N samples, the distortion takes the form given in (2) which we denote as $J_{[0,T]}(\tau_1, \tau_2, \ldots \tau_N)$. If we assume that the optimal choice of $N-1$ deterministic samples over $[T_1, T_2]$ is the uniform one, that is

$$\tau_i = T_1 + i\frac{T_2 - T_1}{N},\ i = 1,2,\ldots,N-1,$$

the corresponding minimum distortion becomes:

$$J_{[T_1,T_2]}(\tau_1,\ldots,\tau_{N-1}) = \frac{(T_2-T_1)^2}{2N}.$$

Notice now that the minimum distortion over the set of N sampling times can be written as

$$\begin{aligned}
&\min_{\tau_1,\tau_2,\ldots\tau_N} J_{[0,T]}(\tau_1,\tau_2,\ldots\tau_N) \\
&= \min_{\tau_1}\left\{\int_0^{\tau_1}(x_s - \hat{x}_s)^2 ds + \min_{\tau_2,\tau_2,\ldots\tau_N} J_{[\tau_1,T]}(\tau_2,\ldots\tau_N)\right\} \\
&= \min_{\tau_1}\left\{\frac{{\tau_1}^2}{2} + \frac{(T-\tau_1)^2}{2N}\right\} = \frac{T^2}{2(N+1)},
\end{aligned}$$

the minimum being achieved for $\tau_1 = T/(N+1)$. This proves the assertion about the optimality of uniform sampling.

2) Level triggered sampling: Here, the sampling times are defined with the help of N thresholds $\eta_i \geq 0,\ i=1,\ldots,N$ as follows

$$\tau_{i,\eta_i} = \inf_{t\geq\tau_{i-1}} \left\{t : |x_t - x_{\tau_{i-1}}| \geq \eta_i\right\},$$
$$\tau_i = \min\left\{\tau_{i,\eta_i}, T\right\},\ \tau_0 = 0.$$

Based on the discussion in section(II-A.2), we can write down the minimal distortion for the level-trigerred scheme with a single sample over $[T_1, T]$. allowed. It is:

$$c_1 \frac{\left[(T-T_1)^+\right]^2}{2} = 0.3952 \frac{\left[(T-T_1)^+\right]^2}{2}.$$

Like in the single sample case, we will show that the expected distortion over $[0,T]$ given at most N samples is of the form

$$c_N \frac{T^2}{2}.$$

The minimal expected cost over $[0,T]$ for an optimal stopping problem of the type (7) below g, is of the form of the terminal cost expression in (7). Let τ_η be the level-crossing time of equation (5). Then, given a positive real number α, the following minimal cost

$$\min_{\eta\geq 0} J(\eta) = \min_{\eta\geq 0} \mathbb{E}\left[\int_0^{\tau_\eta\wedge T} x_s^2 ds + \alpha\left[(T-\tau_\eta)^+\right]^2\right] \quad (7)$$

turns out to be of the form $\beta\left[(T-\tau_\eta)^+\right]^2$, where $\beta > 0$ depends only on α.

We will now prove this useful fact. Notice that:

$$d\left[(T-t)x_t^2\right] = -x_t^2 dt + 2(T-t)x_t dx_t + (T-t)dt,$$

and that,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{\tau_\eta\wedge T} x_s^2 ds\right] \\
&= \mathbb{E}\left[(T-\tau_\eta\wedge T)x_{\tau_\eta\wedge T}^2 + \frac{T^2}{2} - \frac{1}{2}(T-\tau_\eta\wedge T)^2\right] \\
&= \frac{T^2}{2} - \mathbb{E}\left[\eta^2(T-\tau_\eta)^+ + \frac{1}{2}\left[(T-\tau_\eta)^+\right]^2\right].
\end{aligned}$$

Thus, the cost (7) becomes:

$$J(\eta) = \frac{T^2}{2}\left[\varphi(\lambda) + \left(\frac{1}{2}-\alpha\right)\psi(\lambda)\right],$$

where we have followed the notation of section(II-A.2) with λ and φ defined in (6). Function ψ is also given as a series expansion in the following relation

$$\psi(\lambda) = \tfrac{16}{\pi\lambda^2}\textstyle\sum_{k=0}^{\infty}(-1)^k \tfrac{e^{-(2k+1)^2\lambda} - 1 + (2k+1)^2\lambda - (1/2)(2k+1)^4 x^2}{(2k+1)^5}.$$

Then we have the optimal cost (7) as:

$$\min_{\eta\geq 0} J(\eta) = \frac{T^2}{2}\inf_\lambda\left\{\varphi(\lambda) + (0.5-\alpha)\psi(\lambda)\right\}.$$

Based on the above discussion, we can define c_k recursively as follows: For $k \geq 2$,

$$\begin{aligned}
c_k &= \inf_\lambda\left\{\varphi(\lambda) + (0.5 - c_{k-1})\psi(\lambda)\right\}, \\
\lambda_k^* &= \arg\inf_\lambda\left\{\varphi(\lambda) + (0.5 - c_{k-1})\psi(\lambda)\right\}, \\
\rho_k &= \frac{\pi}{2\sqrt{2\lambda_k^*}}.
\end{aligned}$$

The optimal set of thresholds are given by:

$$\eta_k^* = \rho_{N-k+1}\sqrt{T-\tau_{k-1}}.$$

3) Optimal multiple sampling: Exactly like in the previous discussion on multiple level-triggered sampling, we will obtain a parametric expression for the minimal expected distortion given at most N samples. Analogous to equation (7), consider the stopping cost:

$$J(\tau) = \mathbb{E}\left[\int_0^{\tau\wedge T} x_s^2 ds + \frac{\alpha}{2}\left[(T-\tau)^+\right]^2\right] \quad (8)$$

where $\alpha \geq 0$ is a given constant. We can rewrite this as:

$$\frac{1}{2}\left\{T^2 - \mathbb{E}\left[2x_{\tau\wedge T}^2(T-\tau)^+ + (1-\alpha)\left[(T-\tau)^+\right]^2\right]\right\}.$$

Note that there is no change in optimality by permitting τ to take values bigger than T. In fact the optimal τ even with this relaxation will a.s. be less than T. Like in the single sample case, let us pay attention to the part of the above expression which depends on τ and define the following optimal stopping problem:

$$\min_\tau \mathbb{E}\left[2x_\tau^2(T-\tau) + (1-\alpha)(T-\tau)^2\right].$$

Consider the candidate *maximum expected reward function*:

$$g(x,t) = A\left\{(T-t)^2 + 2x^2(T-t) + \frac{x^4}{3}\right\}.$$

where A is a constant chosen such that $g(x,t) - 2x^2(T-t) - (1-\alpha)(T-t)^2$ becomes a perfect square. The only possible value for A then is $A = \frac{(5+\alpha)-\sqrt{(5+\alpha)^2-24}}{4}$. Then the optimal stopping time is given by:

$$\begin{aligned}
\tau^* &= \inf_t\left\{t : g(x_t,t) \leq 2x_t^2(T-t) + (1-\alpha)(T-t)^2\right\}, \\
&= \inf_t\left\{t : x_t^2 \geq \sqrt{\frac{3(A-1+\alpha)}{A}}(T-t)\right\},
\end{aligned}$$

and the corresponding optimal distortion J becomes

$$J = (1-A)\frac{T^2}{2}.$$

We obtain the explicit stopping rules and the corresponding minimal distortions for different values of the sample budget N by defining recursively κ_N, γ_N:

$$\begin{aligned}
\kappa_N &= 1 - \frac{(5+\kappa_{N-1}) - \sqrt{(5+\kappa_{N-1})^2 - 24}}{4}, \\
\gamma_N &= \sqrt{\frac{3(\kappa_{N-1}-\kappa_N)}{1-\kappa_N}}.
\end{aligned}$$

The $(k+1)^{\text{th}}$ sampling time is chosen as:

$$\tau_{k+1} = \inf_{t\geq\tau_k}\left\{t : \left(x_t - x_{\tau_k}\right)^2 \geq \gamma_{N-k+1}\sqrt{T-t}\right\}.$$

4) Comparisons: We list below a numerical comparison of the aggregate filtering distortions incurred by the three sampling strategies on the same time interval $[0,T]$. We obtained the distortions for all sampling strategies as a product of $T^2/2$ and a positive coefficient. The numbers listed in the table are these coefficients.

N	1	2	3	4
Deterministic	0.5	0.333	0.25	0.2
Level-triggered	0.3953	0.3471	0.3219	0.3078
Optimal	0.3660	0.2059	0.1388	0.1032

It is rather surprising that deterministic sampling outperforms the best level triggering scheme when we use more than one samples.

III. Sampling an Ornstein Uhlenbeck process

Now we will turn to the case when the signal is an Ornstein Uhlenbeck process satisfying the SDE

$$dx_t = ax_t dt + dW_t, \quad t \in [0,T], \tag{9}$$

with $x_0 = 0$ and W_t, as before, being a standard Wiener process. For the signal estimate $\hat{x}_t$ we have that

$$\hat{x}_t = \mathbb{E}[x_t|\mathscr{F}_{\tau_i}] = x_{\tau_i}e^{a(t-\tau_i)}, \tag{10}$$

and the quality of this estimate is measured by the aggregate squared error distortion defined in (2). Let us examine the three sampling policies.

A. Optimum deterministic sampling

As in the Wiener case, we will show through induction that uniform sampling on the interval $[0,T]$ is the optimal deterministic choice of N samples at $0 \leq \tau_1 \leq \cdots \leq \tau_N \leq T$. Denoting by $J_{[0,T]}(\tau_1,\ldots,\tau_N)$ the distortion using N samples, let us assume that the optimal choice of $N-1$ deterministic samples over $[T_1,T_2]$ is the set of uniformly spaced samples between T_1 and T_2, i.e. $\tau_i = T_1 + i(T_2 - T_1)/N$ which yields the following distortion

$$J_{[T_1,T_2]}(\tau_1,\ldots,\tau_{N-1}) = \frac{N}{4a^2}\left(e^{2a\frac{T_2-T_1}{N}} - 1\right) - \frac{1}{2a}(T_2-T_1).$$

Using this for our induction, we have that the minimum distortion over the set of N sampling times is:

$$\begin{aligned}
&\min_{\tau_1,\tau_2,\ldots\tau_N} J_{[0,T]}(\tau_1,\tau_2,\ldots\tau_N)\\
&= \min_{\tau_1}\left\{\int_0^{\tau_1}(x_s-\hat{x}_s)^2 ds + \min_{\tau_2,\tau_2,\ldots\tau_N} J_{[\tau_1,T]}(\tau_2,\ldots\tau_N)\right\}\\
&= \min_{\tau_1}\left\{\frac{1}{4a^2}\left(e^{2a\tau_1}-1\right) + \frac{N}{4a^2}\left(e^{2a\frac{T-\tau_1}{N}}-1\right) - \frac{1}{2a}T\right\}\\
&= \frac{N+1}{4a^2}\left(e^{2a\frac{T}{N+1}}-1\right) - \frac{1}{2a}T,
\end{aligned}$$

the minimum being achieved for $\tau_1 = T/(N+1)$. Thus, we have the uniform sampling scheme being the optimal one here as well.

B. Optimum level-triggered sampling

Let us first address the single sample case. From (3) and using (10) we conclude that

$$\begin{aligned}
J(\tau_1) &= \mathbb{E}\left[\int_0^T x_t^2\,dt - \int_{\tau_1}^T (\hat{x}_t)^2\,dt\right]\\
&= \frac{e^{2aT}-1-2aT}{4a^2} - \mathbb{E}\left[x_{\tau_1}^2\frac{e^{2a(T-\tau_1)}-1}{2a}\right]\\
&= T^2\left\{\frac{e^{2aT}-1-2aT}{4(aT)^2} - \mathbb{E}\left[\frac{x_{\tau_1}^2}{T}\frac{e^{2(aT)(1-\tau_1/T)}-1}{2(aT)}\right]\right\}\\
&= T^2\left\{\frac{e^{-2\bar{a}}-1+2\bar{a}}{4\bar{a}^2} - \mathbb{E}\left[-\bar{x}_{\bar{\tau}_1}^{\;2}\frac{e^{2\bar{a}(1-\bar{\tau}_1)}-1}{2\bar{a}}\right]\right\}
\end{aligned}$$

where

$$\bar{t} = \frac{t}{T}, \quad \bar{a} = aT, \quad \bar{x}_{\bar{t}} = \frac{x_{\frac{t}{T}}}{\sqrt{T}}. \tag{11}$$

We have $\bar{x}$ satisfying the following SDE:

$$d\bar{x}_{\bar{t}} = -\bar{a}\bar{x}g_{\bar{t}}d\bar{t} + dW_{\bar{t}}.$$

This suggests that, without loss of generality, we can limit ourselves to the normalized case $T=1$ since the case $T \neq 1$ can be reduced to it by using the transformations in (11) and thus solve the multiple sampling problem on $[0,1]$.

We carry over the definitions for threshold sampling times from section II-B.2 We do not have series expansions like for the case of the Wiener process. Instead we have a computational procedure that involves solving a PDE initial and boundary value problem. The distortion corresponding to a chosen η_1 is given by:

$$\begin{aligned}
J(\eta_1) &= \frac{1}{4a^2}\left(e^{2a}-1\right) - \frac{1}{2a} - \frac{\eta_1^2}{2a}\mathbb{E}\left[e^{2a(1-\tau_1)}-1\right]\\
&= \frac{1}{4a^2}\left(e^{2a}-1\right) - \frac{1}{2a} - \frac{\eta_1^2}{2a}\left(e^{2a}\left(1+2aU^1(0,0)\right)-1\right),
\end{aligned}$$

where the function $U^1(x,t)$ satisfies the PDE:

$$\frac{1}{2}U^1_{xx} + axU^1_x + U^1_t + e^{-2at} = 0,$$

along with the boundary and initial conditions:

$$\begin{cases} U^1(-\eta_1,t) = U^1(\eta_1,t) = 0 & \text{for } t \in [0,1],\\ U^1(x,1) = 0 & \text{for } x \in [-\eta_1,\eta_1]. \end{cases}$$

We choose the optimal η_1 by computing the performances for values of η_1 for progressively increasing from 0. We stop when the cost stops decreasing and starts increasing. Note that the solution $U^1(0,t)$ to the PDE furnishes us with the performance of the η_1-triggered sampling over $[t,1]$. We will use this to solve the multiple sampling problem.

We use the numerical computation of the optimal distortion for the N sample case to compute the performance of the optimal $N+1$ sample threshold sampling policy. Let the optimal policy of choosing N levels for sampling over $[T_1,1]$ be given where $0 \leq T_1 \leq 1$. Let the resulting distortion be also known as a function of T_1. Let this known distortion over $[T_1,1]$ given N level-triggered samples be denoted by $G_N(1-T_1)$. Then, the $N+1$ sampling problem can be solved as follows. Let $U^{N+1}(x,t)$ satisfy the PDE:

$$\frac{1}{2}U^{N+1}_{xx}+axU^{N+1}_x+U^{N+1}_t=0,$$

along with the boundary and initial conditions:

$$\begin{cases} U^{N+1}(-\eta_1,t)=U^{N+1}(\eta_1,t)=G_N(1-t), \text{ for } t\in[0,1], \\ U^{N+1}(x,T)=0, \text{ for } x\in[-\eta_1,\eta_1]. \end{cases}$$

Then the distortion we are seeking to minimize over η_1 is given by:

$$\begin{aligned} J(\eta_1) &= \frac{1}{4a^2}\left(e^{2a}-1\right)-\frac{1}{2a}+\mathbb{E}\left[G_N(1-\tau_1)\right] \\ &-\frac{\eta_1^2}{2a}\mathbb{E}\left[e^{2a(1-\tau_1)}-1+\frac{1}{4a^2}\left(e^{2a(1-\tau_1)}-1\right)-\frac{1}{2a}(1-\tau_1)\right] \\ &=\frac{1}{4a^2}\left(e^{2a}-1-2a\right)-\frac{\eta_1^2}{2a}\{e^{2a}[1+2aU^1(0,0)]-1\}-U^{N+1}. \end{aligned}$$

We choose the optimal η_1 by computing the resultant distortion for increasing values of η_1 and stopping when the distortion stops decreasing.

C. *Optimal sampling*

We do not have analytic expressions for the minimum distortion like in the Wiener process case, neither we can reduce the problem to a PDE with well defined boundary conditions. Unfortunately the PDE we obtain is of *free boundary* type, a fact that makes it difficult to solve the problem even numerically using standard PDE solvers. We therefore reduce the problem to discrete time by finely discretizing time and solving the corresponding discrete-time optimal stopping problems.

By discretizing time, we get random variables $x_0,x_1,\ldots,x_M$, that satisfy the AR(1) model below. We have $x_0=0$ and for $1\le n\le M$

$$x_n=e^{a\delta}x_{n-1}+w_n, \quad w_n\sim\mathcal{N}\left(0,\frac{e^{2a\delta}-1}{2a}\right); \quad 1\le n\le M.$$

The sequence $\{w_n\}$ is an i.i.d. Gaussian sequence.

Sampling once in discrete time means selecting a sample x_ν from the set of $M+1$ sequentially available random variable $x_0,\ldots,x_M$, with the help of a stopping time $\nu\in\{0,1,\ldots,M\}$. The optimum cost to go can be analyzed as follows. For $n=M,M-1,\ldots,0$, denote the minimum distortion incurred by using only one sample in $[0,T]$ by $V^1_n(x)$. The superscript refers to the number of samples allowed and the subscript to the minimum distotion incurred by sampling at discrete time instants no less than n given that $x_n=x$.

$$\begin{aligned} V^1_n(x) &= \sup_{n\le\nu\le M}\mathbb{E}\left[x_\nu^2\frac{e^{2a\delta(M-\nu)}-1}{2a}\Big|x_n=x\right] \\ &= \max\left\{x^2\frac{e^{2a\delta(M-n)}-1}{2a},\mathbb{E}[V^1_{n+1}(x_{n+1})|x_n=x]\right\}. \end{aligned}$$

The above equation provides a (backward) recurrence relation for the computation of the single sampling cost function $V^1_n(x)$. Notice that for values of x for which the l.h.s. exceeds the r.h.s. we stop and sample, otherwise we continue to the next time instant. We can prove by induction that the optimum policy is a *time-varying threshold* one. Specifically for every time n there exists a threshold λ_n such that if $|x_n|\ge\lambda_n$ we sample, otherwise we go to the next time instant. The minimum expected distortion for this single sampling problem is:

$$\frac{e^{2aT}-1-2aT}{4a^2}-V^1_0(0).$$

For obtaining the solution to the $N+1$-sampling problem, we use the solution to the N-sampling problem. For $n=M,M-1,\ldots 0$, the minimal distortion of the $N+1$ sampling problem is obtained using the vector$\{V^N_i\}^M_0$, and choosing the optimal stopping time ν for the stopping problem below:

$$\begin{aligned} V^{N+1}_n(x) &= \sup_{n\le\nu\le M}\mathbb{E}\left[V^N_\nu(0)+x_\nu^2\frac{e^{2a\delta(M-\nu)}-1}{2a}\Big|x_n=x\right] \\ &= \max\left\{V^N_n(0)+x^2\frac{e^{2a\delta(M-n)}-1}{2a},\right. \\ &\qquad \left. V^N_{n+1}(0)+\mathbb{E}\left[V^1_{n+1}(x_{n+1})|x_n=x\right]\right\}. \end{aligned}$$

In Figs. 1 through 4 we can see the relative performance of the three sampling schemes for values of the parameter $a=-10,-5,-1,1$ (value $a=0$ corresponds to the Wiener case).

IV. CONCLUDING REMARKS

We have furnished methods to obtain good sampling policies for the finite horizon filtering problem. When the signal to be kept track of is a Wiener process, we have analytic solutions. When the signal is an Ornstein-Uhlenbeck process, we have provided computational recipes to determine the best sampling policies and their performance.

We will report elsewhere on the solution to the case when the sensor has access only to noisy observations of the signal instead of perfect observations. This leads us to some simple multi-sensor sampling and filtering problems which can be solved in the same way.

The case where the samples are not reliably transmitted but can be lost in transmission is computationally more involved. There, the relative performances of the three sampling strategies is unknown. However, in principle, the best policies and their performances can be computed using nested optimization routines like we have used in this paper.

Another set of unanswered questions involves the performance of these sampling policies when the actual objective is not filtering but control or signal detection based on the samples. It will be very useful to know the extent to which the overall performance is reduced by using sampling designs that achieve merely good filtering performance.

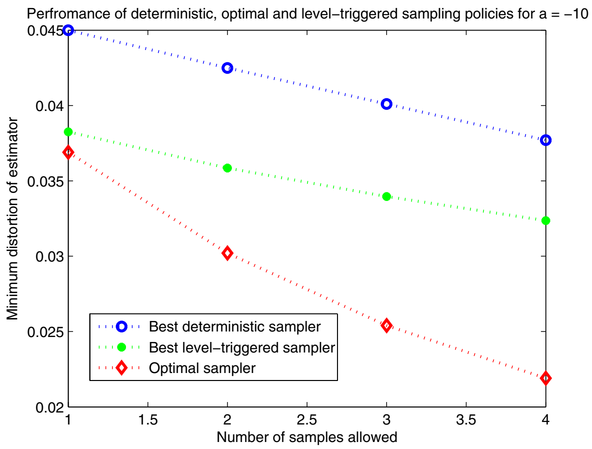

Fig. 1. Relative distortions incurred by the three samplings schemes when $a = -10$.

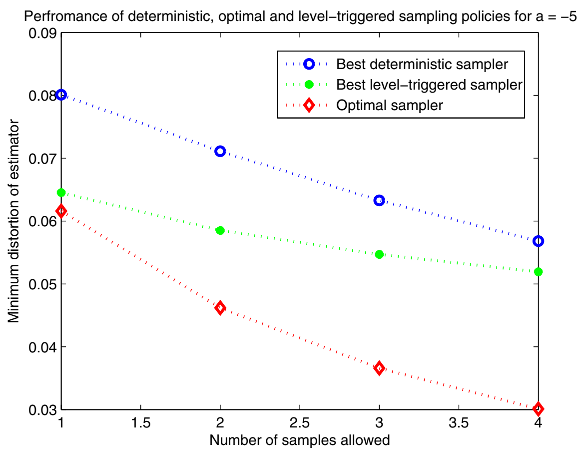

Fig. 2. Relative distortions incurred by the three samplings schemes when $a = -5$.

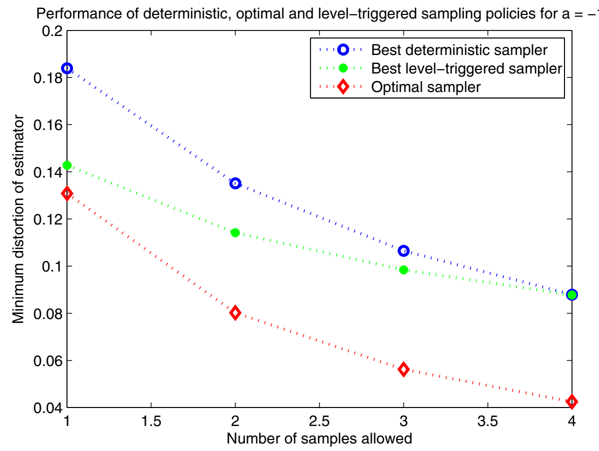

Fig. 3. Relative distortions incurred by the three samplings schemes when $a = -1$.

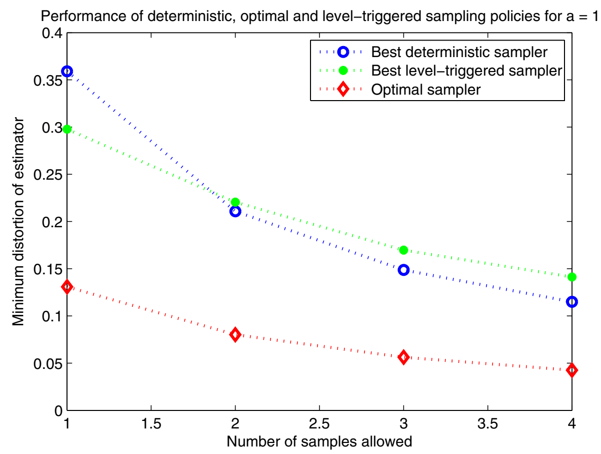

Fig. 4. Relative distortions incurred by the three samplings schemes when $a = 1$.

References

[1] Richard S. Murray, "Control in an information rich world", *IEEE Control Systems Magazine*, vol. 23, no. 2, pp. 20–33, 2003.

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

[3] Farokh Marvasti, Ed., *Nonuniform sampling*, Information Technology: Transmission, Processing and Storage. Kluwer Academic/Plenum Publishers, New York, 2001, Theory and practice, With 1 CD-ROM.

[4] Stamatis Cambanis and Elias Masry, "Sampling designs for the detection of signals in noise", *IEEE Trans. Inform. Theory*, vol. 29, no. 1, pp. 83–104, 1983.

[5] Karl Johan Åström and Bo Bernhardsson, "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)*. 2002, pp. 2011–2016, IEEE Control Systems Society.

[6] Orhan C. Imer and Tamer Basar, "Optimal estimation with limited measurements", in *Proceedings of the 44rd IEEE conference on Decision and Control and European Control Conference (Seville, Spain, 2004)*. 2005, pp. 1029–1034, IEEE Control Systems Society.

[7] Maben Rabi and John S. Baras, "Sampling of diffusion processes for real-time estimation", in *Proceedings of the 43rd IEEE conference on Decision and Control (Paradise Island Bahamas, 2004)*. 2004, pp. 4163–4168, IEEE Control Systems Society.

[8] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, vol. 113 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, second edition, 1991.

[9] A. N. Shiryaev, *Optimal stopping rules*, Springer-Verlag, 1978, translated from the Russian *Statisticheskii posledovatelnyi analiz*.