

# Gaussian Distortion-Rate Function under Sub-Nyquist Nonuniform Sampling

Alon Kipnis\*, Andrea J. Goldsmith\* and Yonina C. Eldar†

**Abstract**—A bound on the amount of distortion in the reconstruction of a stationary Gaussian process from its rate-limited samples is derived. The bound is based on a combined sampling and source coding problem in which a Gaussian stationary process is described from a compressed version of its values on an infinite discrete set. We show that the distortion in reconstruction cannot be lower than the distortion-rate function based on optimal uniform filter-bank sampling using a sufficient number of sampling branches. This can be seen as an extension of Landau’s theorem on a necessary condition for optimal recovery of a signal from its samples, in the sense that it describes both the error as a result of sub-sampling and the error incurred due to lossy compression of the samples.

## I. INTRODUCTION

The process of digitally storing information associated with the representation of a continuous time stochastic process involves the operations of sampling and quantization. The number of bits per second needed to describe an information source to some given average distortion is one of the basic problems in information theory, while the sampling and reconstruction of a stationary stochastic signal is a classic problem in signal processing. In this work we are interested in a combined problem where the source needs to be reconstructed from a rate-limited version of its samples.

The quantities of merit in this problem are the sampling set  $\Lambda \subset \mathbb{R}$ , the information rate  $R$  and the average distortion  $D$ . If the set  $\Lambda$  is dense enough such that the signal can be fully reconstructed from the samples, then the trade-off is described by the classic distortion-rate function (DRF) of the source. The other extreme is where the information rate  $R$  is infinite, in which case we are left with the reconstruction problem of an undersampled signal.

An expression for the indirect distortion rate function (iDRF) of a Gaussian stationary process given its rate-limited uniform samples was derived in [1] and [2]. In particular, a lower bound  $D_l(f_s, R)$  which depends only on the average sampling frequency  $f_s$  and the power spectral density (PSD) of the source was obtained there. This bound is obtained by waterfilling over the part of the PSD of the source with the highest energy. It equals zero if and only if the average sampling frequency of the samples is higher than the support of the PSD of the source, which is in accordance with Landau’s condition for perfect recovery of signals from their samples [3]. But the bound from [1] does more: it bounds from below the error that can be achieved under non-optimal sampling or when the samples are distorted by quantization or any form of lossy compression.

In this paper we consider an arbitrary discrete sampling set  $\Lambda = \{t_k, k \in \mathbb{Z}\} \subset \mathbb{R}$  and the iDRF of a Gaussian stationary source given its samples on this set. We show that the bound in [1] still holds under this non-uniform sampling setting, where we replace the average sampling frequency with the *Beurling density*  $d(\Lambda)$  of the sampling set. This establishes  $D_l(d(\Lambda), R)$  as a fundamental quantity in information theory and signal processing. In particular, this quantity bounds from below the distortion incurred due to source encoding based on the information in any uniform sampling scheme of a Gaussian stationary source.

## Background and Related Work

The reconstruction of a stationary process from its samples under a mean square error (MSE) criterion was tied in [4] to the problem of describing the auto-correlation function of the process from the values of the auto-correlation function on the sampling set. This establishes the sufficient condition given by the Shannon-Nyquist-Whitaker sampling theorem, that is uniform sampling above the Nyquist rate, as a sufficient condition for perfect reconstruction of a random process from its samples under expected MSE criterion. In the case of non-uniform sampling, the general notion of sampling frequency is replaced by the upper (lower) Beurling density  $d(\Lambda)$  of the sampling set  $\Lambda$ , defined by the largest (smallest) number of points in  $\Lambda$  contained within a single interval divided by the length of the interval, as that length tends to infinity. It follows from the landmark work of Landau [3] that a necessary and sufficient condition for perfect reconstruction of a random stationary signal from its non-uniform samples is that the spectral occupancy of the PSD of the signal, now termed the *Landau rate*, does not exceed the Beurling density of the sampling set.

Recently, the iDRF in reconstructing a Gaussian process given its uniform samples was solved in [1], [2]. This was obtained by reducing the combined sampling and source coding problem into a classical *indirect* source coding problem [5], first considered by [6]. More details on the results in [1] will be given in Section III. The combined sampling and source coding problem of this work can be seen as the dual to the channel coding problem with sampling at the receiver, considered in [7]. This last work was extended to non-uniform sampling in [8].

## Contribution

The main result of this paper is a lower bound  $D_l(d(\Lambda), R)$  on the iDRF of a Gaussian stationary process  $X(\cdot)$  given its samples  $\{X(t_k), t_k \in \Lambda\}$ . This bound depends only on the

\*Department of Electrical Engineering, Stanford University, CA

†Department of EE Technion, Israel Institute of Technology, Haifa, Israel

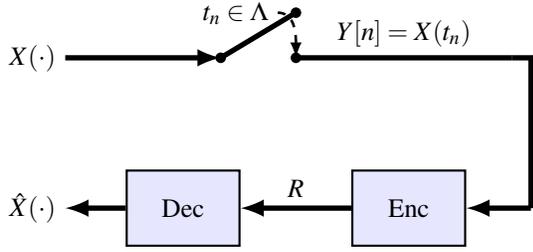


Fig. 1. Combined nonuniform sampling and source coding model.

PSD of the source and the Beurling density  $d(\Lambda)$  of the sampling set. As the information rate  $R$  goes to infinity, the iDRF approaches the optimal MSE in the reconstruction of the process from the samples. This leads to a lower bound on the MSE in estimating a Gaussian stationary process from its uniform samples. The bound  $D_l(d(\Lambda), R)$  extends Landau's characterization of sampling sets for a signal in three aspects:

- 1) It describes the minimal error if the sampling set is not dense enough.
- 2) It describes the minimal amount of excess distortion in the reconstruction due to lossy compression of the samples.
- 3) It considers the case where the process undergoes a general linear time-varying processing before nonuniform sampling.

The rest of this paper is organized as follows: in Section II we present the combined non-uniform sampling and source coding problem. In Section III we review recent results on the iDRF of processes from sub-Nyquist uniform sampling and a pre-sampling filter. In Section IV we prove our main results. An example for the case of a Gauss-Markov process is given in Section V. Concluding remarks are provided in Section VI.

## II. PRELIMINARIES

### A. System Model

We consider a combined sampling and source coding problem as depicted in Fig. 1. The source  $X(\cdot) = \{X(t), t \in \mathbb{R}\}$  is a real Gaussian stationary process with PSD

$$S_X(f) \triangleq \int_{-\infty}^{\infty} \mathbb{E}[X(t+\tau)X(t)] e^{-2\pi i f \tau} d\tau,$$

and variance  $\sigma_X^2 \triangleq \int_{-\infty}^{\infty} S_X(f) df < \infty$ . The sampler receives the process  $X(\cdot)$  as an input and produces a discrete-time process  $Y[n] = X(t_n)$ ,  $t_n \in \Lambda$ . Throughout this paper, we use round brackets and square brackets to distinguish between continuous-time and discrete-time processes.

The fidelity criterion is defined by the MSE between the original source and its reconstruction  $\hat{X}(\cdot) = \{\hat{X}(t), t \in \mathbb{R}\}$ , namely

$$\mathbb{E}d(\hat{X}(\cdot), X(\cdot)) \triangleq \mathbb{E}\|\hat{X}(\cdot) - X(\cdot)\|^2 \quad (1)$$

where  $\|X(\cdot)\|$  is the time-averaged  $L_2$  norm<sup>1</sup> of the process  $X(\cdot)$ , defined by

$$\|X(\cdot)\|^2 \triangleq \limsup_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mathbb{E}(X^2(t)) dt.$$

We denote by  $mmse(\Lambda)$  the optimal MSE estimation error of  $X(\cdot)$  from the process  $Y[\cdot]$ ,

$$mmse(\Lambda) = \|X(t) - \mathbb{E}[X(t)|Y_\Lambda[\cdot]]\|^2.$$

The *Landau rate* of the process  $X(\cdot)$  is defined to be the Lebesgue measure of the support of  $S_X(f)$ , and will be denoted by  $f_L(X)$ .

### Non-uniform sampling

For a discrete set  $\Lambda \subset \mathbb{R}$ , denote by  $\bar{n}_r(\Lambda)$  the maximal number of elements of  $\Lambda$  that belong to a single interval of length  $r$ , namely

$$\bar{n}_r(\Lambda) = \sup_{u \in \mathbb{R}} |\{t \in \Lambda, t \in u + [0, r)\}|.$$

Similarly, define

$$\underline{n}_r(\Lambda) = \inf_{u \in \mathbb{R}} |\{t \in \Lambda, t \in u + [0, r)\}|.$$

The lower and upper Beurling densities of  $\Lambda$  are respectively defined by

$$\underline{d}(\Lambda) = \lim_{r \rightarrow \infty} \frac{\underline{n}_r(\Lambda)}{r},$$

and

$$\bar{d}(\Lambda) = \lim_{r \rightarrow \infty} \frac{\bar{n}_r(\Lambda)}{r}.$$

If  $\bar{d}(\Lambda) = \underline{d}(\Lambda)$ , we say that the set  $\Lambda$  has *Beurling density*  $d(\Lambda) = \bar{d}(\Lambda)$ . Throughout the rest of this paper, only sampling sets  $\Lambda$  with existing Beurling density will be considered.

Given a nonuniform sampling set  $\Lambda$ , whether  $X(\cdot)$  is recoverable from the nonuniform samples sequence  $Y_\Lambda[\cdot]$  is determined by the completeness of the exponential functions  $\mathcal{E}_\Lambda \triangleq \{e^{2\pi i t_n f}, t_n \in \Lambda\}$  in  $\mathbf{L}_2(\mathbb{R})$ . In particular, when  $\Lambda = \mathbb{Z}/f_s = \{n/f_s, n \in \mathbb{Z}\}$ , the set  $\mathcal{E}_\Lambda$  forms a basis of  $\mathbf{L}_2(-f_s/2, f_s/2)$  by the Shannon-Nyquist sampling theorem. The fundamental sampling rate necessary for perfect reconstruction of  $X(\cdot)$  from  $Y_\Lambda[\cdot]$  has been given by Landau [3]. It follows from [3] that perfect reconstruction of  $X(\cdot)$  from  $Y_\Lambda[\cdot]$  is possible if and only if the Beurling density of  $\Lambda$  is higher than the Landau frequency of  $X(\cdot)$ , i.e.,

$$mmse(\Lambda) = 0 \Leftrightarrow d(\Lambda) \geq f_L(X). \quad (2)$$

As in [3], this work is concerned only with the existence of estimators that achieve minimal or zero distortion and we will not be concerned with obtaining a technique for such estimation.

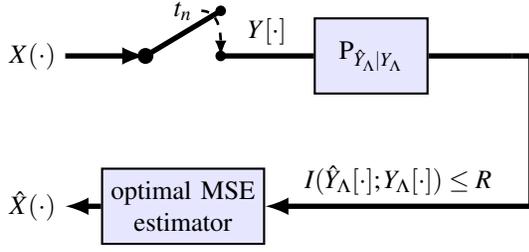


Fig. 2. Informational rate-distortion representation

### Problem Statement

Given a sampling set  $\Lambda$ , we consider the minimum possible distortion that can be attained between  $X(\cdot)$  and  $\hat{X}(\cdot)$ . Classical results in rate-distortion theory [5], [9] imply that this problem has the informational rate-distortion characterization depicted in Fig. 2, in which  $P_{\hat{Y}_\Lambda|Y_\Lambda}$  denotes a ‘test channel’ between  $Y_\Lambda[\cdot]$  and  $\hat{Y}_\Lambda[\cdot] = \{\hat{Y}_\Lambda[n], n \in \mathbb{Z}\}$ , and the reconstruction process  $\hat{X}(\cdot)$  is obtained from the process  $\hat{Y}_\Lambda[\cdot]$  which represents a compressed or a quantized version of  $Y_\Lambda[\cdot]$ . For  $T > 0$  define  $\Lambda_T \triangleq \Lambda \cap [-T, T]$ , and denote by  $\mathcal{A}(\Lambda_T, R)$  the set of all random mappings  $Y_\Lambda[\cdot] \rightarrow \hat{Y}_\Lambda[\cdot] \rightarrow \hat{X}(\cdot)$  measurable with respect to the  $\sigma$ -algebra generated by  $Y_{\Lambda_T}[\cdot]$ , such that the mutual information rate between  $Y_\Lambda[\cdot]$  and  $\hat{Y}_\Lambda[\cdot]$ , defined by

$$I(\hat{Y}_\Lambda[\cdot]; Y_\Lambda[\cdot]) \triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} I(\hat{Y}_{\Lambda_T}; Y_{\Lambda_T}),$$

is limited to  $R$  bits per time unit. The iDRF of  $X(\cdot)$  given  $Y_\Lambda[\cdot]$  is defined by

$$D_\Lambda(R) = \lim_{T \rightarrow \infty} D_{\Lambda_T}(R), \quad (3)$$

where

$$D_{\Lambda_T}(R) = \inf_{\mathcal{A}(\Lambda_T, R)} \mathbb{E}d(\hat{X}(\cdot), X(\cdot)). \quad (4)$$

The facts below follows from the definition of  $D_\Lambda(R)$ :

*Proposition 2.1:* The following holds for all discrete sets  $\Lambda \subset \mathbb{R}$  with Beurling density  $d(\Lambda)$  and  $R \geq 0$ :

- (i)  $D_\Lambda(R) \leq \|X(\cdot)\|^2 = \sigma_X^2$ .
- (ii)  $D_\Lambda(R) \geq mmse(\Lambda)$ , where  $mmse(\Lambda)$  is the minimal MSE error in estimating  $X(\cdot)$  from the samples  $Y_\Lambda[\cdot]$ , namely

$$mmse(\Lambda) = \|X(\cdot) - \tilde{X}(\cdot)\|^2,$$

where

$$\tilde{X}(t) = \mathbb{E}[X(t)|Y_\Lambda[\cdot]]$$

is the minimal MSE estimator.

- (iii) As  $R \rightarrow \infty$ ,  $D_\Lambda(R) \rightarrow mmse(\Lambda)$ .
- (iv) If  $d(\Lambda)$  is bigger than the total measure of the support of  $S_X(f)$ , then  $D_\Lambda(R) = D_X(R)$ , where  $D_X(R)$  is the DRF of the Gaussian stationary process  $X(\cdot)$ .

<sup>1</sup>In our notation, we allow  $\|\hat{X}(\cdot)\|$  to take infinite value if  $\hat{X}(\cdot)$  is not in  $L_2$ .

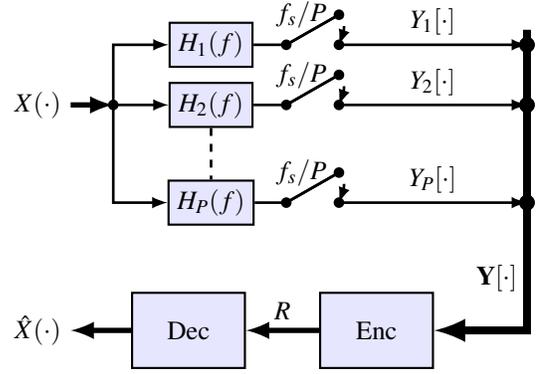


Fig. 3. Combined source coding and uniform multi-branch uniform sampling model.

*Proof:* (i) follows by taking the reconstruction process to be  $\hat{X}(t) = 0$  for all  $t \in \mathbb{R}$ . (ii) and (iii) are due to lossy compression of the samples when  $R$  is finite. (iv) follows from the celebrated result of Landau [3] on the condition for perfect reconstruction of a signal from its non-uniform samples (2).

### III. DISTORTION-RATE FUNCTION UNDER UNIFORM SAMPLING AND PRE-SAMPLING FILTERING

A special case of the set  $\Lambda$  is when there exists  $T \in \mathbb{R}$  such that

$$\Lambda = \Lambda + T \triangleq \{t + T, t \in \Lambda\}, \quad (5)$$

i.e., the set  $\Lambda$  is invariant under shifts by integer multiples of  $T$ . Denote by  $T_\Lambda$  the smallest  $T$  that satisfies (5) and by  $N$  the number of elements of  $\Lambda$  in an interval of length  $T_\Lambda$ . It is easy to verify that the Beurling density of  $\Lambda$  exists and is given by  $d(\Lambda) = \frac{N}{T_\Lambda}$ . A more restricted special case is when the sampling set  $\Lambda$  is uniform:  $\Lambda = \mathbb{Z}T = \{nT, n \in \mathbb{Z}\}$ . This corresponds to a single branch of uniform sampling at frequency  $d(\Lambda) = 1/T$ .

The iDRF under the multi-branch sampling and pre-sample filtering scheme given in Fig. 3 was considered in [1]. The main result there is an expression for the iDRF of  $X(\cdot)$  given the samples  $\mathbf{Y}[\cdot] = (Y_1[\cdot], \dots, Y_P[\cdot])$  obtained using  $P$  uniform sampling branches as in Fig. 3. The  $p^{\text{th}}$  branch is composed of a pre-sampling linear time-invariant (LTI) filter with frequency response  $H_p(f)$  and a uniform sampler at frequency  $f_s/P$ . Note that this setting includes in particular the case where the sampling points constitute a uniform grid or a finite union of uniform grids with the same fundamental gap, which can be achieved by taking each of the pre-sampling filters to be time delay systems, e.g.  $H_p(f) = e^{2\pi i f t_p}$ . This fact will play a crucial role in proving our main theorem. Optimization carried in [1] of  $D(P, f_s, R)$  over the pre-sampling filters leads to an optimal choice of the filters  $H_1^*(f), \dots, H_P^*(f)$ . The resulting distortion  $D^*(P, f_s, R)$  depends only on the number of sampling branches  $P$ , the effective sampling frequency  $f_s$  and the PSD of the source  $S_X(f)$ . As the number of sampling branches  $P$  goes to infinity,  $D^*(P, f_s, R)$  converges (not necessarily monotonically) to

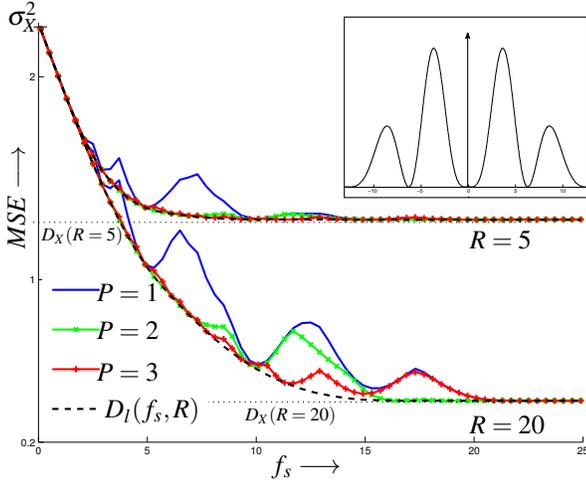


Fig. 4.  $D^*(P, f_s, R)$  as a function of  $f_s$  for two fixed values of the source coding rate  $R$  and source with spectrum given in the small frame. As  $f_s$  increases,  $D^*(P, f_s, R)$  converges to the DRF of the source  $D_X(R)$  (with equality guaranteed for  $f_s$  bigger than the Nyquist rate). As  $P$  increases,  $D^*(P, f_s, R)$  converges to the bound  $D_I(f_s, R)$ .

a lower value,  $D_I(f_s, R)$ , given by

$$D_I(f_s, R) = \sigma_X^2 - \int_{F^*} [S_X(f) - \theta]^+ df, \quad (6)$$

where  $\theta$  is determined by

$$R = \frac{1}{2} \int_{F^*} \log^+ [S_X(f)/\theta] df,$$

and  $F^*$  is the set that maximizes  $\int_F S_X(f) df$  over all measurable subsets  $F \subset \mathbb{R}$  with Lebesgue measure not exceeding  $f_s$ . The function  $D^*(P, f_s, R)$  is illustrated in Fig.4 for various values of  $P$  and  $R$  as a function of the sampling frequency.

As we take  $R \rightarrow \infty$ , (6) gives the following bound on the optimal MSE in estimating  $X(\cdot)$  from  $\mathbf{Y}[\cdot]$ :

$$mmse_I(f_s) = \sigma_X^2 - \int_{\mathbb{R} \setminus F^*} S_X(f) df = \int_{\mathbb{R} \setminus F^*} S_X(f) df. \quad (7)$$

Note that (7) coincides with the expression for the optimal MSE in estimating  $X(\cdot)$  from its filtered version with maximal spectral occupancy of measure  $f_s$ .

*Example 3.1 ( $D_\Lambda(R)$  in uniform sampling):* Under the setting of Fig. 1 and  $\Lambda$  of the form  $\Lambda = \{T_s n, n \in \mathbb{N}\}$  for some  $T_s > 0$ , the main result in [2] leads to the following expression for  $D_\Lambda(R)$ :

$$R_\theta = \frac{1}{2} \int_{-\infty}^{\infty} \log^+ [J(f)/\theta] df, \quad (8a)$$

$$D_\theta(R) = \sigma_X^2 - \int_{-\infty}^{\infty} [J(f) - \theta]^+ df, \quad (8b)$$

where

$$J(f) \triangleq \frac{S_X^2(f)}{\sum_{k \in \mathbb{Z}} S_X(f - kf_s)},$$

and  $f_s = 1/T_s$ . The expression (8) has the waterfilling interpretation described in Fig. 5. If the support of  $S_X(f)$

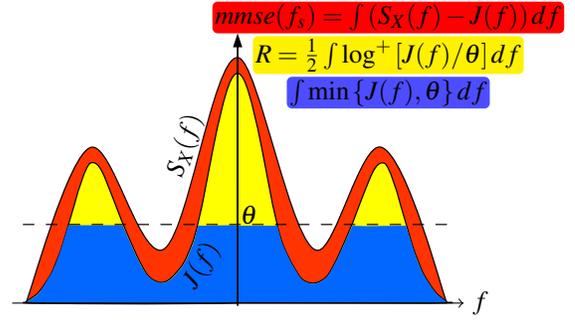


Fig. 5. Water-filling interpretation of (8). The overall distortion is the sum of the red and the blue areas.

is contained within the interval  $(-f_s/2, f_s/2)$ , then  $J(f) = S_X(f)$  and (8) reduces to Shannon-Kolmogorv-Pinsker reverse water-filling [10].

The function  $D_I(f_s, R)$  provides a lower bound on the iDRF of a Gaussian process given its samples using filter bank uniform sampling. This includes in particular a bound on  $D_\Lambda(R)$  where  $\Lambda$  is of the form

$$\Lambda = \bigcup_{p=1}^P (\mathbb{Z}/f_s + t_p), \quad (9)$$

where  $t_1, \dots, t_P \in \mathbb{R}$  and  $f_s > 0$ . This bound depends only on the PSD  $S_X(f)$  and the sampling frequency  $f_s$ . This bound is achievable using multi-branch sampling with LTI pre-sampling filters in each branch as given in Fig. 3, if we allow enough sampling branches  $P$  in our system. In the next section we will show that this bound still holds under the nonuniform sampling setting of Fig. 1, if we replace the sampling frequency with the Beurling density of the sampling set.

#### IV. MAIN RESULT: $D_\Lambda(R) \geq D_I(d(\Lambda), R)$

We first show that the setting in [1, Thm. 14] which leads to  $D(P, f_s, R)$ , includes a periodic nonuniform sampling set as a special case.

*Theorem 4.1 (periodic sampling):* Let  $\Lambda \subset \mathbb{R}$  be a periodic sampling set with period  $T$ , namely  $\Lambda = \Lambda + T\mathbb{Z}$ . The indirect distortion-rate function of  $X(\cdot)$  given the samples  $Y_\Lambda[\cdot] = \{X(t_n), t_n \in \Lambda\}$  satisfies

$$D_\Lambda(R) \geq D_I(d(\Lambda), R), \quad (10)$$

where  $d(\Lambda)$  is the Beurling density of  $\Lambda$ , and  $D_I(f_s, R)$  is the lower bound (7) on the indirect distortion-rate function of  $X(\cdot)$  given its sub-Nyquist uniform samples at rate  $f_s$ .

*Proof:* Since any process with finite variance can be described by its uniform samples at high enough frequency with arbitrarily small error (even if the process is not band-limited), we conclude that  $D_I(f_s, R) \rightarrow D_X(R)$  as  $f_s$  goes to infinity. This implies that if  $d(\Lambda)$  is infinite, then  $D_I(d(\Lambda), R)$  equals the distortion-rate function  $D_X(R)$  of the process  $X(\cdot)$ , which implies  $D_\Lambda(R) \geq D^*(d(\Lambda), R) = D_X(R)$ . If  $d(\Lambda) = 0$ ,

then  $D_l(0, R) = \sigma_X^2 = D_\Lambda(R)$ . We will further assume that  $d(\Lambda)$  is finite and positive, which in particular implies that each interval of finite length contains at most a finite number of points of  $\Lambda$ . In fact, such a set has the general form (9).

Since  $\Lambda$  is periodic, the number of points of  $\Lambda$  in each interval of length  $T$  is constant and is equal to the Beurling density  $d(\Lambda)$  of the set  $\Lambda$  times the length of that interval. Let  $M$  be the number of points in  $\Lambda$  in an interval of length  $T$ , namely  $M = Td(\Lambda)$ . Denote by  $t_1, \dots, t_M$  the members of  $\Lambda$  inside the interval  $[0, T)$ . Since  $\Lambda + T = \Lambda$ , sampling with  $\Lambda$  is equivalent to sampling with  $M$  sampling branches each of sampling frequency  $1/T$  and a pre-sampling filter  $H_m(f) = e^{2\pi i f t_m}$ ,  $m = 1, \dots, M$ . We conclude that  $D_\Lambda(R) = D(M, d(\Lambda), R)$  where  $D(M, d(\Lambda), R)$  is the indirect distortion rate function of  $X(\cdot)$  given its multi-branch sampling which is given by [1, Thm. 14]. Since for any set of filters we have  $D_l(d(\Lambda), R) \leq D(M, d(\Lambda), R)$ , (10) follows.  $\square$

The general case follows from Theorem 4.1 by an approximation argument:

**Theorem 4.2 (main result):** Let  $\Lambda$  be a discrete subset of  $\mathbb{R}$  with finite Beurling density  $d(\Lambda)$ . The indirect distortion-rate function of the Gaussian stationary process  $X(\cdot)$  given the samples  $Y_\Lambda[\cdot] = \{X(t_n), t_n \in \Lambda\}$  satisfies

$$D_\Lambda(R) \geq D_l(d(\Lambda), R). \quad (11)$$

*Proof:* Fix  $r > 0$  and denote  $I_r = [-r/2, r/2]$ . Define  $\Lambda_r \triangleq \Lambda \cap I_r$ . In addition, define the set  $\overline{\Lambda}_r$  to consist of the points in  $\Lambda \cap I_r$  and all their shifts by integers multiples of  $r$ , namely

$$\overline{\Lambda}_r = \Lambda_r + r\mathbb{Z}.$$

It can be verified that  $\overline{\Lambda}_r$  is a periodic sampling set with period  $r$  and Beurling density  $d(\Lambda_r) = \frac{|\Lambda_r|}{r}$ . From Theorem 4.1 we have that  $D_{\overline{\Lambda}_r}(R) \geq D_l(d(\Lambda_r), R)$ , and since  $\Lambda_r \subset \overline{\Lambda}_r$  we conclude that

$$D_{\Lambda_r}(R) \geq D_{\overline{\Lambda}_r}(R) \geq D_l(d(\Lambda_r), R). \quad (12)$$

Since

$$\frac{\overline{n}_r(\Lambda)}{r} \geq \frac{|\Lambda_r|}{r} \geq \frac{n_r(\Lambda)}{r},$$

and both sides converge to  $d(\Lambda)$  as  $r \rightarrow \infty$ , so does  $\frac{|\Lambda_r|}{r}$ . Since  $D_l$  is continuous in its first argument, in order to complete the proof it is left to show that  $D_{\Lambda_r}(R)$  converges to  $D_\Lambda(R)$ . This follows from the definition of  $D_\Lambda(R)$  since

$$D_\Lambda(R) = \lim_{r \rightarrow \infty} \inf_{\mathcal{A}(\Lambda_r, R)} \mathbb{E}d(\hat{X}(\cdot), X(\cdot)) = \lim_{r \rightarrow \infty} D_{\Lambda_r}(R). \quad \square$$

#### Extended Model with Pre-processing

We now wish to extend our system model to consider a pre-sampling filter as described in Fig. 6, where  $q(t, \tau)$  is the system response at time  $\tau$  to an impulse at time  $t$ , namely, the process  $Y_\Lambda[\cdot]$  is given by

$$Y_{\Lambda, q}[n] = Z(t_n) \triangleq \int_{-\infty}^{\infty} X(\tau) q(t_n, t_n - \tau) d\tau, \quad t_n \in \Lambda, \quad n \in \mathbb{Z}, \quad (13)$$

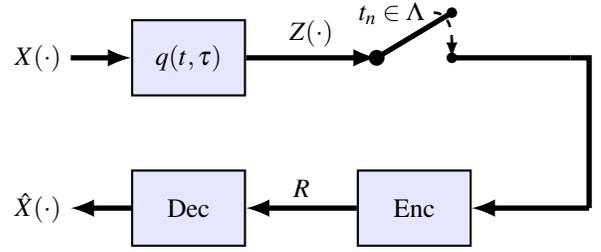


Fig. 6. Extended system model

where we assume that for all  $t_n \in \Lambda$ , the integral in (13) is finite for almost all realizations of  $X(\cdot)$ .  $\mathcal{A}(\Lambda_T, R)$  and the iDRF of  $X(\cdot)$  given  $Y_{\Lambda, q}[\cdot]$  are now defined in the same way as in Section II using  $Y_{\Lambda, q}[\cdot]$ .

In [8, Prop. 1] it was shown that any multi-branch uniform sampling scheme with possibly different time-varying pre-sampling filters at every branch is equivalent to the single branch sampling system of Fig. 6. In particular, this means that the system in Fig. 6 includes the system of Fig. 3 as a special case.

We have the following extension of Theorem 4.2:

**Theorem 4.3:** Let  $\Lambda$  be a discrete subset of  $\mathbb{R}$  with finite Beurling density  $d(\Lambda)$ . The indirect distortion-rate function of the Gaussian stationary process  $X(\cdot)$  given the samples  $Y_{\Lambda, q}[\cdot] = \{Z(t_n), t_n \in \Lambda\}$  satisfies

$$D_{\Lambda, q}(R) \geq D_l(d(\Lambda), R).$$

*Proof:* The proof goes along a similar line as the proof of Theorem 4.2. We first consider the case of a periodic sampling set  $\Lambda = \Lambda + T$  which has Beurling density  $d(\Lambda) = MT$ , where  $M$  is the number of points in  $\Lambda$  in an interval of length  $T$ . We also add the assumption (that will later be removed) that  $q(t + Tk, Tk + \tau) = q(t, \tau)$  for all  $k \in \mathbb{Z}$ , i.e.  $q(t, \tau)$  is periodic in  $t$  and  $\tau$  with period  $T$ . Denote by  $t_0, \dots, t_{M-1}$  the members of  $\Lambda$  inside the interval  $[0, T)$ . By the periodicity of  $\Lambda$ ,  $t_{m+Mk} = t_m + Tk$  for all  $m = 0, \dots, M-1$  and  $k \in \mathbb{Z}$ . For  $n = m + kM$  we have,

$$\begin{aligned} Y_{\Lambda, q}[n] &= Z(t_{m+Mk}) = \int_{-\infty}^{\infty} q(t_m + Tk, t_m + Tk - \tau) X(\tau) d\tau \\ &= \int_{-\infty}^{\infty} q(t_m, t_m - \tau) X(\tau) d\tau. \end{aligned}$$

This shows that sampling with the set  $\Lambda$  and the pre-processing system  $q(t, \tau)$  is equivalent to  $M$  uniform sampling branches each of sampling frequency  $1/T$  and pre-sampling filter

$$H_m(f) \triangleq e^{2\pi i f t_m} Q_t(f),$$

where for  $t \in \mathbb{R}$ ,  $Q_t(f)$  is the Fourier transform of  $q(t, \tau)$  with respect to  $\tau$ . The iDRF of  $X(\cdot)$  given  $Y[\cdot]$  is given in [1] and denoted  $D(M, d(\Lambda), R)$ . It also follows from [1] that

$$D(M, d(\Lambda), R) \geq D_l(d(\Lambda), R),$$

which completes the proof for periodic sampling with a periodic pre-processing system  $q(t, \tau)$ . The general case follows along the same line as in the proof of Theorem 4.2: We first

consider the sampling set  $\Lambda_r = \Lambda \cap I_r$  where  $I_r = [-r/2, r/2]$ , and its periodic extension  $\overline{\Lambda}_r = \Lambda_r + r\mathbb{Z}$  for a given  $r > 0$ . We also extend  $q(t, \tau)$  periodically as

$$\overline{q}_r(t, t - \tau) \triangleq q(t|_r, t|_r - \tau),$$

where we used the notation  $t|_r$  to denote  $t$  modulo the grid  $\frac{1}{2} + r\mathbb{Z}$ . We denote by  $Y_{\overline{\Lambda}_r, \overline{q}_r}[\cdot]$  the sampling process obtained by sampling with the set  $\overline{\Lambda}_r$  and the pre-processing system  $\overline{q}_r(t, t - \tau)$ , i.e.,

$$\begin{aligned} Y_{\overline{\Lambda}_r, \overline{q}_r}[n] &= \int_{-\infty}^{\infty} \overline{q}_r(t_n, t_n - \tau) X(\tau) d\tau \\ &= \begin{cases} \int_{-\infty}^{\infty} q(t_n, t_n - \tau) X(\tau) d\tau, & t_n \in \Lambda_r, \\ \int_{-\infty}^{\infty} q(t_n|_r, t_n|_r - \tau) X(\tau) d\tau, & t_n \in \Lambda_r + \mathbb{Z}r \setminus I_r, \end{cases} \end{aligned} \quad (14)$$

for all  $n \in \mathbb{Z}$ . Note that the samples  $Y_{\overline{\Lambda}_r, \overline{q}_r}[\cdot]$  were obtained using a periodic sampling set with a periodic pre-processing system with period  $r$ . By the first part of the proof we have

$$D_{\overline{\Lambda}_r, \overline{q}_r}(R) \geq D_l(d(\overline{\Lambda}_r), R) = D_l\left(\frac{|\Lambda_r|}{r}, R\right). \quad (15)$$

From (14) we conclude that  $Y_{\Lambda_r, q}[\cdot] \subset Y_{\overline{\Lambda}_r, \overline{q}_r}[\cdot]$ , which implies

$$D_{\Lambda_r, q}(R) \geq D_{\overline{\Lambda}_r, \overline{q}_r}(R) \geq D_l\left(\frac{|\Lambda_r|}{r}, R\right).$$

The proof is completed by noting that as  $r$  goes to infinity,  $D_{\Lambda_r, q}(R)$  converges to  $D_{\Lambda, q}(R)$  by definition,  $\frac{|\Lambda_r|}{r}$  converges to  $d(\Lambda)$  and  $D_l$  is continuous in its first argument.  $\square$

### Discussion

Theorem 4.2 establishes  $D_l(f_d, R)$  as a fundamental bound on the iDRF of a Gaussian stationary process given its samples. It extends the celebrated result of Landau [3] in two aspects. First, it bounds the MSE in the reconstruction of a process when the conditions for perfect reconstruction are not met. Second, it describes the distortion as a result of rate-limited descriptions of the samples. In addition,  $D_l(f_d, R)$  is monotone decreasing in its first argument  $f_d$ . Monotonicity in the sampling density is a desired property since increasing the sampling density is intuitively associated with increasing the information available about the process. This is in contrast to the case of uniform sampling where the functions  $D^*(P, f_s, R)$  and  $D^*(P, f_s, R)$  derived in [1] are not necessarily monotone in the sampling frequency  $f_s$ . A sampling scheme that achieves the bound but uses filter-bank sampling was considered in [1]. Theorem 4.3 shows that the bound  $D_l$  still holds even if we allow linear pre-processing before sampling. This implies that Landau's condition for perfect reconstruction of a signal from its nonuniform samples (2) cannot be improved by linear pre-processing, and linear time-varying pre-processing cannot improve distortion over multi-branch sampling with optimal LTI pre-sampling filters.

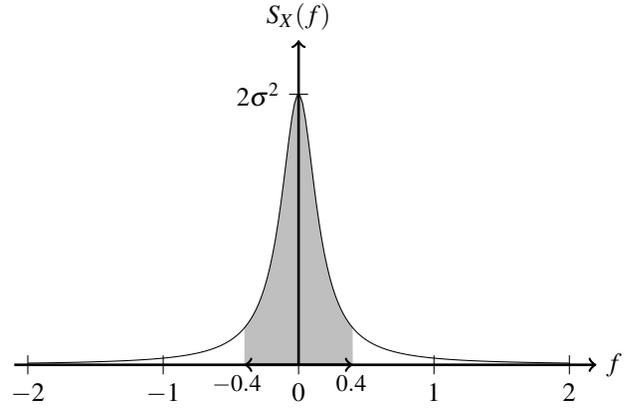


Fig. 7. PSD of a Gauss-Markov process and the set  $F^*$  for  $d(\Lambda) = 0.8$ .

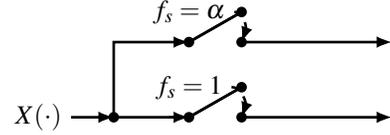


Fig. 8. Sampling system that implements the sampling set  $\Lambda(\alpha) = \mathbb{Z} \cup \mathbb{Z}/\alpha$ .

### V. EXAMPLE: MARKOV-GAUSS SOURCE

In this example we compare the optimal MSE in estimating a Gauss-Markov process from its non-uniform samples to the bound

$$mmse_l(d(\Lambda)) \triangleq \lim_{R \rightarrow \infty} D_l(d(\Lambda), R).$$

Consider the Gaussian stationary process with PSD

$$S_X(f) = \frac{2\sigma^2}{(2\pi f)^2 + 1}. \quad (16)$$

This PSD corresponds to the auto-correlation function

$$K_X(\tau) = \sigma^2 e^{-|\tau|}.$$

Note that the support of (16) is the entire real line, and therefore the Landau rate of  $X(\cdot)$  is infinite. This means that  $mmse(\Lambda) > 0$  and  $mmse_l(d(\Lambda)) > 0$  for any sampling set  $\Lambda$ .

Since  $S_X(f)$  is unimodal, we have that  $F^* = [-d(\Lambda)/2, d(\Lambda)/2]$  for any sampling set  $\Lambda \subset \mathbb{R}$  (see Fig. 7).

The optimal MSE bound (7) is found to be

$$mmse_l(d(\Lambda)) = \sigma^2 \left( 1 - \frac{2}{\pi} \tan^{-1}(2\pi d(\Lambda)) \right), \quad (17)$$

which is given by the white area in Fig. 7.

We consider now the non-uniform sampling set

$$\Lambda(\alpha) = \mathbb{Z} \cup \mathbb{Z}/\alpha,$$

where  $\alpha$  is a real positive number. This sampling set can be implemented by two sampling branches as shown in Fig. 8. For irrational  $\alpha$ , the Beurling density of  $\Lambda$  equals  $1 + \alpha$  and it is less than  $1 + \alpha$  for rational  $\alpha$ . Since a Gaussian process with the PSD (16) is Markovian, for each time

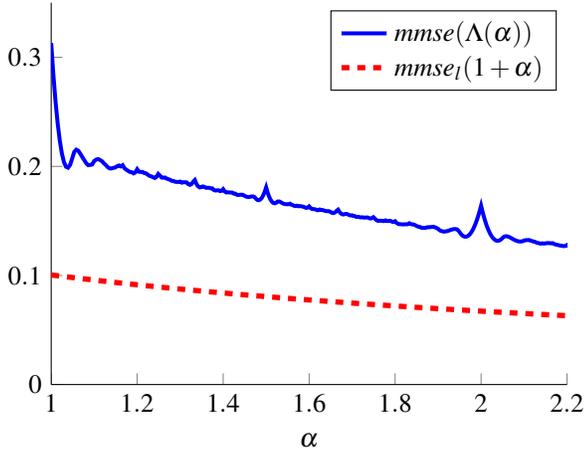


Fig. 9. The minimal estimation error of the Gauss-Markov process  $X(\cdot)$  from the sampling set  $\Lambda(\alpha) = \mathbb{Z} \cup \mathbb{Z}/\alpha$ . This is compared to the bound  $mmse_l(1+\alpha)$  given by (17). Note that for  $\alpha$  rational the true Beurling density of  $\Lambda(\alpha)$  is smaller than  $1+\alpha$ . For example  $d(\Lambda(\alpha)) = 1$  at  $\alpha = 1$  and the MSE in this case is the error in reconstruction of  $X(\cdot)$  from its samples on the integers.

instant  $t \in \mathbb{R}$  an estimator of  $X(t)$  from the samples  $Y_{\Lambda(\alpha)}[\cdot]$  only depends on the two closest points in  $\Lambda(\alpha)$  below and above  $t$ , respectively, namely  $n_t^- = \max \Lambda(\alpha) \cap (-\infty, t]$  and  $n_t^+ = \min \Lambda(\alpha) \cap [t, \infty)$ . Since  $X(\cdot)$  is Gaussian, the optimal MSE estimator of  $X(t)$  from  $Y_{\Lambda(\alpha)}[\cdot]$  is given by the linear projection of  $X(t)$  onto the sub-space spanned by  $X(n_t^+)$  and  $X(n_t^-)$ . This leads to the following expression for the instantaneous minimal estimation error

$$mmse(t, \Lambda(\alpha)) = \sigma^2 - \begin{pmatrix} K_X(t-t^+) & K_X(t-t^-) \end{pmatrix} \times \begin{pmatrix} \sigma^2 & K_X(t^+ - t^-) \\ K_X(t^- - t^+) & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} K_X(t^+ - t) \\ K_X(t^- - t) \end{pmatrix},$$

which can be evaluated numerically. The true average error  $mmse(\Lambda(\alpha))$  can be computed by averaging  $mmse(t, \Lambda(\alpha))$  over a large time interval. The plot in Fig. 9 compares  $mmse(\Lambda(\alpha))$  to the bound (17) evaluated at  $1+\alpha$  for various values of  $\alpha$ .

Note that the bound  $mmse_l(1+\alpha)$  can be achieved using a single branch uniform sampler at frequency  $f_s = 1+\alpha$  and an anti-aliasing filter. It is not clear, however, what can be achieved using the sampling set  $\Lambda(\alpha)$  and a pre-sampling filter, since pre-processing may change the Markov property of the process and make it hard to evaluate the MSE.

## VI. CONCLUSIONS

We derived a lower bound  $D_l(d(\Lambda), R)$  on the indirect distortion-rate function of a Gaussian stationary source  $X(\cdot)$  given its samples over an arbitrary set  $\{X(t_k), t_k \in \Lambda\}$ . This bound is given only in terms of the power spectral density of the source and the upper Beurling density of the sampling set  $\Lambda$ , and therefore describes a fundamental quantity in information theory and signal processing associated with any Gaussian stationary source. It was shown that this bound still holds even if we allow linear pre-processing before

nonuniform sampling. As an example, we computed the bound on the MSE for the case of a Gauss-Markov process. This bound was compared to the true MSE of this process in its reconstruction from its samples on a non-uniform sampling grid.

In previous work [1] we have shown that the proposed bound can be achieved using filter-bank uniform sampling. We leave open the question on how to choose the sampling set  $\Lambda$  for a prescribed Beurling density, to achieve minimal distortion.

## ACKNOWLEDGMENT

This work was supported in part by the NSF Center for Science of Information (CSoI) under grant CCF-0939370 and BSF Transformative Science Grant 2010505.

## REFERENCES

- [1] A. Kipnis, A. J. Goldsmith, T. Weissman, and Y. C. Eldar, "Distortion-Rate Function of Sub-Nyquist Sampled Gaussian Sources," May 2014. [Online]. Available: <http://arxiv.org/abs/1405.5329>
- [2] A. Kipnis, A. Goldsmith, T. Weissman, and Y. Eldar, "Rate-distortion function of sub-nyquist sampled gaussian sources corrupted by noise," in *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013.
- [3] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, 1967. [Online]. Available: <http://dx.doi.org/10.1007/BF02395039>
- [4] A. Balakrishnan, "A note on the sampling principle for continuous signals," *Information Theory, IRE Transactions on*, vol. 3, no. 2, pp. 143–146, June 1957.
- [5] T. Berger, *Rate-Distortion Theory*. Wiley Online Library, 1971.
- [6] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *Information Theory, IRE Transactions on*, vol. 8, no. 5, pp. 293–304, 1962.
- [7] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets nyquist: Capacity of sampled gaussian channels," *Information Theory, IEEE Transactions on*, vol. 59, no. 8, pp. 4889–4914, 2013.
- [8] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, "Channel capacity under sub-nyquist nonuniform sampling," *ArXiv e-prints*, Apr. 2012.
- [9] R. Gray, D. Neuhoff, and J. Omura, "Process definitions of distortion-rate functions and source coding theorems," *Information Theory, IEEE Transactions on*, vol. 21, no. 5, pp. 524–532, Sep 1975.
- [10] T. Berger and J. Gibson, "Lossy source coding," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2693–2723, 1998.