1 Age of Information and Remote Estimation

Tasmeen Zaman Ornee and Yin Sun

Department of Electrical and Computer Engineering, Auburn University, Auburn, AL

Abstract: In this chapter, we discuss the relationship between Age of Information and signal estimation error in real-time signal sampling and reconstruction. Consider a remote estimation system, where samples of a scalar Gauss-Markov signal are taken at a source node and forwarded to a remote estimator through a channel that is modeled as a queue. The estimator reconstructs an estimate of the real-time signal value from causally received samples. The optimal sampling policy for minimizing the mean square estimation error is presented, in which a new sample is taken once the *instantaneous estimation error* exceeds a predetermined threshold. When the sampler has no knowledge of current and history signal values, the optimal sampling problem reduces to a problem for minimizing a *nonlinear* Age of Information metric. In the AoI-optimal sampling policy, a new sample is taken once the *expected estimation error* exceeds a threshold. The threshold can be computed by low-complexity algorithms and the insights behind these algorithms are provided. These optimal sampling results were established (i) for general service time distributions of the queueing server, (ii) for both stable and unstable scalar Gauss-Markov signals, and (iii) for sampling problems both with and without a sampling rate constraint. [1]

1.1 Introduction

Information is usually of the greatest value when it is fresh. For example, real-time knowledge about the location, orientation, and speed of motor vehicles is imperative in autonomous driving, and access to timely information about the stock price and interest rate movements is essential for developing trading strategies on the stock market. In recent years, the Age of Information (AoI) has been adopted as a new metric for quantifying the freshness of information in real-time systems and networks. Consider a sequence of information packets that are sent to a receiver. Let U_t be the generation time of the newest packet that has been delivered to the receiver by time t. AoI, as a function of t, is defined as

$$\Delta(t) = t - U_t. \tag{1.1}$$

[1] This work was supported in part by NSF grant CCF-1813050 and ONR grant N00014-17-1-2417.

Besides the linear AoI $\Delta(t)$, nonlinear functions $p(\Delta(t))$ of the AoI has been recently demonstrated to be useful metrics for information freshness in signal estimation, control and, wireless communications (e.g., the freshness of channel state information); see recent surveys [1], [2], and [3] for more details.

In many real-time systems, the information of interest — e.g., the trajectory of UAV mobility, the measurement of temperature sensors, and the price of a stock — is represented by the value of a time-varying signal X_t, that may change slowly at some time and vary more dynamically later. Hence, the time difference described by the AoI $\Delta_t = t - U_t$ or its nonlinear functions cannot fully characterize how much the signal value has varied during the same time period, i.e., $X_t - X_{U_t}$. Hence, the status-update policy that minimizes the AoI is insufficient for minimizing the signal estimation error.

In this chapter, we discuss some recent results on real-time signal sampling and reconstructions. A problem of sampling a scalar Gauss-Markov signal is considered, where the samples forwarded to a remote estimator through a channel in a first-come, first-served (FCFS) fashion. The considered Gauss-Markov signals are Wiener process, stable and stationary Ornstein-Uhlenbeck (OU) process, and unstable and non-stationary OU process. A well-known example of Gauss-Markov process is Wiener process that is a real-valued continuous-time stochastic process and has independent increments. Another well-known example of stationary Gauss-Markov process is OU process. An OU process X_t is the continuous-time analogue of the well-known first-order autoregressive process, i.e., AR(1) process. The OU process is defined as the solution to the stochastic differential equation (SDE) [4], [5]

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \tag{1.2}$$

where μ, $\theta > 0$, and $\sigma > 0$ are parameters and W_t represents a Wiener process. It is the only nontrivial continuous-time process that is stationary, Gaussian, and Markovian [5]. From (1.2), if $\theta \to 0$ and $\sigma = 1$, the OU process becomes a Wiener process. Therefore, the Wiener process is a special case of the OU process. If the parameter of the OU process $\theta < 0$, then the process becomes unstable and non-stationary. The SDE in (1.2) still holds in such an unstable scenario. Examples of first-order systems that can be described as the Gauss-Markov process include interest rates, currency exchange rates, and commodity prices (with modifications) [6], control systems such as node mobility in the mobile ad-hoc networks, robotic swarms, and UAV systems [7], [8], and physical processes such as the transfer of liquids or gases in and out of a tank [9]. Another application is on federated learning where the progression of clients' weights is modeled as an OU process [10].

The samples experience *i.i.d.* random transmission times over the channel, which is caused by random sample size, channel fading, interference, congestions, etc. For example, UAVs flying close to WiFi access points may suffer from long communication delays and instability issues, because they receive strong interference from the WiFi access points [11]. At any time only one sample can be served by the channel. The samples that are waiting to be sent are stored in a queue at the transmitter. Hence, the channel is modeled as an FCFS queue with *i.i.d.* service times. The service time distributions that we consider are quite general: they are only required to have a finite mean. This queueing model is helpful to analyze the robustness of remote estimation systems with occasionally long transmission times.

The estimator utilizes causally received samples to construct an estimate $\hat{X}_t$ of the real-time signal value X_t. The quality of remote estimation is measured by the time-average mean-squared estimation error, i.e.,

$$\mathsf{mse} = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right]. \quad (1.3)$$

The goal is to find the optimal sampling policy that minimizes mse by causally choosing the sampling times. Our results show that if the sampler has no knowledge on the value of the Gauss-Markov signal, the optimal sampling strategy is to minimize the time-average of a nonlinear AoI function; however, by exploiting causal knowledge of the signal values, it is possible to achieve a smaller estimation error. We have also considered the sampling problem that is subject to a maximum sampling rate constraint. In practice, the cost (e.g., energy, CPU cycle, storage) for state updates increases with the average sampling rate. Hence, the optimum tradeoff between estimation error and update cost should be found.

Our main results are summarized as follows:

The optimal sampling problem for minimizing the mse under a sampling rate constraint is formulated as a constrained continuous-time Markov decision process (MDP) with an uncountable state space. Because of the curse of dimensionality, such problems are often lack of low-complexity solutions that are arbitrarily accurate. However, this MDP is solved exactly: The optimal sampling policy is proven to be a threshold policy on *instantaneous* estimation error, where the threshold is a non-linear function $v(\beta)$ of a parameter β. The value of β is equal to the summation of the optimal objective value of the MDP and the optimal Lagrangian dual variable associated with the sampling rate constraint. If there is no sampling rate constraint, the Lagrangian dual variable is zero and hence β is exactly the optimal objective value.

By comparing the optimal sampling policies of the Wiener process, stable and unstable OU processes, we find that the threshold function $v(\beta)$ changes according to the signal model, whereas the parameter β is determined in the same way for all three signal models.

Further, for a class of signal-agnostic sampling policies, the sampling times are determined without using knowledge of the observed process. The optimal signal-agnostic sampling problem is equivalent to an MDP for minimizing the time-average of a nonlinear age function $p(\Delta_t)$. The age-optimal sampling policy is a threshold policy on *expected* estimation error, where the threshold function is simply $v(\beta) = \beta$ and the parameter β is determined in the same way as above.

The above results hold for (i) general service time distributions of the queueing server, (ii) both stable and unstable scalar Gauss-Markov signals, and (ii) sampling problems both with and without a sampling rate constraint. Numerical results suggest that the optimal sampling policy is better than zero-wait sampling and classic uniform sampling.

The optimal sampling results for the Wiener process and the OU process were proven in [12], [13]. The proofs for the unstable OU process will be provided in a paper that is currently under preparation.

The rest of the chapter is organized as follows: In Section 1.2, we discuss the related work on the Age of Information and remote estimation. In Section 1.3, the model of remote estimation systems and the formulation of optimal sampling problems are presented. The solutions to signal-aware and signal-agnostic sampling problems are provided in Sections 1.3-1.4, which include both the cases with and without the sampling rate constraint. The numerical results are shown in Section 1.5 and a summary of the chapter is given in Section 1.6.

1.2 Related Work and Contributions

In this section, we will present a brief survey of the related work in the area of age of information and remote estimation.

1.2.1 Age of Information

The results presented in this chapter are significantly related to recent studies on the age of information Δ_t, e.g., [1, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. In [14], the authors provided a simple example about a status updating system, where samples of a Wiener process W_t are forwarded to a remote estimator. The age of the delivered sample is $\Delta_t = t - U_t$ if U_t is the generation time of the latest received sample. Furthermore, the MMSE estimate of W_t is $\hat{W}_t = W_{U(t)}$ and the variance of this estimator is $\mathbb{E}[(W_t - \hat{W}_t)^2] = \Delta_t$. In [15], the authors proposed a sampling policy for a discrete-time source process by incorporating mutual information as a measure for maximizing the information freshness. The results in [15] were further extended for both continuous and discrete-time source processes in [1] where the non-linear functions of the age had been used to measure data freshness. In [16], sampling and scheduling policy for the multi-source system was studied by analyzing the peak-age and peak-average-age. In [17], the authors analyzed status age when the message may take various routes in the network for queueing systems. In [18], the optimal control for information updates traveled from a source to a remote destination was studied and optimal tradeoff between the update policy and the age of information was found. The authors also showed that in many cases, the optimal policy is to wait a certain amount before sending the next update. The average age and average peak age have been analyzed for various queueing systems in, e.g., [14, 17, 19, 20]. The optimality of the

Last-Come, First-Served (LCFS) policy, or more generally the Last-Generated, First-Served (LGFS) policy, was established for various queueing system models in [23, 24, 25, 29]. Optimal sampling policies for minimizing non-linear age functions were developed in, e.g., [1, 15, 18, 34]. Age-optimal transmission scheduling of wireless networks were investigated in [21, 22, 26, 27, 28, 30, 31, 36, 37]. In [35], a game-theoretic perspective of the age was studied and the authors proposed a sampling policy by studying the timeliness of the status update where an attacker sabotages the system by jamming the channel and maximizing age-of-information. which does not have a signal model. A broad survey in the area of Age of Information is presented in [2].

1.2.2 Remote Estimation

The results in this chapter also have a tight connection with the area of remote estimation, e.g., [9, 38, 39, 40, 41, 42, 43] by adding a queue between the sampler and estimator. In [9], remote state estimation in first-order linear time-invariant (LTI) discrete-time systems was considered with a quadratic cost function and finite time horizon. They showed that a time-dependent threshold-based sampler and Kalman-like estimator are jointly optimal. In [38], the authors investigated the joint optimization of paging and registration policies in cellular networks, which is essentially the same as a joint sampling and estimation optimization problem with an indicator-type cost function and an infinite time horizon. They used majorization theory and Riesz's rearrangement inequality to show that, if the state process is modeled as a symmetric or Gaussian random walk, a threshold-based sampler and a nearest distance estimator are jointly optimal. This is the first study pointing out that the sampler and estimator have different information patterns. In [39], The authors considered a remote estimation problem with an energy-harvesting sensor and a remote estimator, where the sampling decision at the sensor is constrained by the energy level of the battery. They proved that an energy- level dependent threshold-based sampler and a Kalman-like estimator are jointly optimal. In [40], [41], optimal sampling of Wiener processes was studied, where the transmission time from the sampler to the estimator is zero. Optimal sampling of OU processes was also considered in [40], which is solved by discretizing time and using dynamic programming to solve the discrete-time optimal stopping problems. In [13], optimal sampler of OU processes is obtained analytically. In the optimal sampling policy, sampling is suspended when the server is busy and is reactivated once the server becomes idle. In addition, the threshold precisely was also characterized. The optimal sampling policy for the Wiener process in [12] is a limiting case. Remote estimation of the Wiener process with random two-way delay was considered in [44]. Remote estimation over several different channel models was recently studied in, e.g., [42, 43]. In [9, 13, 38, 39, 40, 41, 42, 43], the optimal sampling policies were proven to be threshold policies. Because of the queueing model, the optimal sampling policy in [13] has a different structure from those in [9, 38, 39, 40, 41, 42, 43]. Specifically,

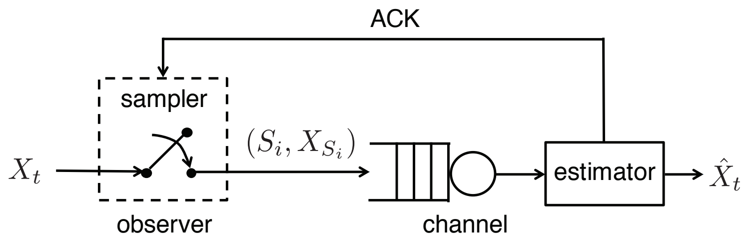

Figure 1.1 System model.

in [45], a jointly optimal sampler, quantizer, and estimator design was found for a class of continuous-time Markov processes under a bit-rate constraint. In [46], the quantization and coding schemes on the estimation performance are studied. A recent survey on remote estimation systems was presented in [47].

1.3 System Model and Problem Formulation

1.3.1 System model

Let us consider the remote estimation system illustrated in Fig. 1.1, where an observer takes samples from a Gauss-Markov process X_t and forwards the samples to an estimator through a communication channel. The channel is modeled as a single-server FCFS queue with *i.i.d.* service times. The samples experience random service times in the channel due to fading, interference, congestions, etc. The service times are *i.i.d.* and only one sample can be delivered through the channel at a time.

The system starts to operate at time $t = 0$. The i-th sample is generated at time S_i and is delivered to the estimator at time D_i with a service time Y_i, which satisfy $S_i \leq S_{i+1}$, $S_i + Y_i \leq D_i$, $D_i + Y_{i+1} \leq D_{i+1}$, and $0 < \mathbb{E}[Y_i] < \infty$ for all i. Each sample packet (S_i, X_{S_i}) contains the sampling time S_i and the sample value X_{S_i}. Let $U_t = \max\{S_i : D_i \leq t\}$ be the sampling time of the latest received sample at time t. The *age of information* or simply *age*, at time t is defined as [14], [48]

$$\Delta_t = t - U_t = t - \max\{S_i : D_i \leq t\}, \tag{1.4}$$

which is shown in Fig. 1.2. Because $D_i \leq D_{i+1}$, Δ_t can be also expressed as

$$\Delta_t = t - S_i, \text{ if } t \in [D_i, D_{i+1}), \; i = 0, 1, 2, \ldots \tag{1.5}$$

The initial state of the system is assumed to satisfy $S_0 = 0$, $D_0 = Y_0$, X_0 and Δ_0 are finite constants. The parameters μ, θ, and σ in (1.2) are known at both the sampler and estimator.

Let $I_t \in \{0, 1\}$ represent the idle/busy state of the server at time t. We assume that whenever a sample is delivered, an acknowledgement is sent back to the sampler with zero delay. By this, the idle/busy state I_t of the server is known at

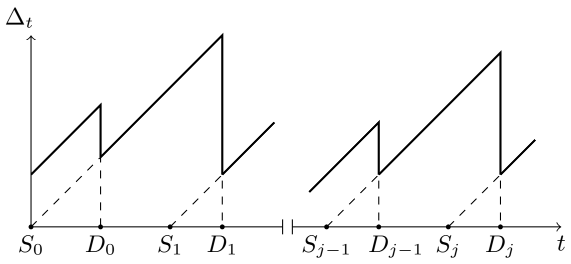

Figure 1.2 Evolution of the age Δ_t over time.

the sampler. Therefore, the information that is available at the sampler at time t can be expressed as $\{X_s, I_s : 0 \leq s \leq t\}$.

1.3.2 Sampling Policies

In causal sampling policies, each sampling time S_i is chosen by using the up-to-date information available at the sampler. S_i is a *stopping time* with respect to the filtration $\{\mathcal{N}_t^+, t \geq 0\}$ (a non-decreasing and right-continuous family of σ-fields) such that

$$\{S_i \leq t\} \in \mathcal{N}_t^+, \ \forall t \geq 0. \tag{1.6}$$

Let $\pi = (S_1, S_2, ...)$ represent a sampling policy and Π represent the set of *causal* sampling policies.

If the inter-sampling times form a regenerative process [49] and (1.6) is satisfied for each sampling policy $\pi \in \Pi$, we can obtain that S_i is finite almost surely for all i. We assume that the OU process $\{X_t, t \geq 0\}$ and the service times $\{Y_i, i = 1, 2, \dots\}$ are mutually independent, and do not change according to the sampling policy.

A sampling policy $\pi \in \Pi$ is said to be *signal-agnostic* (*signal-aware*), if π is (not necessarily) independent of $\{X_t, t \geq 0\}$. Let $\Pi_{\text{signal-agnostic}} \subset \Pi$ denote the set of signal-agnostic sampling policies, defined as

$$\Pi_{\text{signal-agnostic}} = \{\pi \in \Pi : \pi \text{ is independent of } \{X_t, t \geq 0\}\}. \tag{1.7}$$

1.3.3 MMSE Estimator

According to (1.6), S_i is a finite stopping time. By using [50, Eq. (3)] and the strong Markov property of the Gauss-Markov process [51, Eq. (4.3.27)], X_t is expressed as

$$\begin{aligned} X_t = & X_{S_i} e^{-\theta(t-S_i)} + \mu\big[1 - e^{-\theta(t-S_i)}\big] \\ & + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-S_i)} W_{e^{2\theta(t-S_i)}-1}, \text{ if } t \in [S_i, \infty). \end{aligned} \tag{1.8}$$

At any time $t \geq 0$, the estimator uses causally received samples to construct an estimate $\hat{X}_t$ of the real-time signal value X_t. The information available to the estimator consists of two parts: (i) $M_t = \{(S_i, X_{S_i}, D_i) : D_i \leq t\}$, which contains the sampling time S_i, sample value X_{S_i}, and delivery time D_i of the samples that have been delivered by time t and (ii) the fact that no sample has been received after the last delivery time $\max\{D_i : D_i \leq t\}$. Similar to [40, 52, 12], we assume that the estimator neglects the second part of information. This assumption can be removed by considering a joint sampler and estimator design problem. Specifically, it was shown in [9, 38, 39, 42, 43] that when the sampler and estimator are jointly optimized in discrete-time systems, the optimal estimator has the same expression no matter with or without the second part of the information. As pointed out in [38, p. 619], such a structure property of the MMSE estimator can be also established for continuous-time systems. Our goal is to find the closed-form expression of the optimal sampler under this assumption. The remaining task of finding the jointly optimal sampler and estimator design can be done by further using the majorization techniques developed in [9, 38, 39, 42, 43]; see [45] for a recent treatment on this task. Then the minimum mean square error (MMSE) estimator is determined by

$$\begin{aligned}\hat{X}_t = \mathbb{E}[X_t|M_t] =& X_{S_i} e^{-\theta(t-S_i)} + \mu\big[1 - e^{-\theta(t-S_i)}\big], \\ & \text{if } t \in [D_i, D_{i+1}),\ i = 0, 1, 2, \ldots \end{aligned} \tag{1.9}$$

1.3.4 Problem Formulation

We study the optimal sampling policy that minimizes the mean-squared estimation error subject to an average sampling-rate constraint, which is formulated as the following problem:

$$\mathsf{mse}_{\text{opt}} = \inf_{\pi \in \Pi} \ \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right] \tag{1.10}$$

$$\text{s.t. } \liminf_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{n} (S_{i+1} - S_i)\right] \geq \frac{1}{f_{\max}}, \tag{1.11}$$

where $\mathsf{mse}_{\text{opt}}$ is the optimum value of (1.10) and $f_{\max}$ is the maximum allowed sampling rate. When $f_{\max} = \infty$, this problem becomes an unconstrained problem.

Problem (1.10) is a constrained continuous-time MDP with a continuous state space. However, we found an exact solution to this problem.

1.4 Optimal Signal Sampling Policies

The optimal sampling policies for Wiener process, OU process, and unstable OU process are presented in the following theorems, respectively.

1.4.1 Signal-Aware Sampling Without A Sampling Rate Constraint

We first consider the unconstrained optimal sampling problem, i.e., $f_{\max} = \infty$, such that the sampling rate constraint (1.11) can be removed.

1.4.1.1 Wiener Process

If the signal being sampled, $X_t = W_t$ is a Wiener process, then the optimal sampling policy is provided in the following theorem.

THEOREM 1.1 *If X_t is a Wiener process with $f_{\max} = \infty$, and the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \ldots)$ with a parameter β is an optimal solution to (1.10), where*

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \left|X_t - \hat{X}_t\right| \geq v(\beta)\right\}, \tag{1.12}$$

$D_i(\beta) = S_i(\beta) + Y_i$, $v(\beta)$ is defined by

$$v(\beta) = \sqrt{3(\beta - \mathbb{E}[Y_i])}. \tag{1.13}$$

and β is the unique root of

$$\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right] - \beta\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 0, \tag{1.14}$$

The optimal objective value to (1.10) is given by

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.15}$$

Furthermore, β is exactly the optimal value to (1.10), i.e., $\beta = \mathsf{mse}_{\text{opt}}$.

The optimal sampling policy in Theorem 1.1 has a nice structure. Specifically, the $(i+1)$-th sample is taken at the earliest time t satisfying two conditions: (i) The i-th sample has already been delivered by time t, i.e., $t \geq D_i(\beta)$, and (ii) the estimation error $|X_t - \hat{X}_t|$ is no smaller than a pre-determined threshold $v(\beta)$, where $v(\cdot)$ is a non-linear function defined in (1.13).

1.4.1.2 Ornstein-Uhlenbeck Process (Stable Case, $\theta > 0$)

To present the optimal sampler when X_t is an OU process, we need to introduce another OU process O_t with the initial state $O_t = 0$ and parameter $\mu = 0$. According to (1.8), O_t can be expressed as

$$O_t = \frac{\sigma}{\sqrt{2\theta}} e^{-\theta t} W_{e^{2\theta t} - 1}. \tag{1.16}$$

Define

$$\mathsf{mse}_{Y_i} = \mathbb{E}[O_{Y_i}^2] = \frac{\sigma^2}{2\theta}\mathbb{E}[1 - e^{-2\theta Y_i}], \tag{1.17}$$

$$\mathsf{mse}_{\infty} = \mathbb{E}[O_{\infty}^2] = \frac{\sigma^2}{2\theta}. \tag{1.18}$$

In [13], it is shown that mse_{Y_i} and mse_∞ are the lower and upper bounds of $\mathsf{mse}_{\text{opt}}$, respectively. We will also need to use the following function

$$G(x) = \frac{e^{x^2}}{x}\int_0^x e^{-t^2}dt = \frac{e^{x^2}}{x}\frac{\sqrt{\pi}}{2}\operatorname{erf}(x),\ x \in [0,\infty), \tag{1.19}$$

where if $x = 0$, $G(x)$ is defined as its right limit $G(0) = \lim_{x\to 0^+} G(x) = 1$, and $\operatorname{erf}(\cdot)$ is the error function [53], defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt. \tag{1.20}$$

In this scenario, the optimal sampler is provided in the following theorem.

THEOREM 1.2 *If X_t is a stable and stationary OU process with $f_{\max} = \infty$ and the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \ldots)$ with a parameter β is an optimal solution to* (1.10)*, where*

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \left|X_t - \hat{X}_t\right| \geq v(\beta)\right\}, \tag{1.21}$$

$D_i(\beta) = S_i(\beta) + Y_i$, $v(\beta)$ is defined by

$$v(\beta) = \frac{\sigma}{\sqrt{\theta}}G^{-1}\left(\frac{\mathsf{mse}_\infty - \mathsf{mse}_{Y_i}}{\mathsf{mse}_\infty - \beta}\right), \tag{1.22}$$

$G^{-1}(\cdot)$ is the inverse function of $G(\cdot)$ in (1.19) *and β is the unique root of*

$$\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t - \hat{X}_t)^2dt\right] - \beta\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 0. \tag{1.23}$$

The optimal objective value to (1.10) *is given by*

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t - \hat{X}_t)^2dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.24}$$

Similar to Theorem 1.1, $v(\cdot)$ in (1.22) is also a non-linear function. In [13], it is shown that $\mathsf{mse}_{Y_i} \leq \beta < \mathsf{mse}_\infty$. Further, it is not hard to show that $G(x)$ is strictly increasing on $[0,\infty)$ and $G(0) = 1$. Hence, its inverse function $G^{-1}(\cdot)$ and the threshold $v(\beta)$ are properly defined and $v(\beta) \geq 0$.

1.4.1.3 Ornstein-Uhlenbeck Process (Unstable Case, $\theta < 0$)

Let $\rho = -\theta$, then an unstable OU process with initial state O_0, parameters $\mu = 0$, $\sigma > 0$, and $\theta < 0$ can be expressed as

$$O_t = \frac{\sigma}{\sqrt{2\rho}}e^{\rho t}W_{1-e^{-2\rho t}}. \tag{1.25}$$

Define

$$\mathsf{umse}_{\mathsf{Y_i}} = \mathbb{E}[O_{Y_i}^2] = \frac{\sigma^2}{2\rho}\mathbb{E}[e^{2\rho Y_i} - 1], \tag{1.26}$$

$$\mathsf{umse}_\infty = \mathbb{E}[O_\infty^2] \to \infty. \tag{1.27}$$

where umse_{Y_i} and umse_∞ are the lower and upper bounds of $\mathsf{mse}_{\text{opt}}$, respectively. We will also need to use the following function

$$K(x) = \frac{e^{-x^2}}{x}\int_0^x e^{t^2}dt = \frac{e^{-x^2}}{x}\frac{\sqrt{\pi}}{2}\,\text{erfi}(x),\ \ x \in [0,\infty), \tag{1.28}$$

where if $x = 0$, $K(x)$ is defined as its right limit $K(0) = \lim_{x\to 0^+} K(x) = 1$, and $\text{erfi}(\cdot)$ is the error function [53], defined as

$$\text{erfi}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{t^2}dt. \tag{1.29}$$

Note that $K(x)$ is a strictly decreasing function on $x \in [0,\infty)$ and its inverse $K^{-1}(\cdot)$ is properly defined.

The optimal sampler is then provided in the following theorem.

THEOREM 1.3 *If X_t is an unstable and non-stationary OU process with $f_{\max} = \infty$ and the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \ldots)$ with a parameter β is an optimal solution to* (1.10)*, where*

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \left|X_t - \hat{X}_t\right| \geq v(\beta)\right\}, \tag{1.30}$$

$D_i(\beta) = S_i(\beta) + Y_i$, $v(\beta)$ is defined by

$$v(\beta) = \frac{\sigma}{\sqrt{\rho}}K^{-1}\left(\frac{\mathsf{umse}_{Y_i} + \frac{\sigma^2}{2\rho}}{\frac{\sigma^2}{2\rho} + \beta}\right), \tag{1.31}$$

$K^{-1}(\cdot)$ is the inverse function of $K(\cdot)$ in (1.28) *and β is the unique root of*

$$\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t - \hat{X}_t)^2 dt\right] - \beta\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 0. \tag{1.32}$$

The optimal objective value to (1.10) *is given by*

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.33}$$

The results for unstable OU process will be provided in a paper that is currently under preparation.

1.4.1.4 Low-complexity Algorithms for Computing β

We now present three algorithms, e.g., bisection search, Nweton's method, and fixed-point iterations algorithms for computing the root of (1.14), (1.23), and (1.32). Because the $S_i(\beta)$'s are stopping times, numerically calculating the expectations in (1.14), (1.23), and (1.32) appears to be a difficult task. Nonetheless, this challenge can be solved by resorting to the following lemma, which is obtained by using Dynkin's formula [54, Theorem 7.4.1] and the optional stopping theorem.

LEMMA 1.4 *In Theorems 1.2 and 1.3, it holds that*

$$\begin{aligned}&\mathbb{E}[D_{i+1}(\beta)-D_i(\beta)]\\ =&\mathbb{E}[\max\{R_1(v(\beta))-R_1(O_{Y_i}),0\}]+\mathbb{E}[Y_i],\end{aligned} \tag{1.34}$$

$$\begin{aligned}&\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t-\hat{X}_t)^2dt\right]\\ =&\mathbb{E}[\max\{R_2(v(\beta))-R_2(O_{Y_i}),0\}]\\ &+\mathsf{mse}_\infty[\mathbb{E}(Y_i)-\gamma]+\mathbb{E}\left[\max\{v^2(\beta),O_{Y_i}^2\}\right]\gamma,\end{aligned} \tag{1.35}$$

where

$$\gamma=\frac{1}{2\theta}\mathbb{E}[1-e^{-2\theta Y_i}], \tag{1.36}$$

$$R_1(v)=\frac{v^2}{\sigma^2}\,{}_2F_2\left(1,1;\frac{3}{2},2;\frac{\theta}{\sigma^2}v^2\right), \tag{1.37}$$

$$R_2(v)=-\frac{v^2}{2\theta}+\frac{v^2}{2\theta}\,{}_2F_2\left(1,1;\frac{3}{2},2;\frac{\theta}{\sigma^2}v^2\right). \tag{1.38}$$

In (1.37) and (1.38), we have used the generalized hypergeometric function, which is defined by [55, Eq. 16.2.1]

$$\begin{aligned}&{}_pF_q(a_1,a_2,\cdots,a_p;b_1,b_2,\cdots b_q;z)\\ =&\sum_{n=0}^{\infty}\frac{(a_1)_n(a_2)_n\cdots(a_p)_n}{(b_1)_n(b_2)_n\cdots(b_p)_n}\frac{z^n}{n!},\end{aligned} \tag{1.39}$$

where

$$(a)_0=1, \tag{1.40}$$

$$(a)_n=a(a+1)(a+2)\cdots(a-n+1),\ n\geq 1. \tag{1.41}$$

Using Lemma 1.4, the expectations in (1.14), (1.23), and (1.32) can be evaluated by Monte Carlo simulations of one-dimensional random variables O_{Y_i} and Y_i, which is much simpler than directly simulating the entire random process $\{O_t,t\geq 0\}$.

For notational simplicity, we rewrite (1.14), (1.23), and (1.32) as

$$f(\beta)=f_1(\beta)-\beta f_2(\beta)=0, \tag{1.42}$$

where $f_1(\beta)=\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t-\hat{X}_t)^2dt\right]$ and $f_2(\beta)=\mathbb{E}[D_{i+1}(\beta)-D_i(\beta)]$. The function $f(\beta)$ has several nice properties, which are asserted in the following lemma.

LEMMA 1.5 *The function $f(\beta)$ has the following properties:*
(i) $f(\beta)$ is concave, continuous, and strictly decreasing on β,
(ii) $f(\mathsf{mse}_{Y_i})>0$ and $\lim_{\beta\to\mathsf{mse}_\infty^-}f(\beta)=-\infty$.

The uniqueness of β follows immediately from Lemma 1.5.

Now we are ready to present the associated algorithms for solving β.

Algorithm 1 Bisection search method for finding β

---

given $l = \mathsf{mse}_{Y_i}$, $u = \mathsf{mse}_{\infty}$, tolerance $\epsilon > 0$.
repeat
  $\beta := (l + u)/2$.
  $o := f_1(\beta) - \beta f_2(\beta)$.
  if $o \geq 0$, $l := \beta$; **else**, $u := \beta$.
until $u - l \leq \epsilon$.
return β.

Algorithm 2 Newton's method for finding β

---

given tolerance $\epsilon > 0$.
Pick initial value $\beta_0 \in [\mathsf{mse}_{\text{opt}}, \mathsf{mse}_{\infty})$.
repeat
  $\beta_{k+1} := \beta_k - \frac{f(\beta_k)}{f'(\beta_k)}$.
until $|\frac{f(\beta_k)}{f'(\beta_k)}| \leq \epsilon$.
return β_{k+1}.

Bisection Search

Because $f(\beta)$ is decreasing and has a unique root, one can use a bisection search method to solve (1.14), (1.23), and (1.32), which is illustrated in Algorithm 1. The bisection search method has a global linear convergence speed.

Newton's method

To achieve an even faster convergence speed, we can use Newton's method [56]

$$\beta_{k+1} = \beta_k - \frac{f(\beta_k)}{f'(\beta_k)} \tag{1.43}$$

to solve (1.23), as shown in Algorithm 2. We suggest choosing the initial value β_0 of Newton's method from the set $[\mathsf{mse}_{\text{opt}}, \mathsf{mse}_{\infty})$, i.e., β_0 is larger than the root $\mathsf{mse}_{\text{opt}}$. Such an initial value β_0 can be found by taking a few bisection search iterations. Because $f(\beta)$ is a concave function, the choice of the initial value $\beta_0 \in [\mathsf{mse}_{\text{opt}}, \mathsf{mse}_{\infty})$ ensures that β_k is a decreasing sequence converging to $\mathsf{mse}_{\text{opt}}$ [57]. Since Newton's method is a fixed-point iterative algorithm, it has a global linear convergence speed. In addition, Newton's method is known to have a quadratic convergence speed in the neighborhood of the root $\mathsf{mse}_{\text{opt}}$ [56, Chapter 2].

Fixed-point iterations

Newton's method requires to compute the gradient $f'(\beta_k)$, which can be solved by a finite-difference approximation, as in the secant method [56]. In the sequel, we introduce another approximation approach of Newton's method, which is of

Algorithm 3 Fixed-point iterations for finding β

given tolerance $\epsilon > 0$.
Pick initial value $\beta_0 \in [\mathsf{mse}_{\text{opt}}, \mathsf{mse}_\infty)$.
repeat
$\beta_{k+1} := \frac{f_1(\beta_k)}{f_2(\beta_k)}$.
until $|\beta_{k+1} - \frac{f_1(\beta_k)}{f_2(\beta_k)}| \le \epsilon$.
return β_{k+1}.

independent interest. In Theorem 1.2, we have shown that $\mathsf{mse}_{\text{opt}}$ is the optimal solution for $f_1(\beta)/f_2(\beta)$. Hence, the gradient of $f_1(\beta)/f_2(\beta)$ is equal to zero at the optimal solution $\beta = \mathsf{mse}_{\text{opt}}$, which leads to

$$f_1'(\mathsf{mse}_{\text{opt}})f_2(\mathsf{mse}_{\text{opt}}) - f_1(\mathsf{mse}_{\text{opt}})f_2'(\mathsf{mse}_{\text{opt}}) = 0. \tag{1.44}$$

Therefore,

$$\mathsf{mse}_{\text{opt}} = \frac{f_1(\mathsf{mse}_{\text{opt}})}{f_2(\mathsf{mse}_{\text{opt}})} = \frac{f_1'(\mathsf{mse}_{\text{opt}})}{f_2'(\mathsf{mse}_{\text{opt}})}. \tag{1.45}$$

Because $f_1(\beta)$ and $f_2(\beta)$ are smooth functions, when β_k is in the neighborhood of $\mathsf{mse}_{\text{opt}}$, (1.45) implies that $f_1'(\beta_k) - \beta_k f_2'(\beta_k) \approx f_1'(\mathsf{mse}_{\text{opt}}) - \mathsf{mse}_{\text{opt}} f_2'(\mathsf{mse}_{\text{opt}}) = 0$. Substituting this into (1.43), yields

$$\begin{aligned}\beta_{k+1} =& \beta_k - \frac{f_1(\beta_k) - \beta_k f_2(\beta_k)}{f_1'(\beta_k) - f_2(\beta_k) - \beta_k f_2'(\beta_k)} \\ \approx& \beta_k - \frac{f_1(\beta_k) - \beta_k f_2(\beta_k)}{-f_2(\beta_k)} \\ =& \frac{f_1(\beta_k)}{f_2(\beta_k)},\end{aligned} \tag{1.46}$$

which is a fixed-point iterative algorithm that was recently proposed in [58]. Similar to Newton's method, the fixed-point updates in (1.46) converge to $\mathsf{mse}_{\text{opt}}$ if the initial value $\beta_0 \in [\mathsf{mse}_{\text{opt}}, \mathsf{mse}_\infty)$. Moreover, (1.46) has a global linear convergence speed and a local quadratic convergence speed. See [58] for a proof of these results.

A comparison of these three algorithms are shown in [13]. One can observe that the fixed-point updates and Newton's method converge faster than bisection search.

More insights

We note that although (1.14), (1.23), and (1.32), and equivalently (1.42), has a unique root $\mathsf{mse}_{\text{opt}}$, the fixed-point equation

$$h(\beta) = \frac{f_1(\beta)}{f_2(\beta)} - \beta = \frac{f_1(\beta) - \beta f_2(\beta)}{f_2(\beta)} = 0 \tag{1.47}$$

has two roots $\mathsf{mse}_{\text{opt}}$ and mse_{∞} if the signal process X_t is stable and stationary OU process. This is because the denominator $f_2(\beta)$ in (1.47) increases to ∞ as $\beta \rightarrow \mathsf{mse}_{\infty}$, which creates an extra root at $\beta = \mathsf{mse}_{\infty}$. See [13, Fig. 6] for an illustration of the two roots of $h(\beta)$. It is showed in [13] that the correct root for computing the optimal threshold is $\mathsf{mse}_{\text{opt}}$. Interestingly, Algorithms 1-2 converge to the desired root $\mathsf{mse}_{\text{opt}}$, instead of mse_{∞}. Finally, we remark that these three algorithms can be used to find the optimal threshold in the age-optimal sampling problem studied in, e.g., [1, 15].

1.4.2 Signal-aware Sampling With A Sampling Rate Constraint

When the sampling rate constraint (1.11) is taken into consideration, a solution to (1.10) for the three processes discussed above can be expressed in the following theorems:

1.4.2.1 Wiener Process

THEOREM 1.6 *If X_t is a Wiener process with Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then* (1.12)-(1.14) *is an optimal solution to* (1.10). *The value of $\beta \geq 0$ is determined in two cases: β is the unique root of* (1.14) *if the root of* (1.14) *satisfies*

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] > 1/f_{\max}; \tag{1.48}$$

otherwise, β is the unique root of

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 1/f_{\max}. \tag{1.49}$$

The optimal objective value to (1.10) *is given by*

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.50}$$

One can see that Theorem 1.1 is a special case of Theorem 1.6 when $f_{\max} = \infty$.

1.4.2.2 Ornstein-Uhlenbeck Process (Stable Case, $\theta > 0$)

THEOREM 1.7 *If X_t is a stable and stationary OU process with the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then* (1.21)-(1.23) *is an optimal solution to* (1.10). *The value of $\beta \geq 0$ is determined in two cases: β is the unique root of* (1.23) *if the root of* (1.23) *satisfies*

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] > 1/f_{\max}; \tag{1.51}$$

otherwise, β is the unique root of

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 1/f_{\max}. \tag{1.52}$$

The optimal objective value to (1.10) *is given by*

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.53}$$

Theorem 1.2 is also a special case of Theorem 1.7 when $f_{\max} = \infty$.

1.4.2.3 Ornstein-Uhlenbeck Process (Unstable Case, $\theta < 0$)

THEOREM 1.8 *If X_t is a stable and stationary OU process with the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then (1.30)-(1.32) is an optimal solution to (1.10). The value of $\beta \geq 0$ is determined in two cases: β is the unique root of (1.32) if the root of (1.32) satisfies*

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] > 1/f_{\max}; \tag{1.54}$$

otherwise, β is the unique root of

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 1/f_{\max}. \tag{1.55}$$

The optimal objective value to (1.10) is given by

$$\mathsf{mse}_{\text{opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.56}$$

Theorem 1.3 is also a special case of Theorem 1.8 when $f_{\max} = \infty$.

In Theorems 1.6, 1.7, and 1.8 the calculation of β falls into two cases: In one case, β can be computed via Algorithms 1-3.

For this case to occur, the sampling rate constraint (1.11) needs to be inactive at the root of (1.14), (1.23), and (1.32). Because $D_i(\beta) = S_i(\beta) + Y_i$, we can obtain $\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = \mathbb{E}[S_{i+1}(\beta) - S_i(\beta)]$ and hence (1.48), (1.51), and (1.54) holds when the sampling rate constraint (1.11) is inactive.

In the other case, β is selected to satisfy the sampling rate constraint (1.11) with equality, as required in (1.49), (1.52), and (1.55). Before we solve (1.49), (1.52), and (1.55), let us first use $f_2(\beta)$ to express (1.49), (1.52), and (1.55) as

$$g(\beta) = \frac{1}{f_{\max}} - f_2(\beta) = 0. \tag{1.57}$$

LEMMA 1.9 *The function $g(\beta)$ has the following properties:*
(i) $g(\beta)$ is continuous and strictly decreasing on β,
(ii) $g(\mathsf{mse}_{Y_i}) \geq 0$ and $\lim_{\beta \to \mathsf{mse}_\infty^-} g(\beta) = -\infty$.

According to Lemma 1.9, (1.49), (1.52), and (1.55) has a unique root in $[\mathsf{mse}_{Y_i}, \mathsf{mse}_\infty)$, which is denoted as β^*. In addition, the numerical results also suggest that $g(\beta)$ should be concave [13].

The root β^* can be solved by using bisection search and Newton's method, which are explained in Algorithms 4-5, respectively. Similar to the discussions in Section 1.4.1.4, the convergence of Algorithm 4 is ensured by Lemma 1.9. Moreover, if $g(\beta)$ is concave and $\beta_0 \in [\beta^*, \mathsf{mse}_\infty)$, β_k in Algorithm 5 is a decreasing sequence converging to the root β^* of (1.49), (1.52), and (1.55) [57].

Algorithm 4 Bisection search method for finding β with rate constraint

given $l = \mathsf{mse}_{Y_i}$, $u = \mathsf{mse}_\infty$, tolerance $\epsilon > 0$.
repeat
  $\beta := (l+u)/2$.
  $o := \mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]$.
  if $o \geq 1/f_{\max}$, $u := \beta$; **else**, $l := \beta$.
until $u - l \leq \epsilon$.
return β.

Algorithm 5 Newton's method for finding β with rate constraint

given tolerance $\epsilon > 0$.
Pick initial value $\beta_0 \in [\beta^*, \mathsf{mse}_\infty)$.
repeat
  $\beta_{k+1} := \beta_k - \frac{g(\beta_k)}{g'(\beta_k)}$.
until $|\frac{g(\beta_k)}{g'(\beta_k)}| \leq \epsilon$.
return β_{k+1}.

1.4.3 Signal-Agnostic Sampling

In signal-agnostic sampling policies, the sampling times S_i are determined based only on the service times Y_i, but not on the observed process $\{X_t, t \geq 0\}$, and the following lemma can be introduced.

LEMMA 1.10 *If $\pi \in \Pi_{signal\text{-}agnostic}$, then the mean-squared estimation error of the Wiener process X_t at time t is*

$$p(\Delta_t) = \mathbb{E}\left[(X_t - \hat{X}_t)^2 \Big| \pi, Y_1, Y_2, \ldots\right] = \Delta_t, \tag{1.58}$$

where Δ_t is the age of information at time t. On the other hand, if X_t is an OU process, then the mean-squared estimation error at time t is

$$p(\Delta_t) = \mathbb{E}\left[(X_t - \hat{X}_t)^2 \Big| \pi, Y_1, Y_2, \ldots\right] = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\Delta_t}\right), \tag{1.59}$$

which a strictly increasing function of the age Δ_t.

According to Lemma 1.10 and Fubini's theorem, for every policy $\pi \in \Pi_{\text{signal-agnostic}}$,

$$\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right] = \mathbb{E}\left[\int_0^T p(\Delta_t) dt\right]. \tag{1.60}$$

Hence, minimizing the mean-squared estimation error among signal-agnostic sampling policies can be formulated as the following MDP for minimizing the expected time-average of the nonlinear age function $p(\Delta_t)$:

$$\mathsf{mse}_{\text{age-opt}} = \inf_{\pi \in \Pi_{\text{signal-agnostic}}} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T p(\Delta_t)dt\right] \tag{1.61}$$

$$\text{s.t.} \quad \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(S_{i+1}-S_i)\right] \geq \frac{1}{f_{\max}}, \tag{1.62}$$

where $\mathsf{mse}_{\text{age-opt}}$ is the optimal value of (1.61). By (1.59), $p(\Delta_t)$ and $\mathsf{mse}_{\text{age-opt}}$ are bounded. Because $\Pi_{\text{signal-agnostic}} \subset \Pi$, it follows immediately that $\mathsf{mse}_{\text{opt}} \leq \mathsf{mse}_{\text{age-opt}}$.

Problem (1.61) is one instance of the problems recently solved in Corollary 3 of [1] for general strictly increasing functions $p(\cdot)$.

1.4.3.1 Signal-Agnostic Sampling Without A Sampling Rate Constraint

From this, a solution to (1.61) for signal-agnostic sampling without rate constraint is given in the following theorem.

THEOREM 1.11 *If X_t is a Gauss-Markov process with $f_{\max} = \infty$ and the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \ldots)$ with a parameter β is an optimal solution to* (1.61)*, where*

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \mathbb{E}[(X_{t+Y_{i+1}} - \hat{X}_{t+Y_{i+1}})^2] \geq \beta\right\}, \tag{1.63}$$

$D_i(\beta) = S_i(\beta) + Y_i$ and the optimal threshold β is the root of

$$\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right] - \beta\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 0, \tag{1.64}$$

The optimal objective value to (1.61) *is given by*

$$\mathsf{mse}_{\text{age-opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}(X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.65}$$

1.4.3.2 Signal-Agnostic Sampling With A Sampling Rate Constraint

On the other hand, a solution to (1.61) when (1.62) is taken into consideration is given in the following theorem.

THEOREM 1.12 *If X_t is a Gauss-Markov process with the Y_i's are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then* (1.63)-(1.64) *is an optimal solution to* (1.61)*. The value of threshold $\beta \geq 0$ is determined in two cases: β is the root of* (1.64)*, if the root of* (1.64) *satisfies*

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] > \frac{1}{f_{\max}}; \tag{1.66}$$

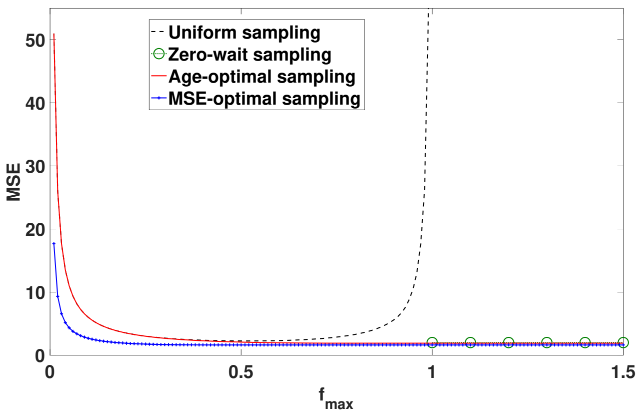

Figure 1.3 MSE vs $f_{\max}$ tradeoff for *i.i.d.* exponential service time for Wiener process with $\mathbb{E}[Y_i] = 1$.

otherwise, β is the root of

$$\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = \frac{1}{f_{\max}}. \tag{1.67}$$

The optimal objective value to (1.61) *is given by*

$$\mathsf{mse}_{\text{age-opt}} = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{1.68}$$

Theorem 1.12 follows from Corollary 3 of [1] and Lemma 1.10. Similar to the case of signal-aware sampling, the roots of (1.64) and (1.67) can be solved by using Algorithms 1-5. In fact, Algorithms 1-5 can be used for minimizing general monotonic age penalty functions. [1].

1.5 Numerical Results

In this section, we evaluate the estimation error achieved by the following four sampling policies: 1. Uniform sampling, 2. Zero-wait sampling [18, 14], 3. Age-optimal sampling [1], and 4. MSE-optimal sampling.

Let $\mathsf{mse}_{\text{uniform}}$, $\mathsf{mse}_{\text{zero-wait}}$, $\mathsf{mse}_{\text{age-opt}}$, and $\mathsf{mse}_{\text{opt}}$, be the MSEs of uniform sampling, zero-wait sampling, age-optimal sampling, MSE-optimal sampling, respectively. We can obtain

$$\begin{aligned} &\mathsf{mse}_{Y_i} \leq \mathsf{mse}_{\text{opt}} \leq \mathsf{mse}_{\text{age-opt}} \leq \mathsf{mse}_{\text{uniform}} \leq \mathsf{mse}_{\infty}, \\ &\mathsf{mse}_{\text{age-opt}} \leq \mathsf{mse}_{\text{zero-wait}} \leq \mathsf{mse}_{\infty}, \end{aligned} \tag{1.69}$$

whenever zero-wait sampling is feasible, which fit with our numerical results. The expectations in (1.37) and (1.38) are evaluated by taking the average over

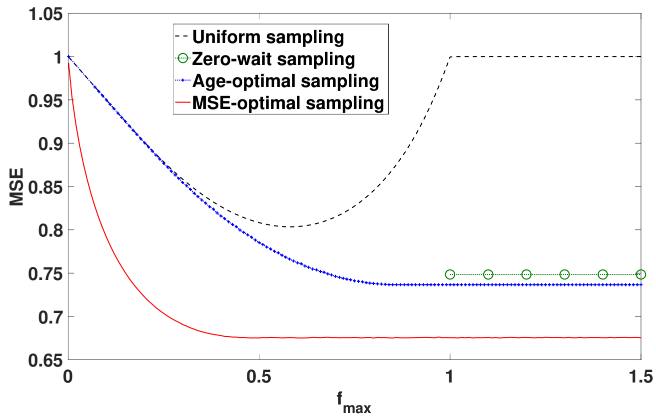

Figure 1.4 MSE vs $f_{\max}$ tradeoff for *i.i.d.* exponential service time with $\mathbb{E}[Y_i] = 1$, where the parameters of the OU process are $\sigma = 1$ and $\theta = 0.5$.

1 million samples. For OU process, the parameters are given by $\sigma = 1$, $\theta = 0.5$, and μ can be chosen arbitrarily because it does not affect the estimation error.

Figures 1.3 and 1.4 illustrate the tradeoff between the MSE and $f_{\max}$ for *i.i.d.* exponential service times with mean $\mathbb{E}[Y_i] = 1$. Because $\mathbb{E}[Y_i] = 1$, the maximum throughput of the queue is 1. For Figure 1.4, the lower bound mse_{Y_i} is 0.5 and the upper bound mse_{∞} is 1. In fact, as Y_i is an exponential random variable with mean 1, $\frac{\sigma^2}{2\theta}(1 - e^{-2\theta Y_i})$ has a uniform distribution on $[0, 1]$. Hence, $\mathsf{mse}_{Y_i} = 0.5$. For small values of $f_{\max}$, age-optimal sampling is similar to uniform sampling, and hence $\mathsf{mse}_{\text{age-opt}}$ and $\mathsf{mse}_{\text{uniform}}$ are close to each other in the regime. However, as $f_{\max}$ grows, $\mathsf{mse}_{\text{uniform}}$ reaches the upper bound mse_{∞} and remains constant for $f_{\max} \geq 1$. This is because the queue length of uniform sampling is large at high sampling frequencies. In particular, when $f_{\max} \geq 1$, the queue length of uniform sampling is infinite. On the other hand, $\mathsf{mse}_{\text{age-opt}}$ and $\mathsf{mse}_{\text{opt}}$ decrease with respect to $f_{\max}$. The reason behind this is that the set of feasible sampling policies satisfying the constraint in (1.10) and (1.61) becomes larger as $f_{\max}$ grows, and hence the optimal values of (1.10) and (1.61) are decreasing in $f_{\max}$. As we expected, $\mathsf{mse}_{\text{zero-wait}}$ is larger than $\mathsf{mse}_{\text{opt}}$ and $\mathsf{mse}_{\text{age-opt}}$. Moreover, all of them are between the lower bound mse_{Y_i} and upper bound mse_{∞}.

Figures 1.5 and 1.6 depict the MSE of *i.i.d.* normalized log-normal service time for $f_{\max} = 0.8$. Figures 1.7 and 1.8 depict the MSE of *i.i.d.* normalized log-normal service time for $f_{\max} = 1.2$, where $Y_i = e^{\alpha X_i}/\mathbb{E}[e^{\alpha X_i}], \alpha > 0$ is the scale parameter of log-normal distribution, and $(X_1, X_2, \dots)$ are *i.i.d.* Gaussian random variables with zero mean and unit variance. Because $\mathbb{E}[Y_i] = 1$, the maximum throughput of the queue is 1. In Figures 1.5 and 1.6, since $f_{\max} < 1$, zero-wait sampling is not feasible and hence is not plotted. As the scale parameter α grows, the tail of the log-normal distribution becomes heavier.

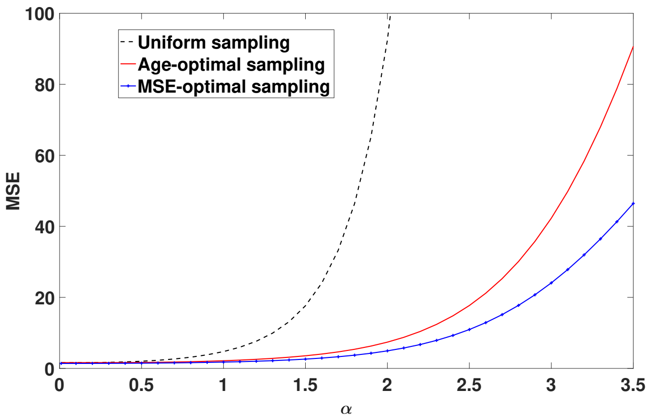

Figure 1.5 MSE vs. the scale parameter α of *i.i.d.* normalized log-normal service time distribution for Wiener process with $\mathbb{E}[Y_i] = 1$ and $f_{\max} = 0.8$. Zero-wait sampling is not feasible here as $f_{\max} < 1/\mathbb{E}[Y_i]$ and hence is not plotted.

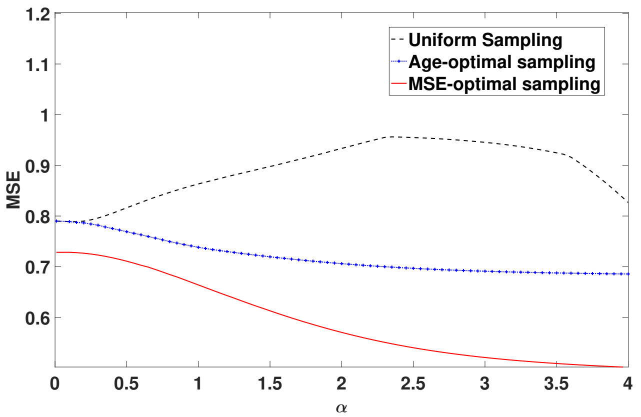

Figure 1.6 MSE vs. the scale parameter α of *i.i.d.* normalized log-normal service time distribution with $\mathbb{E}[Y_i] = 1$ and $f_{\max} = 0.8$, where the parameters of the OU process are $\sigma = 1$ and $\theta = 0.5$. Zero-wait sampling is not feasible here as $f_{\max} < 1/\mathbb{E}[Y_i]$ and hence is not plotted.

In Figures 1.5 and 1.7, $\mathsf{mse}_{\text{age-opt}}$ and $\mathsf{mse}_{\text{opt}}$ grow quickly in α. But in Figures 1.6 and 1.8, $\mathsf{mse}_{\text{age-opt}}$ and $\mathsf{mse}_{\text{opt}}$ drop with α. This phenomenon may look surprising at first sight. To understand this phenomenon, let us consider the age penalty function $p(\Delta_t)$ in (1.59) for the OU process. As the scale parameter α grows, the service time tends to become either shorter or much longer than the mean $\mathbb{E}[Y_i]$, rather than being close to $\mathbb{E}[Y_i]$. When Δ_t is short, $p(\Delta_t)$ reduces quickly in Δ_t; meanwhile, when Δ_t is quite long, $p(\Delta_t)$ cannot increase much because it is upper bounded by mse_∞. Because of these two reasons, the aver-

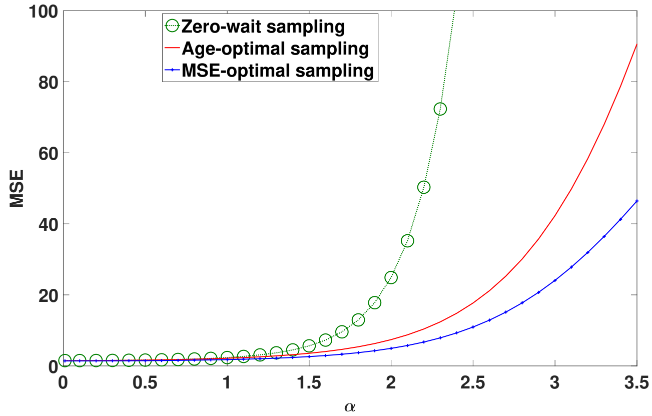

Figure 1.7 MSE vs. the scale parameter α of *i.i.d.* normalized log-normal service time distribution for Wiener process $\mathbb{E}[Y_i] = 1$ and $f_{\max} = 1.2$. Uniform sampling is not feasible here and hence is not plotted.

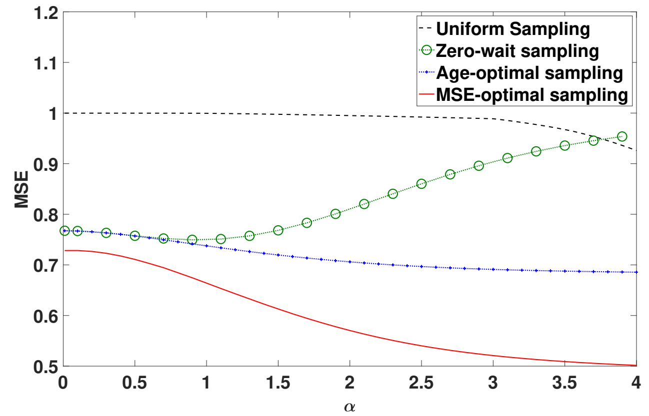

Figure 1.8 MSE vs. the scale parameter α of *i.i.d.* normalized log-normal service time distribution $\mathbb{E}[Y_i] = 1$ and $f_{\max} = 1.2$, where the parameters of the OU process are $\sigma = 1$, $\theta = 0.5$.

age age penalty $\mathsf{mse}_{\text{age-opt}}$ decreases in α. The dropping of $\mathsf{mse}_{\text{opt}}$ in α can be understood in a similar fashion.

On the other hand, the age penalty function of the Wiener process is $p(\Delta_t) = \Delta_t$, which is quite different from the case considered here. We also observe that in all the four figures, the gap between $\mathsf{mse}_{\text{opt}}$ and $\mathsf{mse}_{\text{age-opt}}$ increases as α grows.

1.6 Summary

In this chapter, we have described the optimal sampling policies for minimizing the estimation error for three scalar Gauss-Markov signal processes. The optimal sampler design problem is solved with and without a sampling rate constraint. A smaller estimation error has been achieved by exploiting causal knowledge of the signal values. On the other hand, when the sampler has no knowledge about the signal process, the optimal sampling policy is to optimize a non-linear age function. The problem is formulated as a continuous-time (constrained) MDP with a continuous state space. The optimal sampling policy is a threshold policy on instantaneous estimation error and the threshold is found exactly. The optimal sampling policies can be computed by low-complexity algorithms, and the curse of dimensionality is circumvented.

References

[1] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[2] R. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *ArXiv*, vol. abs/2007.08564, 2020.

[3] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the internet of things," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, Dec. 2019.

[4] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion," *Phys. Rev.*, vol. 36, pp. 823–841, Sept. 1930.

[5] J. L. Doob, "The Brownian movement and stochastic equations," *Annals of Mathematics*, vol. 43, no. 2, pp. 351–369, 1942.

[6] L. Evans, S. Keef, and J. Okunev, "Modelling real interest rates," *Journal of Banking and Finance*, vol. 18, no. 1, pp. 153 – 165, 1994.

[7] A. Cika, M. Badiu, and J. Coon, "Quantifying link stability in Ad Hoc wireless networks subject to Ornstein-Uhlenbeck mobility," in *IEEE ICC*, 2019.

[8] H. Kim, J. Park, M. Bennis, and S. Kim, "Massive UAV-to-ground communication and its stable movement control: A mean-field approach," in *IEEE SPAWC*, June 2018, pp. 1–5.

[9] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Trans. Auto. Control*, vol. 56, no. 9, pp. 2013–2025, Sept. 2011.

[10] M. Ribero and H. Vikalo, "Communication-efficient federated learning via optimal client sampling," *ArXiv*, vol. abs/2007.15197, 2020.

[11] E. Vinogradov, H. Sallouha, S. D. Bast, M. M. Azari, and S. Pollin, "Tutorial on UAV: A blue sky view on wireless communication," 2019, coRR, abs/1901.02306.

[12] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, Feb 2020.

[13] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," *CoRR*, vol. abs/1902.03552, 2019.

[14] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM*, 2012.

[15] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *IEEE SPAWC Workshop*, 2018.

[16] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in *ACM MobiHoc*, 2019.

[17] C. Kam, S. Kompella, G. D. Nguyen, and A. Ephremides, "Effect of message transmission path diversity on status age," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1360–1374, Mar. 2016.

[18] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[19] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6419–6428, Sept. 2018.

[20] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, Mar. 2019.

[21] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, July 2018.

[22] C. Joo and A. Eryilmaz, "Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2556–2568, Dec 2018.

[23] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of the information through queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5215–5232, Aug 2019.

[24] ——, "The age of information in multihop networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, June 2019.

[25] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *IEEE INFOCOM AoI Workshop*, 2018.

[26] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *IEEE INFOCOM*, April 2018, pp. 1844–1852.

[27] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *ACM MobiHoc*, 2018.

[28] N. Lu, B. Ji, and B. Li, "Age-based scheduling: Improving data freshness for wireless real-time traffic," in *ACM MobiHoc*, 2018.

[29] A. Maatouk, Y. Sun, A. Ephremides, and M. Assaad, "Status updates with priorities: Lexicographic optimality," in *IEEE/IFIP WiOpt*, 2020.

[30] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the Internet of things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7468–7482, Nov 2019.

[31] ——, "Minimum age of information in the Internet of things with non-uniform status packet sizes," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1933–1947, 2020.

[32] A. Kosta, N. Pappas, and V. Angelakis, *Age of Information: A New Concept, Metric, and Tool*. Now Publishers Inc, 2018.

[33] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of Information: A New Metric for Information Freshness*. Morgan & Claypool, 2019.

[34] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Optimal sampling and scheduling for timely status updates in multi-source networks," 2020, https://arxiv.org/abs/2001.09863.

[35] Y. Xiao and Y. Sun, "A dynamic jamming game for real-time status updates," in *IEEE INFOCOM AoI Workshop*, April 2018, pp. 354–360.

[36] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in rf-powered communication systems," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4747–4760, Aug. 2020.

[37] ——, "Aoi-optimal joint sampling and updating for wireless powered communication systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14 110–14 115, Nov. 2020.

[38] B. Hajek, K. Mitzel, and S. Yang, “Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm,” *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 608–622, Feb 2008.

[39] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, “Optimal strategies for communication and remote estimation with an energy harvesting sensor,” *IEEE Trans. Auto. Control*, vol. 58, no. 9, pp. 2246–2260, Sept. 2013.

[40] M. Rabi, G. V. Moustakides, and J. S. Baras, “Adaptive sampling for linear state estimation,” *SIAM Journal on Control and Optimization*, vol. 50, no. 2, pp. 672–702, 2012.

[41] K. Nar and T. Başar, “Sampling multidimensional Wiener processes,” in *IEEE CDC*, Dec. 2014, pp. 3426–3431.

[42] X. Gao, E. Akyol, and T. Başar, “Optimal communication scheduling and remote estimation over an additive noise channel,” *Automatica*, vol. 88, pp. 57 – 69, 2018.

[43] J. Chakravorty and A. Mahajan, “Remote estimation over a packet-drop channel with Markovian state,” *IEEE Trans. Auto. Control*, vol. 65, no. 5, pp. 2016–2031, 2020.

[44] C.-H. Tsai and C.-C. Wang, “Unifying AoI minimization and remote estimation: Optimal sensor/controller coordination with random two-way delay,” in *IEEE INFOCOM*, 2020.

[45] N. Guo and V. Kostina, “Optimal causal rate-constrained sampling for a class of continuous Markov processes,” in *IEEE ISIT*, 2020.

[46] A. Arafa, K. A. Banawan, K. G. Seddik, and H. V. Poor, “Timely estimation using coded quantized samples,” *ArXiv*, vol. abs/2004.12982, 2020.

[47] V. Jog, R. J. La, and N. C. Martins, “Channels, learning, queueing and remote estimation systems with a utilization-dependent component,” 2019, coRR, abs/1905.04362.

[48] X. Song and J. W. S. Liu, “Performance of multiversion concurrency control algorithms in maintaining temporal consistency,” in *Proceedings., Fourteenth Annual International Computer Software and Applications Conference*, Oct 1990, pp. 132–139.

[49] S. M. Ross, *Applied Probability Models with Optimization Applications*. San Francisco, CA: Holden-Day, 1970.

[50] R. A. Maller, G. Müller, and A. Szimayer, “Ornstein-Uhlenbeck processes and extensions,” in *Handbook of Financial Time Series*, T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 421–437.

[51] G. Peskir and A. N. Shiryaev, *Optimal Stopping and Free-Boundary Problems*. Basel, Switzerland: Birkhäuswer Verlag, 2006.

[52] T. Soleymani, S. Hirche, and J. S. Baras, “Optimal information control in cyber-physical systems,” *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 1 – 6, 2016.

[53] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Academic Press, 2007.

[54] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 5th ed. Springer-Verlag Berlin Heidelberg, 2000.

[55] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[56] J. H. Mathews and K. K. Fink, *Numerical Methods Using Matlab*. Simon & Schuster, Inc., 1998.

[57] M. Spivak, *Calculus*, 4th ed. Publish or Perish, 2008.

[58] C.-H. Tsai and C.-C. Wang, "Age-of-information revisited: Two-way delay and distribution-oblivious online algorithm," in *IEEE ISIT*, 2020.