



MIT Open Access Articles

Locally adaptive sampling

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Feizi, Soheil, Vivek K Goyal, and Muriel Medard. "Locally Adaptive Sampling." IEEE, 2010. 152–159. © Copyright 2010 IEEE
As Published	http://dx.doi.org/10.1109/ALLERTON.2010.5706901
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/72184
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Locally Adaptive Sampling

Soheil Feizi

RLE at MIT

Email: sfeizi@mit.edu

Vivek K Goyal

RLE at MIT

Email: vgoyal@mit.edu

Muriel Médard

RLE at MIT

Email: medard@mit.edu

Abstract—In this paper, we introduce a class of *Locally Adaptive Sampling* schemes. In this sampling family, time intervals between samples can be computed by using a function of previously taken samples, called a sampling function. Hence, though it is a non-uniform sampling scheme, we do not need to keep sampling times. The aim of LAS is to have the average sampling rate and the reconstruction error satisfy some requirements. We propose four different schemes of LAS. The first two are designed for deterministic signals. First, we derive a Taylor Series Expansion (TSE) sampling function, which only assumes the third derivative of the signal is bounded, but requires no other specific knowledge of the signal. Then, a Discrete Time-Valued (DTV) sampling function is proposed, where the sampling time intervals are chosen from a lattice. Next, we consider stochastic signals. We propose two sampling methods based on linear prediction filters: a Generalized Linear Prediction (GLP) sampling function, and a Linear Prediction sampling function with Side Information (LPSI). In GLP method, we only assume the signal is locally stationary. However, LPSI is specifically designed for a known signal model.

I. INTRODUCTION

Taking samples from a signal satisfying some requirements on the sampling rate and the reconstruction error is one of the main problems in signal processing. The considered signal can be deterministic (either band-limited or non-bandlimited) or stochastic. In this paper, we propose a sampling family which can be applied on both deterministic and stochastic signals, satisfying sampling and reconstruction requirements.

For band-limited deterministic signals, this problem is well-studied. For example, the Nyquist sampling theorem proposes an appropriate uniform sampling setup which leads to a zero reconstruction error. A non-uniform sampling scheme based on level-crossings with iterative decoding is considered in [1], while reference [2] presents an approach, based on level crossings with a filtering technique, which adapts the sampling rate and filter order by analyzing the input signal variations online. Also, two adaptive sampling schemes for band-limited deterministic signals are proposed in [3]. These schemes are using some definitions of local bandwidth, based on linear time-varying low pass filters [4], and time-warping of bandlimited signals [5].

Many of these results for band-limited signals have been extended to the case of non bandlimited signals [6]. For example, it has been shown that some splines which are non bandlimited signals can be reconstructed from uniformly spaced samples similarly to bandlimited signals. However, it requires to use a non causal IIR filter in some techniques. Reference [7] considers joint sampling of amplitude and derivative of spline-like signals, needing only FIR filters.

For a discrete stochastic signal, one scheme is to take samples uniformly at a high rate and then, use source coding to compress these samples approximately to their entropy rate. This technique has two parts: uniform sampling, and source coding. To have an appropriate performance, we need long blocks of samples to be able to use source coding efficiently, especially if statistical properties of the signal vary slowly in time. This block-based approach may lead to a large delay on reconstruction side. Instead, our proposed scheme is a real-time compression scheme. It adaptively compresses the signal by using its local properties causally.

In this paper, we introduce a new family of adaptive non-uniform sampling schemes. In this sampling family, time intervals between samples can be computed by using a function of previously taken samples. This function is called a *sampling function*. We refer to this sampling family as *Locally Adaptive Sampling* (LAS). The aim in this sampling process is to take samples of a discrete or continuous signal in a way that balances the reconstruction distortion and the average sampling rate.

Consider a continuous signal $X(t)$. Suppose the i^{th} sample is taken at time t_i . Define $T_i \triangleq t_{i+1} - t_i$ and $\Delta_i \triangleq X(t_{i+1}) - X(t_i)$. Then, we take the $(i+1)^{\text{th}}$ sample after a time interval with length $T_i = f(\bigcup_{j=i-M}^{i-1} \{T_j, \Delta_j\})$, where f is called the sampling function and M is called the order of the sampling function (Figure 1). The sampling function is known on both the sampling and the reconstruction sides. Similarly, we can take samples from a discrete time signal, by using a suitable sampling function. Hence, LAS can be applied on both discrete and continuous signals.

This sampling structure is non-uniform except in trivial cases when the sampling function is a constant-valued function. However, the key characteristic of our approach is that, unlike traditional non-uniform sampling procedures, keeping sampling times is not necessary. Rather, these times can be recovered by using the sampling function and previously taken samples. In the above example, we have $t_{i+1} = t_i + f(\bigcup_{j=i-M}^{i-1} \{T_j, \Delta_j\})$.

In the above example, LAS is causal because the next sampling time depends on samples taken before that time. In general, it can be designed to be non-causal. But, in this paper, we consider only causal sampling schemes. Note that the reconstruction method can be causal or non-causal. LAS is an adaptive process, because the sampling function, f , depends on local characteristics of the signal. Finding an appropriate sampling function of LAS depends on sampling requirements such as the sampling rate, the distortion requirement, etc.

TABLE I
TAXONOMY OF PROPOSED SAMPLING FUNCTIONS

Sampling Functions (LAS)	Blind	With Side Information
Deterministic	Taylor Series Expansion (TSE)	Discrete Time-Valued (DTV)
Stochastic	Generalized Linear Prediction (GLP)	Linear Prediction With Side Information (LPSI)

The aim of LAS is to balance between the average sampling rate and the reconstruction distortion. Note that, this objective is different from the one considered in *change point analysis* or *active learning*. There, the objective is to find points of the signal at which the statistical behaviors of the signal change, by causal or non-causal sampling, respectively (e.g., [8] and [9]).

We proceed by introducing four examples of LAS: for deterministic and stochastic signals, with/without side information about the signal. The taxonomy of sampling functions we propose is given in Table I. In section II, we use Taylor's theorem to derive a suitable sampling function for deterministic signals. This method is called Taylor Series Expansion (TSE) method. Then, a Discrete Time-Valued (DTV) method for this signal family is proposed in Section III, where the sampling time intervals are chosen from a lattice. For time-varying stochastic signals, we propose two sampling methods based on linear prediction filters: Generalized Linear Prediction (GLP) method in Section IV, and Linear Prediction with Side Information (LPSI) in Section V. For TSE and GLP methods, we have a general condition on the considered signal (bounded third derivative for TSE, and being locally stationary for GLP). These methods are called blind sampling methods. However, DTV and LPSI methods are non-blind (with side information), because the sampling scheme is specifically designed for a known signal model.

II. TAYLOR SERIES EXPANSION METHOD OF LAS

In this section, we use Taylor's theorem to derive a suitable sampling function for deterministic signals. Various instantiations of LAS can be derived by using Taylor's theorem, depending on signals of interest, the allowed complexity of the sampling function f , and the distortion requirement. In this section, by an example, we explain how a sampling function can be computed for a specific setup. Suppose $X(t)$ is a continuous signal. We consider the following assumptions:

- (C₁) The order of the sampling function is two. In other words, the sampling function is a function of three previously taken samples ($T_i = f(\bigcup_{j=i-2}^{i-1} \{T_j, \Delta_j\})$).
- (C₂) The reconstruction method is connecting two adjacent samples by a line. We call this non-causal linear interpolation.

Also, the distortion requirement is assumed to be as follows:

- (C₃) Suppose $\hat{X}(t)$ is the reconstructed signal. We want to have $|X(t) - \hat{X}(t)| < D_1$, for all t , where D_1 is a sampling parameter.

The upper bound on the absolute value of the third derivative is used a parameter in the sampling function and is also used in the analysis.

Theorem 1. *In the Taylor Series Expansion method of LAS, if $|X'''(t)|$ is uniformly bounded by a constant M , under assumptions C_1 - C_2 , the following sampling function satisfies the sampling requirement C_3 :*

$$T_i = \arg \max_T \quad T$$

$$s.t. \quad (c_1 T + c_2) T^2 \leq D_1 \quad (1)$$

where c_1 and c_2 are constants, defined as follows,

$$c_1 = \frac{M}{3}$$

$$c_2 = \frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|}{\frac{T_{i-1} + T_{i-2}}{2}} + \frac{M}{3} \frac{(T_{i-1} + T_{i-2})^2 + (T_{i-1})^2}{T_{i-2}}$$

Before proceeding by the proof, it is insightful to investigate the behavior of T_i , with respect to different involved parameters in (1).

- Increasing D_1 leads to an increase in T_i . Intuitively, the higher the allowed distortion, the lower the sampling rate, and the larger the sampling time intervals.
- The first term of c_2 (i.e.,

$$|\Delta_{i-1}/T_{i-1} - \Delta_{i-2}/T_{i-2}| / ((T_{i-1} + T_{i-2})/2)$$

can be viewed as an approximation of $|X'''(t)|$ at time t_i . Since the reconstruction method is a first order linear filter, the higher the second derivative, the faster changes of the signal. Therefore, by increasing $|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|$, T_i should decrease.

We present the proof of Theorem 1 in the following.

Proof: By Taylor's theorem, for each t in $[t_i, t_{i+1})$, there exists s_1 in (t_i, t) such that,

$$X(t) = X(t_i) + X'(t_i)(t - t_i) + X''(t_i) \frac{(t - t_i)^2}{2} + X'''(s_1) \frac{(t - t_i)^3}{6}. \quad (2)$$

Since $|X'''(t)| \leq M$ for all t , we have $X_l(t) \leq X(t) \leq X_u(t)$ where,

$$X_u(t) = X(t_i) + X'(t_i)(t - t_i) + X''(t_i) \frac{(t - t_i)^2}{2} + M \frac{(t - t_i)^3}{6}$$

$$X_l(t) = X(t_i) + X'(t_i)(t - t_i) + X''(t_i) \frac{(t - t_i)^2}{2} - M \frac{(t - t_i)^3}{6}. \quad (3)$$

For the reconstruction, according to assumption A_2 , we connect adjacent samples by a line. Hence, for $t \in [t_i, t_{i+1})$, we have,

$$\hat{X}(t) = X(t_i) + (t - t_i) \frac{\Delta_i}{T_i}. \quad (4)$$

By Taylor's theorem, there exists $s_2 \in (t_i, t_{i+1})$ such that,

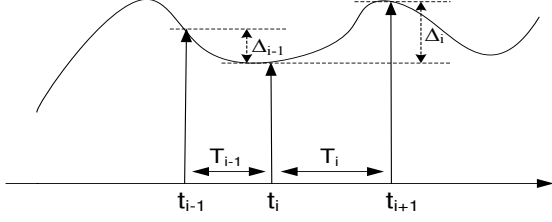


Fig. 1. LAS setup

$$\begin{aligned}
 X(t_{i+1}) &= X(t_i) + X'(t_i)(t_{i+1} - t_i) \\
 &+ X''(t_i) \frac{(t_{i+1} - t_i)^2}{2} + X'''(s_2) \frac{(t_{i+1} - t_i)^3}{6}.
 \end{aligned}$$

Hence,

$$\frac{\Delta_i}{T_i} = X'(t_i) + X''(t_i) \frac{T_i}{2} + X'''(s_2) \frac{T_i^2}{6}. \quad (5)$$

Having (3), (4) and (5), for $t \in [t_i, t_{i+1})$, we can write,

$$\begin{aligned}
 |\tilde{X}(t)| &= |X(t) - \hat{X}(t)| \\
 &\leq (t - t_i)[(t - t_i + T_i)|X''(t_i)|/2 \\
 &\quad + ((t - t_i)^2 + T_i^2)M/6] \\
 &\leq D_1.
 \end{aligned} \quad (6)$$

The right-hand side of (6) is an increasing function of t for $t \in [t_i, t_{i+1})$. Hence, its supremum happens at $t = t_{i+1}$. Thus, we have,

$$\left(\frac{M}{3}T_i + |X''(t_i)|\right)T_i^2 \leq D_1. \quad (7)$$

By using Taylor's theorem at times $t = t_{i-1}$ and $t = t_{i-2}$, centered at $t = t_i$, there exist s_3 and s_4 such that $t_{i-1} < s_3 < t_i$ and $t_{i-2} < s_4 < t_i$, and,

$$\begin{aligned}
 \frac{X(t_{i-1}) - X(t_i)}{T_{i-1}} &= -X'(t_i) + X''(t_i) \frac{T_{i-1}}{2} \\
 &\quad - X'''(s_3) \frac{T_{i-1}^2}{6}
 \end{aligned} \quad (8)$$

$$\begin{aligned}
 \frac{X(t_{i-2}) - X(t_i)}{T_{i-1} + T_{i-2}} &= -X'(t_i) + X''(t_i) \frac{T_{i-1} + T_{i-2}}{2} \\
 &\quad - X'''(s_4) \frac{(T_{i-1} + T_{i-2})^2}{6}.
 \end{aligned} \quad (9)$$

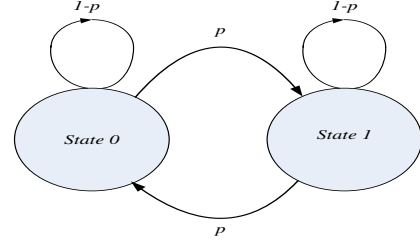


Fig. 2. A hidden Markov chain for θ_n

Hence,

$$\begin{aligned}
 |X''(t_i)| &\leq \frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2} + \Delta_{i-1}}{T_{i-2} + T_{i-1}}|}{\frac{T_{i-2}}{2}} \\
 &\quad + \frac{M}{3} \frac{(T_{i-1} + T_{i-2})^2 + (T_{i-1})^2}{T_{i-2}} \\
 &= \frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|}{\frac{T_{i-1} + T_{i-2}}{2}} \\
 &\quad + \frac{M}{3} \frac{(T_{i-1} + T_{i-2})^2 + (T_{i-1})^2}{T_{i-2}}. \quad (10)
 \end{aligned}$$

Combining (7) and (10) completes the proof. ■

A counterpart of this method can be derived for discrete signals. For example, suppose $X[n]$ is a discrete signal, where n is a non-negative integer. The i^{th} sample is taken at time n_i . Let us define $N_i \triangleq n_{i+1} - n_i$ and $\Delta_i \triangleq X[n_{i+1}] - X[n_i]$. Then, the $(i+1)^{\text{th}}$ sample is taken after $N_i = f(\bigcup_{j=i-2}^{i-1} \{N_j, \Delta_j\})$ samples.

Let us define $X'[n] \triangleq X[n] - X[n-1]$, $X''[n] \triangleq X[n] - 2X[n-1] + X[n-2]$ and $X'''[n] \triangleq X[n] - 3X[n-1] + 3X[n-2] - X[n-3]$. Suppose $|X'''[n]|$ is uniformly bounded by a constant M . Also, consider the following assumptions as counterparts for assumptions C_1 , C_2 and C_3 :

(C₄) The sampling function is a function of three previously taken samples (i.e., $N_i = f(\bigcup_{j=i-2}^{i-1} \{N_j, \Delta_j\})$).

(C₅) The reconstruction method is a linear interpolation among taken samples.

(C₆) Suppose $\hat{X}[n]$ is the reconstructed signal. We want to have $|X[n] - \hat{X}[n]| < D_1$, where D_1 is a sampling parameter.

Theorem 2. In the Taylor Series Expansion method of LAS for discrete signals, if $|X'''[n]|$ is bounded by a constant M , under assumptions C₄-C₅, the following sampling function satisfies the sampling requirement C₆:

$$\begin{aligned}
 N_i &= \arg \max [N] \\
 \text{s.t.} &\quad (c_1 N + c_2) N^2 \leq D_1
 \end{aligned} \quad (11)$$

where c_1 and c_2 are constants, defined as follows,

$$\begin{aligned}
 c_1 &= \frac{M}{3} \\
 c_2 &= \frac{|\frac{\Delta_{i-1}}{N_{i-1}} - \frac{\Delta_{i-2}}{N_{i-2}}|}{\frac{N_{i-1} + N_{i-2}}{2}} + \frac{M}{3} \frac{(N_{i-1} + N_{i-2})^2 + (N_{i-1})^2}{N_{i-2}}.
 \end{aligned}$$

If instead of assumption C_3 , the normalized \mathcal{L}_1 norm of the error signal is considered as a distortion measure (i.e., $\frac{1}{T} \int_{t=0}^T |\tilde{X}(t)| dt \leq D_1$), the following sampling function can be derived:

Corollary 3. *In the Taylor Series Expansion method of LAS, if $|X'''(t)|$ is bounded by a constant M , under assumptions C_1 - C_2 , the following sampling function satisfies the \mathcal{L}_1 norm of the error signal to be less than D_1 :*

$$T_i = \arg \max_T \quad T \quad (12)$$

s.t. $(c_1 T + c_2) T^2 \leq D_1$

where c_1 and c_2 are constants, defined as follows,

$$c_1 = \frac{M}{8},$$

$$c_2 = \frac{5}{12} \left(\frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|}{\frac{T_{i-1} + T_{i-2}}{2}} + \frac{M (T_{i-1} + T_{i-2})^2 + (T_{i-1})^2}{3 T_{i-2}} \right).$$

A. Numerical Evaluations

While our proof has considered deterministic signals, since the proposed sampling method is blind, it renders this method a well-suited choice for uncertain signals. Hence, in this section, we empirically evaluate the performance of our scheme when the signal is deterministic over stochastically varying time intervals. To do this, consider an underlying hidden Markov chain (MC) with a transition probability p , depicted in Figure 2. At each state, the signal is drawn by a third order spline, up-sampled by a factor of u_s , where s represents the state of MC. Thus, if the state of MC is s at time n , $X[n]$ and $X[n + u_s + 1]$ are drawn from a uniform distribution over $[-A, A]$. Signal values at other times are determined by a third order spline interpolation. Figure 3-a shows a sample of this signal family, where $p = 0.001$, $u_0 = 1$, $u_1 = 20$ and $A = 0.5$.

We use the normalized \mathcal{L}_1 norm of the error as a distortion measure. The sampling function is given by (12).

Since our sampling scheme is blind, in our simulations, the signal model is unknown and we are only allowed to use the sampling function of (12). To do this, we need M , an upper bound for the third derivative of the signal. From taken samples, we approximate the third derivative as follows. First, we define,

$$w(t_i) \triangleq \frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|}{\frac{T_{i-1} + T_{i-2}}{2}}. \quad (13)$$

Then, $\zeta(t_i) \triangleq \frac{|w(t_i) - w(t_{i-1})|}{(T_{i-1} + T_{i-2} + T_{i-3})/3}$ is used as an approximation of the third derivative of the signal by Taylor's theorem. At each time, M is chosen as the maximum of $\{\zeta(t_i), \zeta(t_{i-1}), \dots, \zeta(t_{i-W})\}$. In our simulations, we pick $W = 10$.

Figure 3-b shows the number of samples (i) versus time (n_i) when $D_1 = 0.2$. Hence, at each time n_i , the slope of the tangent line on the curve represents the sampling rate at that time. We draw state transitions of the underlying hidden Markov chain in Figure 3-a to clarify the rate adaptation in

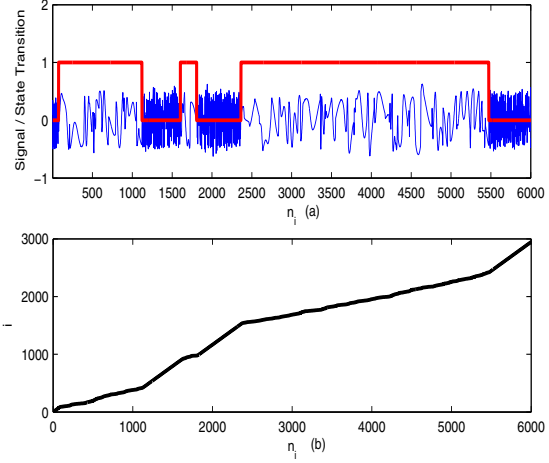


Fig. 3. a) A sample signal and its MC state transitions used in Section (II-A). b) Sampling rate versus time for this signal with TSE method.

different states. In state 0, in which the signal model is an up-sampled third-order spline by a factor $u_0 = 1$, the sampling rate is greater than the one of state 1, where $u_1 = 20$. The overall sampling rate is 0.65 with error 0.005 for this signal sample. Note that although the sampling rate seems roughly the same within each state, it has some fluctuations.

III. DISCRETE TIME-VALUED METHOD OF LAS

In this section, we introduce a family of LAS where sampling time intervals are discrete values. In this method, the sampling rate is being adapted based on an estimation of local variations of the signal. This method can be applied to both continuous and discrete signals. In this section, we consider its continuous time version; while its discrete time version can be derived similarly to Theorem 2.

If $|X'''(t)|$ is uniformly bounded by a constant M , under assumptions C_1 and C_2 mentioned in Section II, along with (10), an estimate of local variation of the signal at time t_i , by using previously taken samples, can be written as follows,

$$w(t_i) \triangleq \frac{|\frac{\Delta_{i-1}}{T_{i-1}} - \frac{\Delta_{i-2}}{T_{i-2}}|}{\frac{T_{i-1} + T_{i-2}}{2}}. \quad (14)$$

By a similar argument as the (10), the error of this estimation can be bounded as follows,

$$|w(t_i) - X''(t_i)| \leq \frac{M (T_{i-1} + T_{i-2})^2 + (T_{i-1})^2}{3 T_{i-2}}. \quad (15)$$

Consider the following heuristic sampling function,

$$T_i = \begin{cases} f_1(T_{i-1}) & w(t_i) < th_1 \text{ and } T_i > T^{min} \\ T_{i-1} & th_1 \leq w(t_i) < th_2 \\ f_2(T_{i-1}) & w(t_i) \geq th_2 \text{ and } T_i < T^{max} \end{cases} \quad (16)$$

where $f_1(T_{i-1}) > T_{i-1}$ and $f_2(T_{i-1}) < T_{i-1}$. T^{min} and T^{max} make sampling time intervals to be bounded.

Thresholds th_1 and th_2 depend on signal characteristics and sampling requirements. If $w(t_i)$ is smaller than a threshold, the signal's slope variations are small and we can decrease the sampling rate in this case, since $f_1(T_{i-1}) > T_{i-1}$. An analogous argument can be expressed when $w(t_i)$ is greater than a threshold.

An example of a sampling function (16) with linear increase or decrease of T_i can be expressed as follows,

$$T_i = \begin{cases} T_{i-1} + \epsilon_1 & w(t_i) < th_1 \text{ and } T_i > T^{min} \\ T_{i-1} & th_1 \leq w(t_i) < th_2 \\ T_{i-1} - \epsilon_2 & w(t_i) \geq th_2 \text{ and } T_i < T^{max} \end{cases} \quad (17)$$

where ϵ_1 and ϵ_2 are positive constants. Note that, given T_{i-1} , there are only three possibilities for T_i , so, the sampling time intervals are discrete.

IV. GENERALIZED LINEAR PREDICTION METHOD OF LAS

In this section, we introduce another family of LAS whose sampling function is based on a generalized linear prediction filter. This sampling method is for discrete stochastic signals. Here, we need the signal to be *locally stationary* ([10] and [11]). In Section IV-B, we explain this family of signals in details.

The used reconstruction method and the distortion measure are as follows,

- (C₇) The reconstruction method is using a generalized linear prediction filter.
(C₈) If $\hat{X}[n]$ is the reconstructed signal, we want to have $E[(X[n] - \hat{X}[n])^2] \leq D_2$, where D_2 is a sampling parameter.

In the following, we introduce a generalized linear prediction filter for stationary signals.

A. Generalized Linear Prediction

Suppose $X[n]$ is a stationary signal. Assume we have M samples of $X[n]$ at times $n - m_1, n - m_2, \dots$, and $n - m_M$. Our aim is to predict linearly $X[n]$ by using these known samples so that the expected mean square error is minimized (MMSE predictor).

$$\begin{aligned} \min_{w_{m_1}, \dots, w_{m_M}} \quad & E[|\tilde{X}[n]|^2] \\ \text{subject to} \quad & \hat{X}[n] = \sum_{k=1}^M w_{m_k} X[n - m_k] \\ & \tilde{X}[n] = X[n] - \hat{X}[n]. \end{aligned} \quad (18)$$

Let us call a solution of this linear optimization, $w_{m_i}^*$, for $1 \leq i \leq M$. The only difference of this setup with a traditional linear prediction ([12]) is to predict $X[n]$ by a set of non-uniform samples.

In an optimal scheme, the error term should be orthogonal to all known samples.

$$E[X[n - m_k] \tilde{X}^*[n]] = 0 \quad (19)$$

for $k = 1, \dots, M$, where $\tilde{X}^*[n] = X[n] - \sum_{k=1}^M w_{m_k}^* X[n - m_k]$. An auto-correlation function of $X[n]$ can be written as follows,

$$r[i] = E[X[n] X^c[n - i]] \quad (20)$$

where $X^c[n]$ is the conjugate of $X[n]$. Since in this paper, we deal with real signals, without loss of generality, we ignore conjugation effects. Hence, by using (19) and (20), we have the following set of linear equations,

$$r[-m_k] = \sum_{i=1}^M w_{m_i}^* r[m_i - m_k] \quad (21)$$

for $k = 1, \dots, M$. Sometimes, it is easier to express (21) in a matrix form.

Let us define the following matrices,

$$\begin{aligned} \mathbf{m} &= [m_1, \dots, m_M]^T \\ \mathbf{X}_m^n &= [X[n - m_1], \dots, X[n - m_M]]^T \\ \mathbf{p} &= [r[-m_1], \dots, r[-m_M]]^T \\ \mathbf{w}_m^* &= [w_{m_1}^*, \dots, w_{m_M}^*]^T \\ \mathbf{R} &= E[(\mathbf{X}_m^n)(\mathbf{X}_m^n)^T]. \end{aligned} \quad (22)$$

Thus, linear equations of (21) can be written as a matrix multiplication,

$$\mathbf{p} = \mathbf{R} \mathbf{w}_m^*.$$

For $X[n]$ with zero mean, define,

$$\begin{aligned} \sigma_X^2 &= E[|X[n]|^2] = r[0] \\ \sigma_{\tilde{X}^*}^2 &= E[|\tilde{X}^*[n]|^2]. \end{aligned}$$

Theorem 4. $\sigma_{\tilde{X}^*}^2 = r[0] - \mathbf{p}^T \mathbf{w}_m^*$.

Proof: Since $X[n]$ has zero mean, and $X[n] = \hat{X}[n] + \tilde{X}^*[n]$, by using (19), we can write,

$$\begin{aligned} \sigma_{\tilde{X}^*}^2 &= \sigma_X^2 - \sigma_{\hat{X}}^2 \\ &= r[0] - \sigma_{\hat{X}}^2 \end{aligned} \quad (23)$$

where $\hat{X}[n] = (\mathbf{w}_m^*)^T (\mathbf{X}_m^n)$. Therefore,

$$\begin{aligned} \sigma_{\hat{X}}^2 &= E[|\hat{X}[n]|^2] \\ &= (\mathbf{w}_m^*)^T E[(\mathbf{X}_m^n)(\mathbf{X}_m^n)^T] \mathbf{w}_m^* \\ &= (\mathbf{w}_m^*)^T \mathbf{R} \mathbf{w}_m^* \\ &= \mathbf{p}^T \mathbf{w}_m^*. \end{aligned} \quad (24)$$

Equations (23) and (24) establish the theorem. \blacksquare

B. LAS with a generalized linear prediction filter

In this section, we assume that the signal $X[n]$ is a *locally stationary* signal. Locally stationary processes can be used as a tool to model systems where their statistical behavior varies with time. We use a definition of locally stationary processes presented in [10] and [11]. Intuitively, a locally stationary process is a process where we can approximate its local covariance coefficients within an error. Reference [10] approximates the covariance of a locally stationary process

by a covariance which is diagonal in basis of cosine packets. While, [11] proposes a method to estimate the covariance from sampled data. For simplicity, we assume a fixed window size. But, this can also vary by time.

Let us define a window $W_L[n]$ with length L as follows,

$$W_L[n] = \begin{cases} 1 & 0 \leq n \leq L-1 \\ 0 & \text{O.W.} \end{cases} \quad (25)$$

$X_L^{n_i}[n]$ is a truncated version of $X[n]$ which has its samples over $n_i - L + 1 \leq n \leq n_i$. I.e.,

$$X_L^{n_i}[n] = X[n] \cdot W_L[n - n_i + L - 1]. \quad (26)$$

For $X_L^{n_i}[n]$, auto-correlation coefficients can be written as follows,

$$r_L^{n_i}[k] = E[X_L^{n_i}[n]X_L^{n_i}[n-k]]. \quad (27)$$

By using these coefficient, for $\mathbf{m} = [0, 1, \dots, M-1]^T$, where $M < L$, we define matrices $(\mathbf{X}_m^{n_i})_L$, $\mathbf{p}_L^{n_i}$, $(\mathbf{w}_m^*)_L^{n_i}$ and $\mathbf{R}_L^{n_i}$, similarly to (22). Since $X[n]$ is assumed to be locally stationary, for any time n_0 and any given ϵ , according to [10] and [11], there exists an appropriate L such that,

$$|E[|\tilde{X}[n_i+1]|^2] - r_L^{n_i}[0] + (\mathbf{p}_L^{n_i})^T (\mathbf{w}_m^*)_L^{n_i}| < \epsilon, \quad (28)$$

where $\tilde{X}[n_i+1] = X[n_i+1] - ((\mathbf{w}_m^*)_L^{n_i})^T (\mathbf{X}_m^{n_i})_L$. We refer to L as a window size of the locally stationary signal. Define L_0 and L_1 as the minimum and the maximum allowed window sizes, respectively. The stochastic nature of the signal affects L_0 and L_1 .

Intuitively, since $X[n]$ is locally stationary, its MMSE linear prediction filter with locally estimated autocorrelation coefficients leads to an approximately same error as the one of stationary signals. Now, we introduce a setup of LAS by using a MMSE generalized linear prediction filter for locally stationary signals. Except being locally stationary, we do not have any other assumptions on the signal. Hence, this method is referred as a blind sampling method.

Suppose $X[n]$ is a locally stationary signal with window size L . Say we have M samples $X[n_i]$, $X[n_i - m_2]$, ..., $X[n_i - m_M]$, where $0 < m_2 < m_3 < \dots < m_M < L$. Now, consider a truncated signal $X_L^{n_i}[n]$ as defined in (26). If we only use taken samples of this truncated signal (i.e., $X[n - m_i]$ for $1 \leq i \leq M$), we can compute approximations for $\mathbf{R}_L^{n_i}$, $\mathbf{p}_L^{n_i}$ and $(\mathbf{w}_m^*)_L^{n_i}$ which we call them $\hat{\mathbf{R}}_L^{n_i}$, $\hat{\mathbf{p}}_L^{n_i}$ and $(\hat{\mathbf{w}}_m^*)_L^{n_i}$, respectively. If L is sufficiently larger than L_0 , we will have enough known samples in the truncated signal and these approximations can be close to $\mathbf{R}_L^{n_i}$, $\mathbf{p}_L^{n_i}$ and $(\mathbf{w}_m^*)_L^{n_i}$.

Then, we linearly predict $X[n_i + N_i]$ by using samples $X[n_i]$, $X[n_i - m_2]$, ..., and $X[n_i - m_M]$. We assume that parameter L_1 of our locally stationary signal is sufficiently large. Hence, by using Theorem 4 and (28),

$$|E[|\tilde{X}[n_i + N_i]|^2] - \tilde{r}_L^{n_i}[0] + (\hat{\mathbf{p}}_L^{n_i})^T (\hat{\mathbf{w}}_m^*)_L^{n_i}| < \epsilon \quad (29)$$

where ϵ is a small positive constant and $\mathbf{m} = [N_i, N_i + m_2, \dots, N_i + m_M]$. The reconstructed signal can be written

as,

$$\hat{X}[n_i + N_i] = ((\hat{\mathbf{w}}_m^*)_L^{n_i})^T \mathbf{X}_m^{n_i + N_i}. \quad (30)$$

A sampling function for this scheme chooses the greatest possible N_i to have the expected error less than a threshold (D_2). Thus, we can write this sampling function as the following linear optimization setup:

$$\begin{aligned} \max \quad & N_i \\ \text{s.t.} \quad & \hat{X}[n_i + N_i] = ((\hat{\mathbf{w}}_m^*)_L^{n_i})^T \mathbf{X}_m^{n_i + N_i} \\ & \hat{\mathbf{p}}_L^{n_i} = \hat{\mathbf{R}}_L^{n_i} (\hat{\mathbf{w}}_m^*)_L^{n_i} \\ & |\tilde{r}_L^{n_i}[0] - (\hat{\mathbf{p}}_L^{n_i})^T (\hat{\mathbf{w}}_m^*)_L^{n_i}| < D_2. \end{aligned} \quad (31)$$

Note that, if the window size L is sufficiently larger than the minimum allowed window size L_0 , we have enough known samples in our window, and these approximations would be appropriate. However, if we do not have enough known samples in our window, we can use autocorrelation coefficients of the previous window.

C. Numerical Evaluation

Analogously to Section II-A, to evaluate the performance of the proposed method, we select a signal model which is an auto regressive model with order one (AR(1)) over stochastically varying intervals. In other words, the used signal model in our simulations is a *Markov jump linear system model* described as follows,

$$X[n] = \alpha_{\theta_n} X[n-1] + Z_{\theta_n}[n], \quad (32)$$

where θ_n is the state of a hidden Markov Chain (MC) with state transition probability p , depicted in Figure 2. At time n , if MC is at state 0, $\theta_n = 0$, otherwise, $\theta_n = 1$. Depending on the value of θ_n , the signal is generated by a first order Auto Regressive model with parameter α_{θ_n} . In our simulations, we assume $\alpha_0 = 0.7$ and $\alpha_1 = 0.97$. Z_{θ_n} is a white Gaussian noise signal with mean 0 and variance $1 - \alpha_{\theta_n}^2$. Hence, in our simulations, we have, $\sigma_{Z_0}^2 \approx 0.5$ and $\sigma_{Z_1}^2 \approx 0.05$. We assume state transition probability of the underlying hidden Markov chain is 0.001 ($p = 0.001$). Note that within each state, this signal is a locally stationary signal. Figure 4-a shows a sample of this signal model.

In our sampling scheme, we assume that we do not know this specific signal model and we are only allowed to use the sampling function (31). We assume $M = 5$ and $L = 100$ throughout the simulations.

Figure 4-b shows the number of taken samples (i) versus time n_i , where $D_2 = 0.4$. Hence, at each time n_i , the slope of the tangent line on the curve represents the sampling rate at that time. State transitions of MC for this signal are depicted in Figure 4-a to clarify the rate adaptation in different states. In state 0, where the noise variance is 0.5, the sampling rate is greater than the one in state 1, where the noise variance is 0.05. The overall sampling rate is 0.55 with $MSE = 0.14$ for this signal sample.

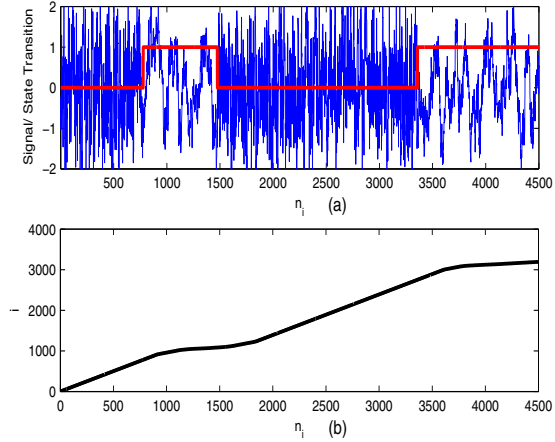


Fig. 4. a) A sample signal and its MC state transitions used in Section IV-C. b) Sampling rate versus time for this signal with GLP method.

V. LINEAR PREDICTION METHOD WITH SIDE INFORMATION

In this section, we propose a sampling function based on linear prediction with side information about the signal. Hence, this method is non-blind. Consider the signal model described in (32). Suppose its parameters (i.e., α_{θ_n}) are known for every n . Also, assume that the transition probability of the underlying MC, p , is small. These parameters form the side information. For reconstruction, we use a linear prediction filter.

Consider $A = \{(R, MSE)\}$ as a set of achievable rate-distortion pairs for this signal model. We consider MSE of the error as a distortion measure. Similarly, define $A_s = \{(R_s, MSE_s)\}$, a set of achievable rate-distortion pairs within state s . The next sample is taken when the prediction error (or, the noise variance) exceeds a threshold D_3 .

Theorem 5. For the signal model described in (32), the following rate-distortion pairs are achievable,

$$A = \{(R, MSE) | (R, MSE) = \frac{1}{2}(1/K_0, MSE_0) + \frac{1}{2}(1/K_1, MSE_1)\} \quad (33)$$

where,

$$MSE_s = \frac{\sum_{i=1}^{K_s-1} (1 - \alpha_s^{2i})}{K_s} \quad (34)$$

and,

$$K_s = \begin{cases} \lceil \frac{\log(1-D_3)}{2 \log(\alpha_s)} \rceil & \alpha_s \notin \{0, 1\} \\ 1 & \alpha_s = 0 \\ 0 & \alpha_s = 1 \end{cases}.$$

$0 \leq D_3 \leq 1$ is a sampling parameter and $s \in \{0, 1\}$ represents the state of the MC.

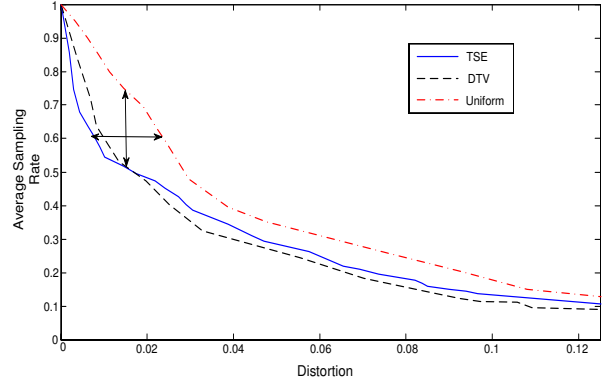


Fig. 5. A rate-distortion comparison between TSE and DTV of FANS and a uniform sampling

Proof: By using (32) for each state s , we have,

$$\begin{aligned} X[n+K] &= \alpha_s^K X[n] + \sum_{i=1}^K Z[n+i] \alpha_j^{K-i} \\ &= \alpha_s^K X[n] + Z_{s,K}. \end{aligned} \quad (35)$$

Hence,

$$\sigma_{Z_{s,K}}^2 = 1 - \alpha_s^{2K}. \quad (36)$$

We take a sample when the prediction error (or, the noise variance) exceeds a threshold D_3 . Hence, for $\alpha_s \notin \{0, 1\}$, we choose the maximum value of K to have $1 - \alpha_s^{2K} \leq D_3$. Hence, at state s , $K_s = \lceil \frac{\log(1-D_3)}{2 \log(\alpha_s)} \rceil$, when $\alpha_s \notin \{0, 1\}$. Cases when $\alpha_s = 0$ or $\alpha_s = 1$ are trivial.

Suppose $MSE_s(l) = E[(X[n+l] - \hat{X}[n+l])^2]$ at state s . Since we take samples at times n and $n+K_s$, $MSE_s(0) = MSE_s(K_s) = 0$. For $l \notin \{0, K_s\}$, $MSE_s(l) = \sigma_{Z_{s,l}}^2 = 1 - \alpha_s^{2l}$. Hence, the average MSE at each state s (called MSE_s) is,

$$MSE_s = \frac{\sum_{l=1}^{K_s-1} (1 - \alpha_s^{2l})}{K_s}. \quad (37)$$

Since the underlying MC is symmetric, with a small transition probability, averaging achievable points over different states of MC establishes the theorem. ■

VI. COMPARISON OF METHODS

In this section, we compare performances of different proposed sampling schemes against uniform sampling. In uniform sampling, the sampling rate is always in the form of $R = \frac{1}{N_s}$, where N_s is a positive integer. To be able to compare the performance of different methods with uniform sampling at different rates, we need to modify the uniform sampling setup to capture all possible sampling rates. To do this, for a given rate $R = \frac{1}{N_s}$ where N_s is not an integer number, we take the i^{th} sample at time $n_i = \lceil Ri \rceil$.

First, we compare sampling methods for deterministic signals (i.e., the Taylor Series Expansion (TSE) method, the Discrete Time Valued (DTV) method, and uniform sampling).

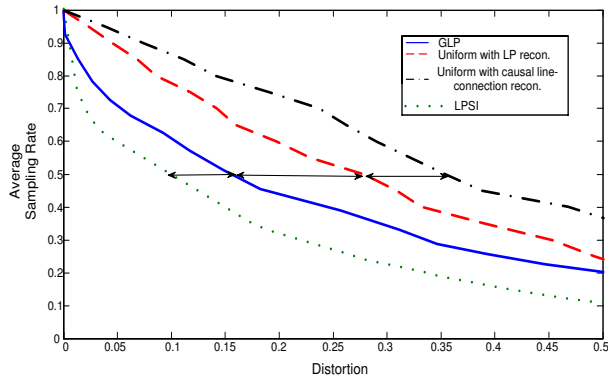


Fig. 6. A rate-distortion comparison between GLP, GA of FANS and uniform sampling with both lp and causal linear reconstructions.

We use the signal model described in Section II-A. For the TSE method, we use a setup explained in Section II-A. For the DTV method, we set $\epsilon_1 = \epsilon_2 = 1$, $T^{min} = 1$ and $T^{max} = 10$. The reconstruction method used for all methods is the first order linear filter. The distortion measure is the \mathcal{L}_1 norm of the error.

Figure 5 shows the average sampling rate versus distortion for different methods. As illustrated in this figure, for low distortion, TSE method, and for high distortion, DTV method outperform the others. For instance, at distortion 0.015, both TSE and DTV methods outperform uniform with a rate factor approximately 1.5. For fixed average sampling rate 0.6, TSE outperforms uniform by distortion factor approximately 3. Note that, non-smoothness of the curves is due to experimental artifacts.

Next, we compare performances of sampling methods for stochastic signals (i.e., the Generalized Linear Prediction (GLP) method, the Linear Prediction with Side Information (LPSI), and uniform sampling). We use the signal model described in (32). For GLP method, we use the setup explained in Section IV-C. For LPSI method, we use the setup explained in Section V. For uniform sampling, we use two reconstruction methods. One is to use a linear prediction filter with the same parameters as one of GLP method. The other one is a causal first order linear filter.

Figure 6 illustrates a rate-distortion function for these methods. Note that LPSI method outperforms the other methods. This is because it is a non-blind LAS and we have assumed we know the underlying signal model. Also, GLP method outperforms uniform sampling with both described reconstruction methods. For example, for a fixed average sampling rate 0.5, GLP, LPSI, and uniform with LP reconstruction outperforms uniform with linear filter reconstruction by distortion factor approximately 3.5, 2.2, and 1.3, respectively.

VII. CONCLUSION

In this paper, we introduced a family of locally adaptive sampling scheme, called LAS. LAS can be applied on both

deterministic (either band-limited or non-bandlimited) and stochastic signals. In this sampling family, time intervals between samples can be computed by using a function of previously taken samples. This function is called a sampling function. Hence, although it is a non-uniform sampling, we do not need to keep sampling times. The aim of this sampling scheme is to have the average sampling rate and the reconstruction error satisfy some requirements. We proposed four different schemes of LAS to explain its different properties: Taylor Series Expansion (TSE) and Discrete Time-Valued (DTV) methods for deterministic signals; Generalized Linear Prediction (GLP) and Linear Prediction with Side Information (LPSI) methods for stochastic signals. TSE and GLP were called blind methods since we have a general condition on the considered signal (bounded third derivative for TSE, and being locally stationary for GLP). However, DTV and LPSI methods are non-blind, because the sampling scheme is specifically designed for a known signal model.

VIII. ACKNOWLEDGMENT

Authors would like to thank Dr. Bruce Suter for pointing out references on change point analysis and active learning. This material is based upon work supported by the Air Force Office of Scientific Research (AFOSR) under award No. 016974-002.

REFERENCES

- [1] F. Marvasti, *Nonuniform sampling: Theory and Practice*. Plenum Publishers Co.
- [2] L. F. S. M. Qaisar and M. Renaudin, "Computationally efficient adaptive rate sampling and filtering," in *15th European Signal Processing Conference (EUSIPCO 07)*, Poznan, Poland, Sep. 2007, pp. 2139–2143.
- [3] D. Wei, "Sampling based on local bandwidth," in *Master Thesis, MIT*, 2007.
- [4] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "On stationary linear time-varying systems," *IEEE Transactions on Circuits and Systems*, vol. 29, no. 3, pp. 169–184, 1982.
- [5] K. Horiuchi, "Sampling principle for continuous signals with time-varying bands," *Information and Control*, vol. 13, no. 1, pp. 53–61, 1968.
- [6] M. Unser and J. Zerubia, "A generalized sampling theory without bandlimiting constraints," *IEEE Trans. Circuits and Systems*, vol. 2, pp. 959–969, 1998.
- [7] P. P. Vaidyanathan and B. Vrcelj, "On sampling theorems for non bandlimited signals," in *Proc. of the ICASSP*, Salt Lake City, May 2001, pp. 3897–3900.
- [8] T. Lai, "Sequential analysis: some classical problems and new challenges," *Statistica Sinica*, vol. 11, no. 2, pp. 303–350, 2001.
- [9] R. Castro and R. Nowak, "Upper and lower error bounds for active learning," in *The 44th Annual Allerton Conference on Communication, Control and Computing*, vol. 2, no. 2.1, 2006, p. 1.
- [10] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Annals of Statistics*, vol. 26, no. 1, pp. 1–47, 1998.
- [11] D. Donoho, S. Mallat, and R. von Sachs, "Estimating covariances of locally stationary processes: rates of convergence of best basis methods," *Statistics, Stanford University, Stanford, California, USA, Tech. Rep.*, 1998.
- [12] S. Haykin, *Adaptive filter theory*. Pearson Education India.