Abstract

The amount of information lost in sub-Nyquist sampling of a continuous-time Gaussian stationary process is quantified. We consider a combined source coding and sub-Nyquist reconstruction problem in which the input to the encoder is a sub-Nyquist sampled version of the analog source. We first derive an expression for the mean square error in reconstruction of the process as a function of the sampling frequency and the average number of bits describing each sample. We define this function as the distortion-rate-frequency function. It is obtained by reverse water-filling over a spectral density associated with the minimum variance reconstruction of an undersampled Gaussian process, plus the error in this reconstruction. Further optimization to reduce distortion is then performed over the sampling structure, and an optimal pre-sampling filter associated with the statistics of the input signal and the sampling frequency is found. This results in an expression for the minimal possible distortion achievable under any uniform sampling scheme. It unifies the Shannon-Nyquist-Whittaker sampling theorem and Shannon rate-distortion theory for Gaussian sources.

Index Terms

Source coding, Gaussian rate-distortion, remote source coding, sub-Nyquist sampling.

I. INTRODUCTION

Consider the task of storing an analog source in digital memory. The trade-off between the bit-rate of the samples and the minimal possible distortion in the reconstruction of the signal from these samples is described by the rate-distortion function of the source [1, Sec. 10]. A key idea in determining the rate-distortion function of an analog source is to map the continuous-time process into a discrete-time process based on sampling above the Nyquist frequency [2, Sec. 4.5.3]. Since wideband signaling and A/D technology limitations can preclude sampling signals at their Nyquist rate [3], an optimal source code based on such discrete-time representation may be impractical in certain scenarios. In addition, some applications may be less sensitive to inaccuracies in the data, which suggests that the sampling frequency can be reduced far above the Nyquist frequency without significantly affecting performance. This motivates us to consider the source coding problem in Fig. 1 in which an analog random signal $X(t)$ needs...
to be reconstructed from rate limited samples of its noisy and filtered version \(Z(\cdot)\). This introduces a combined sampling and source coding problem, which lies at the intersection of information theory and signal processing.

The quantities of merit in this problem are the sampling frequency \(f_s\), the source coding rate \(R\) and the average distortion \(D\). If the sampling frequency is such that the input to the sampler \(Z(\cdot)\) can be fully reconstructed from the samples, then the sampling operation has no effect on distortion and the trade-off between the source coding rate and the distortion is given by the classic rate-distortion theory initiated by Shannon [4, Part V]. The other extreme is where the source coding rate \(R\) goes to infinity, in which case we are left with the signal processing problem of reconstructing an undersampled signal in the presence of noise [5]. The reductions of the general problem in these two special cases are illustrated by the diagram in Fig. 2.

In this work we focus on uniform sampling of Gaussian stationary processes under quadratic distortion, using single branch and multi-branch sampling. We determine the expression for the three-dimensional manifold representing the trade-off among \(f_s\), \(R\) and \(D\) in terms of the power spectral density (PSD) of the source, the noise and the sampling mechanism. In addition, we derive an expression for the optimal pre-sampling filter and the corresponding minimal possible distortion attainable under any uniform sampling scheme. This minimal distortion provides a lower bound for the distortion achieved by any uniform sampling and quantization scheme. In this sense, the distortion-rate function (DRF) associated with our model quantifies the excess distortion incurred due to encoding based on the information in any uniform sampling scheme of Gaussian stationary sources in lieu of the full source information.

Note that under the model in Fig. 1, quantization constraints of the samples are not considered in the sampling operation. This means that for any fixed \(R\), minimal distortion is achieved by taking \(f_s\) beyond the Nyquist frequency of \(Z(\cdot)\) (or to infinity, in the case where \(Z(\cdot)\) is not band-limited) such that \(Z(\cdot)\) can be reconstructed from the samples \(Y[\cdot]\) with zero error. However, technology limitations may confine \(f_s\) to a limited range \([3] f_s \leq f_{max}\). An optimal choice of \(f_s \in (0, f_{max})\) is a result of this work.
Fig. 2: The combined source coding and sampling problem subsumes two classical problems in information theory and signal processing.

A. Related Work

Since in our model the encoder needs to deliver information about the source but cannot observe it directly, the problem of characterizing the DRF belongs to the regime of indirect or remote source coding problems [6], [2, Section 3.5]. A classical indirect distortion-rate problem solved by Dobrushin and Tsybakov in [7], considered the case where the observable process and the source are jointly Gaussian and stationary. We refer to this problem as the stationary Gaussian indirect source coding problem. It can be obtained as a special case from our model when the sampled process can be fully reconstructed from its samples, which is the case for example if \( Z(\cdot) \) is band-limited and \( f_s \) is larger than its Nyquist frequency. In their work, Dobrushin and Tsybakov implicitly showed that the stationary Gaussian indirect source coding problem can be separated into two independent problems: MMSE estimation and standard (direct) source coding. This separation principle was extended to any source distribution under quadratic distortion in [8], and to general distortion measures in [6]. These results are discussed in detail in Section IV.

The other branch of the diagram in Fig. 2 is obtained if we relax the rate constraint in the model in Fig. 1. The distortion at a given sampling frequency is then given by the minimal MSE (MMSE) in estimating \( X(\cdot) \) from its noisy sub-Nyquist samples \( Y[\cdot] \). An expression for this MMSE as well as a description of the optimal pre-sampling filter that minimizes it were derived in [9] for single branch sampling. See also [5] and [10] for a simple derivation of this MMSE. In particular, the MMSE expression establishes the sufficiency of uniform sampling above the Nyquist rate for perfect reconstruction of random stationary signals, a fact which was first noted in [11].

A general necessary and sufficient sampling rate for perfect reconstruction of a sampled signal is termed the Landau rate [12] which is the spectral occupancy of the signal, i.e. the Lebesgue measure of the support of its PSD. Perfect reconstruction from samples at the Landau rate was proven to be possible for the class of signals whose spectral content resides within several sub-bands over a wide spectrum [13] [14]. This can be achieved by multi-branch sampling in which the input is sampled through \( P \) linear filters. This sampling strategy was first analyzed by Papoulis [15]. The MMSE in multi-branch sampling, as well as the optimal pre-sampling filters that minimize it, were derived in [16]. It was shown there that the optimal pre-sampling filters that maximize the capacity of a channel with sampling at the receiver are the same filters that minimize the MMSE in sub-Nyquist sampling. An
extension of this result to non-uniform sampling was considered in [17]. These results on the MMSE in sub-Nyquist sampling will be discussed and extended in Section III.

B. Contributions

The main result of this paper is a closed form expression for the function \( D(f_s, R) \) which represents the minimal quadratic distortion achieved in the reconstruction of any continuous time Gaussian stationary processes from its rate \( R \) uniform noisy samples at frequency \( f_s \). This is shown to be given by a parametric reverse water-filling solution, which in the case of single branch sampling takes the form

\[
R(f_s, \theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left[ J(f) / \theta \right] df,
\]

(1a)

\[
D(f_s, R(\theta)) = \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \left[ J(f) - \theta \right]^+ df,
\]

(1b)

where the function \( J(f) \) is defined in terms of the pre-sampling filter \( H(f) \), the PSD of the source \( S_X(f) \) and the PSD of the noise \( S_\eta(f) \). The proof of (1) relies on an extension of the stationary Gaussian indirect source coding problem considered by Dobrushin and Tsybakov to vector-valued processes, which is given by Theorem 8.

The result of Dobrushin and Tsybakov was obtained for the case where the source and the observable process are jointly Gaussian and stationary. In our setting the observable discrete time process \( Y(\cdot) \) and the analog processes \( X(\cdot) \) and \( Z(\cdot) \) are still jointly Gaussian, but the optimal reconstruction process under quadratic distortion, 

\[ \{ \mathbb{E}[X(t) | Y(\cdot) \} , t \in \mathbb{R} \}, \]

is in general not a stationary process. An easy way to see this is to consider the estimation error at the sampling times \( t \in \mathbb{Z}/f_s \) in the noiseless case \( \eta(\cdot) \equiv 0 \), which must vanish, while the estimation error at any \( t \notin \mathbb{Z}/f_s \) is not necessarily zero (assuming a non-degenerate process). In Section V we present a way to overcome this difficulty. The idea is to use time discretization, after which we can identify a vector-valued process jointly stationary with the samples \( Y(\cdot) \) which contains the same information as the discretized version of \( X(\cdot) \). The result is the indirect DRF at any given sampling frequency in a discrete version of our problem, which converges to \( D(f_s, R) \) under some mild conditions.

In practice, the system designer may choose the parameters of the sampling mechanism to achieve minimal reconstruction error for a given sampling frequency \( f_s \) and source coding rate \( R \). This suggests that for a given source statistic and a sampling frequency \( f_s \), an optimal choice of the pre-sampling filters can further reduce the distortion for a given source coding rate. In the single branch setting, this optimization is carried out in Subsection V-F and leads to the function \( D^*(f_s, R) \), which gives a lower bound on \( D(f_s, R) \). The optimal pre-sampling filter \( H(f) \) is shown to be the filter that passes only one frequency in each discrete aliasing set \( f + f_s \mathbb{Z} \) and suppresses the rest. In other words, minimal distortion is achieved by eliminating aliasing.

We later extend our results to systems with \( P \in \mathbb{N} \) sampling branches. We derive expressions for \( D(P, f_s, R) \) and \( D^*(P, f_s, R) \), which denote the DRF with average sampling frequency \( f_s \) and the DRF under optimal pre-sampling filtering, respectively. As the number of sampling branches \( P \) goes to infinity, \( D^*(P, f_s, R) \) is shown to converge (but not monotonically) to a smaller value \( D_l(f_s, R) \), which essentially describes the minimal distortion achievable under
any uniform sampling scheme. The functions $D^*(P, f_s, R)$ and $D_1(f_s, R)$ depend only on the statistics of the source and the noise. In particular, if we take the noise to be zero, $D^*(P, f_s, R)$ and $D_1(f_s, R)$ can be considered to describe a fundamental trade-off in signal processing and information theory associated with any Gaussian stationary source.

Our main result (1) shows that the function $D(f_s, R)$ is obtained by reverse water-filling over the function $J(f)$ that was initially introduced to calculate the MMSE in sub-Nyquist sampling ($\text{mmse}_{X|Y}(f_s)$) in [9] and [5]. As a result, the optimal pre-sampling filters that minimize $D(f_s, R)$ are the same optimal pre-sampling filters that minimize $\text{mmse}_{X|Y}(f_s)$. In Section III, we prove this result using an approach based on the polyphase components and aliasing-free sets. This allows us to derive the MMSE in sub-Nyquist sampling using multi-branch uniform sampling with the number of branches going to infinity. This result is also used to obtain our main result (1).

C. Organization

The rest of the paper is organized as follows: the combined sampling and source-coding problem is presented in Section II. Sections III and IV are dedicated to the special cases of sub-Nyquist sampling and indirect source coding, as shown in the respective branches in the diagram of Fig. 2. In Section V, we prove our main results for single branch sampling, which are extended to multi-branch sampling in Section VI. Concluding remarks are given in Section VII.

II. Problem Statement

The system model for our combined sampling and source-coding problem is depicted in Fig. 3. The source $X(\cdot) = \{X(t), t \in \mathbb{R}\}$ is a real Gaussian stationary process with variance $\sigma^2_X \triangleq \int_{-\infty}^{\infty} S_X(f) \, df < \infty$, and power spectral density (PSD)

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

The noise $\eta(\cdot)$ is a real Gaussian stationary process independent of the source with spectral density $S_\eta(f)$. The sampler receives the noisy source as an input, and produces a discrete time process $Y[\cdot]$ at a rate of $f_s$ samples per time unit. The process $Y[\cdot]$ is in general a complex vector-valued process since pre-sampling operations that result in a complex valued process are allowed in the sampler. The encoder represents the samples $Y[\cdot]$ in an average rate of no more than $R$ bits per time unit.

Assuming the noise is additive and independent poses no limitation on the generality since for any jointly stationary and Gaussian process pairs $X(\cdot)$ and $Z(\cdot)$, this relationship can be created via a linear transformation. This linear transformation can be seen as a part of the specific sampler structure we consider below.

Throughout this paper, we use round brackets and square brackets to distinguish between continuous-time and discrete-time processes. Vectors and matrices are denoted by bold letters. In addition, we use the word ‘rate’ to indicate information rate rather than sampling rate, and use ‘sampling frequency’ for the latter. In some cases it is more convenient to measure the information rate in bits per sample, which is given by $\bar{R} \triangleq R/f_s$. 
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

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

where $\|X(\cdot)\|$ is the $L_2$ norm\footnote{In our notation, we allow $\|\hat{X}(\cdot)\|$ to take the value $\infty$ if $\hat{X}(\cdot)$ is not in $L_2$.} of the process $X(\cdot)$, defined by

$$\|X(\cdot)\|^2 \triangleq \limsup_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \mathbb{E}\left[\left(\hat{X}(t)\right)^2\right] dt.$$ 

The main problem we consider is as follows: given a sampling scheme with sampling frequency $f_s$ and an information rate $R \geq 0$, what is the minimum possible distortion that can be attained between $X(\cdot)$ and $\hat{X}(\cdot)$? Classical results in rate-distortion theory [2], [18] imply that this problem has the informational rate-distortion characterization depicted in Fig. 4, in which $P_{\hat{Y}|Y}$ denotes a ‘test channel’ between $Y(\cdot)$ and $\hat{Y}(\cdot) = \{\hat{Y}[n], n \in \mathbb{Z}\}$, and the reconstruction process $\hat{X}(\cdot)$ is obtained from the process $\hat{Y}(\cdot)$. Denote by $\mathcal{A}(f_s, R)$ the set of all random mappings $Y(\cdot) \to \hat{Y}(\cdot) \to \hat{X}(\cdot)$ measurable with respect to the $\sigma$-algebra generated by $Y(\cdot)$, such that the mutual
information rate between $Y[\cdot]$ and $\hat{Y}[\cdot]$, defined by

$$I(\hat{Y}[\cdot];Y[\cdot]) \triangleq \lim_{N \to \infty} \frac{1}{2N+1} \log \left( \frac{\log N}{I(\hat{Y}[-N,\ldots,N-1,N];Y[-N,\ldots,N-1,N])} \right)$$

is limited to $\bar{R} = R/f_s$ bits per sample. We seek to determine the function

$$D = \inf_{\mathcal{A}(f_s,R)} \mathbb{E}d(\hat{X}(\cdot), X(\cdot))$$

(4)

An equivalent problem is to find the rate-distortion function at a given sampling frequency $f_s$,

$$R = f_s \inf_{\mathcal{A}'(f_s,D)} I(\hat{Y}[\cdot];Y[\cdot]),$$

where now $\mathcal{A}'(f_s,D)$ is the set of all random mappings $Y[\cdot] \to \hat{Y}[\cdot] \to \hat{X}(\cdot)$ such that $\mathbb{E}d(\hat{X}(\cdot), X(\cdot)) \leq D$. For convenience, we express the distortion $D$ as a function of $R$ rather than the other way around.

Besides the sampling frequency $f_s$ and the source coding rate $R$, $D$ in (4) depends on the sampling structure. In this work we restrict ourselves to samplers consisting of a pre-sampling filtering operation followed by a pointwise sampler. We focus on two basic structures:

1) **Single-branch uniform sampling** (Fig. 5(a)): $H$ is an LTI system with frequency response $H(f)$ which serves as a pre-sampling filter. This means that $Z(\cdot) = \{Z(t), t \in \mathbb{R}\}$ and $X(\cdot)$ are jointly Gaussian and stationary with joint spectral density

$$S_{XZ}(f) \triangleq \int_{-\infty}^{\infty} \mathbb{E}[X(t+\tau)Z(t)] e^{-2\pi i \tau f} d\tau = S_X(f)H^*(f).$$
Although we allow an arbitrary noise PSD $S_\eta(f)$, in order for the uniform sampling operation to be well defined we require that
\[
\int_{-\infty}^{\infty} S_Z(f)\,df = \int_{-\infty}^{\infty} S_{X+\eta}(f)|H(f)|^2\,df < \infty.
\] (5)

In (5) and henceforth we denote $S_{X+\eta}(f) \triangleq S_X(f) + S_\eta(f)$, which is justified since $X(\cdot)$ and $\eta(\cdot)$ are independent processes. We sample $Z(\cdot)$ uniformly at times $\frac{n}{fs}$, resulting in the discrete time process
\[
Y[n] = Z\left(\frac{n}{fs}\right), \quad n \in \mathbb{Z}.
\]

Recall that the spectral density of $Y[\cdot]$ is given by
\[
S_Y(e^{2\pi\phi}) = \sum_{k \in \mathbb{Z}} \mathbb{E}[Y[n]Y[n+k]] e^{-2\pi ik\phi} = \sum_{k \in \mathbb{Z}} f_s S_Z(f_s(\phi - k)).
\]

We denote by $D(fs,R)$ the DRF using uniform single-branch sampling at frequency $fs$.

2) Multi-branch or filter-bank uniform sampling (Fig. 5-b): For each $p = 1, \ldots, P$, $Z_p(\cdot)$ is the output of the LTI system $H_p$ whose input is the source $X(\cdot)$. The sequence $Y_p[\cdot]$ is obtained by uniformly sampling $Z_p(\cdot)$ at frequency $fs/P$, i.e.
\[
Y_p[n] = Z\left(\frac{np}{fs}\right), \quad p = 1, \ldots, P.
\]
The output of the sampler is the vector $Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot])$. Since each one of the $P$ branches produces samples at rate $fs/P$, the sampling frequency of the system is $fs$. The DRF using $P$ uniform sampling branches with an average sampling frequency $fs$ will be denoted $D(P,fs,R)$ ($D(fs,R) = D(1,fs,R)$ in this notation).

The parameters of the two sampling schemes above are the average sampling frequency $fs$ and the pre-sampling filters $H(f)$ or $H_1(f), \ldots, H_p(f)$. Given an average sampling frequency $fs$ and a source coding rate $R$, we also consider the following question: what are the optimal pre-sampling filters that minimize $D(fs,R)$ and $D(P,fs,R)$?

The values of $D^*(fs,R)$ and $D^*(P,fs,R)$ under an optimal choice of the pre-sampling filters are respectively denoted by $D(fs,R)$ and $D(P,fs,R)$. We also determine the behavior of $D^*(P,fs,R)$ as the number of branches $P$ goes to infinity.

The DRF is bounded from below by the MMSE in estimating $X(\cdot)$ from the samples $Y[\cdot]$. In the next section we begin our discussion in the behavior of this estimation error for a given sub-Nyquist sampling frequency $fs$.

III. MMSE IN SUB-NYQUIST SAMPLING OF GAUSSIAN STATIONARY PROCESSES

If we relax the rate constraint in our combined sampling and source coding problem of Fig. 4 i.e. take $R \to \infty$, our source coding problem reduces to the problem of reconstructing an analog Gaussian stationary processes from its sub-Nyquist samples under an MSE criterion. In addition, we may ask what is the optimal pre-sampling filter that minimizes the MMSE for a given input signal and sampling frequency. These two problems were addressed
in [9] for single branch sampling, where an extension to multi-branch sampling can be implicitly found in [16]. In this section we use a different technique to re-prove these results, and provide a new representation useful in proving our main results in Section V. In particular, we introduce the notion of an aliasing-free set to describe the optimal pre-sampling filter that minimizes the MMSE for a given sampling frequency $f_s$.

The system model is depicted in Fig. 6. The problem we consider is to find the MMSE in the estimation of the source $X(\cdot)$ from the samples $Y[\cdot]$ which we denote as $\text{mmse}_{X|Y}(f_s)$:

$$\text{mmse}_{X|Y}(f_s) \triangleq \mathbb{E}[\|X(\cdot) - \hat{X}(\cdot)\|^2],$$

(6)

where $\hat{X}(t) \triangleq \mathbb{E}[X(t)|Y[\cdot]]$, $t \in \mathbb{R}$, is the MMSE estimator of $X(\cdot)$ from the samples $Y[\cdot]$. Note that (6) can be written as

$$\text{mmse}_{X|Y}(f_s) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \mathbb{E} \left[ \left( X(t) - \hat{X}(t) \right)^2 \right] dt$$

$$= \int_{0}^{1} \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E} \left[ \left( X \left( \frac{n + \Delta}{f_s} \right) - \hat{X} \left( \frac{n + \Delta}{f_s} \right) \right)^2 \right] d\Delta$$

$$= \int_{0}^{1} \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E} \left[ \left( X_{\Delta}[n] - \hat{X}_{\Delta}[n] \right)^2 \right] d\Delta$$

$$= \int_{0}^{1} \|X[\cdot] - \hat{X}_{\Delta}[\cdot]\|^2 d\Delta, \quad (7)$$

where the process $X_{\Delta}[\cdot]$ is the $\Delta$ polyphase component of $X(\cdot)$ [19], defined by

$$X_{\Delta}[n] \triangleq X \left( \frac{n + \Delta}{f_s} \right), \quad n \in \mathbb{Z},$$

(8)

and $\hat{X}_{\Delta}[n] \triangleq \mathbb{E}[X_{\Delta}[n]|Y[\cdot]]$. Since $X_{\Delta}[\cdot]$ and $Y[\cdot]$ are jointly Gaussian and stationary, (7) leads to

$$\text{mmse}_{X|Y}(f_s) = \int_{0}^{1} \text{mmse}_{X_{\Delta}|Y} d\Delta$$

$$= \sigma_X^2 - \int_{0}^{1} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_{\Delta}|Y}(e^{2\pi i \phi}) d\Delta d\phi,$$

(9)

where we denoted $\text{mmse}_{X_{\Delta}|Y} \triangleq \|X_{\Delta} - \hat{X}_{\Delta}[\cdot]\|^2$. We begin by evaluating (9) for the case of a single branch sampler.
A. Single Branch Uniform Sampling

Theorem 1. Consider the model of Fig. 6 where we use the single branch sampler of Fig. 5(a). The MMSE in the estimation of \( X(\cdot) \) from \( Y[\cdot] \) is given by

\[
\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} J(f) df, \quad (10)
\]

where

\[
J(f) \triangleq \frac{\sum_{k \in \mathbb{Z}} S_X^2 (f - f_s k) |H(f - f_s k)|^2}{\sum_{k \in \mathbb{Z}} S_X + \eta (f - f_s k) |H(f - f_s k)|^2}. \quad (11)
\]

Proof: The result is obtained by evaluating (9). We first find \( S_{X|Y}(e^{2\pi \phi}) = \frac{|S_{X|Y}(e^{2\pi \phi})|^2}{S_Y(e^{2\pi \phi})} \). We have

\[
S_Y(e^{2\pi \phi}) = \sum_{k \in \mathbb{Z}} S_{X + \eta} (f_s (\phi - k)) |H(f_s (\phi - k))|^2,
\]

\[
S_{X|Y}(e^{2\pi \phi}) = \sum_{k \in \mathbb{Z}} E \left[ X \left( \frac{n + l + \Delta}{f_s} \right) Z \left( \frac{n}{f_s} \right) \right] e^{-2\pi il\phi}
\]

\[
= \sum_{k \in \mathbb{Z}} S_X (f_s (\phi - k)) H^*(f_s (\phi - k)) e^{2\pi ik\Delta},
\]

where we used the fact that the spectral density of the polyphase component \( X_{\Delta}[\cdot] \) equals

\[
S_{X_{\Delta}}(e^{2\pi \phi}) = \sum_{k \in \mathbb{Z}} S_X (f_s (\phi - k)) e^{2\pi ik\Delta \phi}.
\]

This leads to

\[
S_{X|Y}(e^{2\pi \phi}) = \frac{|S_{X|Y}(e^{2\pi \phi})|^2}{S_Y(e^{2\pi \phi})} = \frac{\sum_{k \in \mathbb{Z}} S_X (f_s (\phi - k)) H^*(f_s (\phi - k)) S_X H(f_s (\phi - m)) e^{2\pi ik(m-k)\Delta}}{\sum_{k \in \mathbb{Z}} S_Y (f_s (\phi - k))}. \quad (12)
\]

Integrating (12) over \( \Delta \) from 0 to 1 gives

\[
\frac{\sum_{k \in \mathbb{Z}} S_X^2 (f_s (\phi - k)) |H(f_s (\phi - k))|^2}{\sum_{k \in \mathbb{Z}} S_{X + \eta} (f_s (\phi - k)) |H(f_s (\phi - k))|^2}. \quad (13)
\]

Substituting (13) into (9) and changing the integration variable from \( \phi \) to \( f/f_s \) leads to (10).

This proof of Theorem 1 also gives a new interpretation to the function \( J(f) \) as the average of spectral densities of the estimators of the polyphase components of \( X(\cdot) \), namely

\[
J(f) = \int_{0}^{1} f_s S_{X|Y}(f/f_s) df. \quad (14)
\]

We note that since the denominator in (11) is periodic in \( f \) with period \( f_s \), (10) can be written as

\[
\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\infty}^{\infty} \frac{S_X^2(f) |H(f)|^2}{\sum_{k \in \mathbb{Z}} S_{X + \eta} (f - f_s k) |H(f - f_s k)|^2} df
\]

\[
= \int_{-\infty}^{\infty} S_X(f) \left( 1 - \frac{S_X(f) |H(f)|^2}{\sum_{k \in \mathbb{Z}} S_{X + \eta} (f - f_s k) |H(f - f_s k)|^2} \right) df. \quad (15)
\]
The function
\[
1 - \frac{S_x(f)|H(f)|^2}{\sum_{k \in \mathbb{Z}} S_x+n(f-f,k)|H(f-f,k)|^2}
\]  
(16)
is non-negative and can be interpreted as the sampling loss factor in terms of MSE as a result of the sampling system.

B. Optimal pre-Sampling Filter

We now consider the pre-sampling filter \(H\) as a part of the system design and ask what is the optimal pre-sampling filter \(H^*\) that minimizes (10): This problem is equivalent to finding the filter that maximizes \(J(f)\) for every frequency \(f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)\) independently, i.e. we are looking to determine
\[
J^*(f) \triangleq \sup_H J(f) = \sup_H \frac{\sum_{k \in \mathbb{Z}} S_X(f-f,k)|H(f-f,k)|^2}{\sum_{k \in \mathbb{Z}} S_X+n(f-f,k)|H(f-f,k)|^2}
\]  
(17)
in the domain \(\left(-\frac{f_s}{2}, \frac{f_s}{2}\right)\).

The optimal pre-sampling filter \(H^*(f)\) that achieves (17) was first described in [9]. In what follows we use a different approach to derive these results, which is obtained by defining a set of frequencies \(F^*\) of minimal Lebesgue measure such that
\[
\int_{F^*} \frac{S_X(f)}{S_X+n(f)} df = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \sup_{H} \frac{S_X^2(f-f,k)}{S_X+n(f-f,k)} df.
\]  
(18)

We will see later that such a set defines an optimal filter \(H^*(f)\). We note that since the integrand in the RHS of (18) is periodic in \(f\) with period \(f_s\), we can conclude that excluding a set of Lebesgue measure zero, the set \(F^*\) will not contain two frequencies \(f_1, f_2 \in \mathbb{R}\) that differ by an integer multiple of \(f_s\) due to its minimality. This property will be given the following name:

**Definition 1** (aliasing-free set). A measurable set \(F \subset \mathbb{R}\) is said to be aliasing-free with respect to the sampling frequency \(f_s\), if for almost all pairs \(f_1, f_2 \in F\), it holds that \(f_1 - f_2 \notin f_s \mathbb{Z} = \{f_s k, k \in \mathbb{Z}\}\).

The aliasing-free property imposes the following restriction on a measure of a bounded set:

**Theorem 2.** Let \(F\) be an aliasing-free set with respect to \(f_s\). If \(F\) is bounded, then the Lebesgue measure of \(F\) does not exceed \(f_s\).

**Proof:** By the aliasing-free property, for any \(n \in \mathbb{Z}\) \(\setminus \{0\}\) the intersection of \(F\) and \(F^* + nf_s\) is empty. It follows that for all \(N \in \mathbb{N}\), \(\mu(\bigcup_{n=1}^{N} (F^* + nf_s)) = N\mu(F^*)\). Now assume \(F^*\) is bounded by the interval \((-M,M)\) for some \(M > 0\). Then \(\bigcup_{n=1}^{N} (F^* + nf_s)\) is bounded by the interval \((-M,M+nf_s)\). It follows that
\[
\frac{\mu(F^*)}{f_s} = \frac{N\mu(F^*)}{Nf_s} = \frac{\mu(\bigcup_{n=1}^{N} (F^* + nf_s))}{Nf_s} \leq \frac{2M + Nf_s}{Nf_s}.
\]

\(^2\)By almost any we mean for all but a set of Lebesgue measure zero.
Letting \( N \to \infty \), we conclude that \( \mu(F^*) \leq f_s \).

We denote by \( AF(f_s) \) the collection of all bounded aliasing free sets with respect to \( f_s \). Note that a process whose spectrum has support in \( AF(f_s) \) admits no aliasing when uniformly sampled at frequency \( f_s \), i.e., such a process can be reconstructed with probability one from its non-noisy uniform samples at frequency \( f_s \). As the following theorem shows, the optimal pre-sampling filter is characterized by an aliasing-free set with an additional maximality property.

**Theorem 3.** The optimal pre-sampling filter \( H^*(f) \) that maximizes \( J(f) \) and minimizes \( \text{mmse}_{X|Y}(f_s) \) for a given \( f_s \) and for all \( R \geq 0 \), is given by

\[
H^*(f) = \begin{cases} 
1 & f \in F^*, \\
0 & \text{otherwise}, 
\end{cases}
\]

where \( F^* = F^*\left(f_s, \frac{S_X(f)}{S_X+\eta(f)}\right) \in AF(f_s) \) satisfies

\[
\int_{F^*} \frac{S_X^2(f)}{S_X+\eta(f)} df = \sup_{F \in AF(f_s)} \int_{F} \frac{S_X^2(f)}{S_X+\eta(f)} df.
\]

The optimal MMSE at a given sampling frequency \( f_s \) is

\[
\text{mmse}_{X|Y}^*(f_s) = \sigma_X^2 - \int \frac{S_X^2(f)}{S_X+\eta(f)} df.
\]

*Proof:* See Appendix VIII.

**Remarks:**

(i) The proof also shows that

\[
\int_{F^*} \frac{S_X^2(f)}{S_X+\eta(f)} df = \int \frac{J^*(f)}{2} df,
\]

where

\[
J^*(f) \triangleq \sup_k \frac{S_X(f-f_k)}{S_X+\eta(f-f_k)},
\]

i.e.

\[
\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int \frac{1}{2} J^*(f) df.
\]

(ii) The filter \( H^*(f) \) can be specified only in terms of its support, i.e., in (19) we can replace 1 by any non-zero value. This value may even vary with \( f \).

Theorem 3 motivates the following definition:
Definition 2. For a given spectral density $S(f)$ and a sampling frequency $f_s$, an aliasing free set $F^* \in AF(f_s)$ that satisfies
\[ \int_{F^*} S(f) df = \sup_{F \in AF(f_s)} \int_F S(f) df, \]
is called a maximal aliasing-free set with respect to $f_s$ and the spectral density $S(f)$, and denoted by $F^*(f_s, S)$.

Roughly speaking, the maximal aliasing free set $F^*(f_s, S)$ can be constructed by going over all frequencies $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$, and include in $F^*(f_s, S)$ the frequency $f^* \in \mathbb{R}$ such that $S(f^*)$ is maximal among all $S(f)$, $f \in f^*-f_s\mathbb{Z}$. A similar construction was described in [9].

Theorem [3] says that the optimal pre-sampling filter is the indicator of the maximal aliasing free set associated with the spectral density $\frac{S_1(f)}{S_X + \epsilon(f)}$ and sampling frequency $f_s$. Intuition for this result is given through the following example.

Example 1 (Joint MMSE Estimation). Let $U_1$ and $U_2$ be two independent Gaussian random variables with variances $C_{U_1}$ and $C_{U_2}$ respectively. We are interested in MMSE estimation of $U = (U_1, U_2)$ from a noisy linear combination of their sum $V = h_1(U_1 + \xi_1) + h_2(U_2 + \xi_2)$, where $h_1, h_2 \in \mathbb{R}$ and $\xi_1, \xi_2$ are another two Gaussian random variables with variances $C_{\xi_1}$ and $C_{\xi_2}$ respectively, independent of $U_1$ and $U_2$ and independent of each other. We have
\[ \text{mmse}_{U|V} = \frac{1}{2} \left( \text{mmse}_{U_1|V} + \text{mmse}_{U_2|V} \right) \]
\[ = \frac{1}{2} \left( C_{U_1} + C_{U_2} - \frac{h_1^2 C_{U_1}^2 + h_2^2 C_{U_2}^2}{h_1^2 (C_{U_1} + C_{\xi_1}) + h_2^2 (C_{U_2} + C_{\xi_2})} \right). \]
The optimal choice of the coefficients vector $h = (h_1, h_2)$ that minimizes (22) is
\[ h = \begin{cases} (c, 0) & \frac{c^2}{C_{U_1} + C_{\xi_1}} > \frac{c^2}{C_{U_2} + C_{\xi_2}} \\ (0, c) & \frac{c^2}{C_{U_1} + C_{\xi_1}} < \frac{c^2}{C_{U_2} + C_{\xi_2}} \end{cases}, \]
where $c$ is a constant different from zero. If $\frac{c^2}{C_{U_1} + C_{\xi_1}} = \frac{c^2}{C_{U_2} + C_{\xi_2}}$, then any non-trivial linear combination results in the same estimation error. This example can be generalized into a countable number of random variables $U = (U_1, U_2, \ldots)$ and respective noise sequence $\xi = (\xi_1, \xi_2, \ldots)$ such that $V = \sum_{i=1}^{\infty} h_i(U_i + \xi_i) < \infty$ with probability one. The optimal coefficient vector $h = (h_1, h_2, \ldots)$ that minimizes $\text{mmse}_{U|V}$ is the indicator for the maximum among $\left\{ \frac{c^2}{C_{U_i} + C_{\xi_i}}, i = 1, 2, \ldots \right\}$.

In the context of optimal pre-sampling filtering, each frequency $f$ in the support of $S_X(f)$ can be seen as an independent component of the process $X(\cdot)$ with spectrum $\approx I_{[f, f + \Delta f]} S_X(f)$ (see for example the derivation of the SKP reverse water filling in [20]). For a given $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$, the analogue for the vector $U$ in our case are the components of the source process that corresponds to the frequencies $f - f_s \mathbb{Z}$, which are folded and summed together due to aliasing in the sampling process: each set of the form $f - f_s \mathbb{Z}$ corresponds to a linear combination of
Fig. 7: Maximal aliasing-free sets with respect to the PSD $S_X^2(f)/S_X + \eta(f)$ and sampling frequency $f_s = f_{Nyquist}/4$ (left) and $f_s = f_{Nyquist}/2$ (right), for 1, 2 and 3 sampling branches. The first, second and third maximal aliasing-free set is given by the frequencies below the blue, green and red areas, respectively. The sets below the total colored area all have Lebesgue measure $f_s$. Assuming $S_\eta(f) \equiv 0$, $mmse(f_s)$ is given by the white area under the graph. By Theorem 6, the case $P \rightarrow \infty$ corresponds to the set $\mathcal{F}^*$ that achieves the RHS of (27).

A countable number of independent Gaussian random variables attenuated by the coefficients $\{H(f - f_s k), k \in \mathbb{Z}\}$. The optimal choice of coefficients that minimizes the MMSE in the joint estimation of all source components are those that pass only the spectral component with maximal $\frac{S_X^2(f')}{S_X + \eta(f')}$ among all $f' \in f - f_s \mathbb{Z}$, and suppress the rest. This means that under the MSE criterion, the optimal choice is to eliminate aliasing at the price of losing all information contained in spectral components other than the maximal one. An example for a maximal aliasing-free set for a specific PSD appears in Fig. 7 under the case $P = 1$. The MMSE with the optimal pre-sampling filter and without it appears in Fig. 8.
C. Multi-Branch Sampling

We now extend Theorem 1 and Theorem 3 to the case of multi-branch sampling. The system model is given by Fig. 6 with the sampler of Fig. 5(b).

**Theorem 4** (MMSE multi-branch sampling). For each $p = 1, \ldots, P$, let $Z_p(\cdot)$ be the process obtained by passing a Gaussian stationary source $X(\cdot)$ corrupted by a stationary Gaussian noise $\eta(\cdot)$ through an LTI system $H_p(f)$. Let $Y_p[\cdot]$, be the samples of the process $Z_p(\cdot)$ at frequency $f_s/P$, namely

$$Y[n] = Z(Pn/f_s) = h_p(\cdot) \ast (X(\cdot) + \eta(\cdot))(Pn/f_s).$$

The MMSE in estimating $X(\cdot)$ from the samples $Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot])$, is given by

$$\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \text{Tr}(J(f)) \, df. \quad (23)$$

Here $J(f)$ is the $P \times P$ matrix defined by

$$J(f) \triangleq \tilde{S}_Y^{-1/2}(f) K(f) \tilde{S}_Y^{-1/2}(f), \quad (24)$$

where the matrices $\tilde{S}_Y(f), K(f) \in \mathbb{C}^{P \times P}$ are given by

$$(\tilde{S}_Y(f))_{i,j} = \sum_{k \in \mathbb{Z}} \{S_{X+\eta} H_i^* H_j\} (f - f_s k),$$
and

\[(K(f))_{i,j} = \sum_{k \in \mathbb{Z}} \{S_k^2 H_i^* H_j\}(f - f,k) .\]

Proof: See Appendix IX.

D. Optimal pre-Sampling Filters Bank

A generalization of Theorem 3 to the case of multi-branch sampling is as follows:

**Theorem 5.** The optimal pre-sampling filters \(H_1^* (f), \ldots, H_P^* (f)\) that maximize the eigenvalues of \(J(f)\) and minimize \(\text{mmse}_{X|Y}(f_s)\) are given by

\[H_p^* (f) = \begin{cases} 1 & f \in F_p^* , \quad p = 1, \ldots, P , \\ 0 & f \notin F_p^* , \end{cases}\]

where the sets \(F_1^*, \ldots, F_P^* \in \subset \mathbb{R}\) satisfy

(i) For all \(p = 1, \ldots, P, F_p^* \in \subset \mathbb{R}\) satisfy

\[\int_{F_p^*} S_{X+\eta}^2(f) df = \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} J_1^*(f) df,
\]

where

\[J_1^*(f) \equiv \sup_{k \in \mathbb{Z}} \frac{S_k^2(f - kf_s/P)}{S_{X+\eta}(f - kf_s/P)} ,\]

and for all \(p = 2, \ldots, P,\)

\[\int_{F_p^*} S_{X+\eta}^2(f) df = \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} J_p^*(f) df ,\]

where

\[J_p^*(f) \equiv \sup_{k \in \mathbb{Z}} \frac{S_k^2(f - kf_s/P)}{S_{X+\eta}(f - kf_s/P)} \mathbf{1}_{\mathbb{R}\{F_1^* \cup \cdots \cup F_{p-1}^*\}} ,\]

The resulting MMSE is

\[\text{mmse}_{X|Y}(P,f_s) = \sigma_X^2 - \sum_{p=1}^{P} \int_{F_p^*} S_{X+\eta}^2(f) df = \sigma_X^2 - \sum_{p=1}^{P} \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} J_p^*(f) df .\]

Proof: See Appendix IX.

Remarks:

(i) As in the single-branch case in Theorem 3, the filters \(H_1^* (f), \ldots, H_P^* (f)\) can be specified only in terms of their support and in (25) we can replace 1 by any non-zero value which may vary with \(p\) and \(f\).
(ii) Condition (ii) for the sets $F_1^*,...,F_P^*$ can be relaxed in the following sense: if $F_1^*,...,F_P^*$ satisfy condition (i) and (ii), then $\text{mmse}^*(P,f_s)$ is achieved by any pre-sampling filters defined as the indicators of the sets $F'_1,...,F'_P$ in $\mathcal{A}(f_s/P)$ for which

$$
\sum_{p=1}^P \int_{F'_p} \frac{S_X^2(f)}{S_{X+\eta}(f)} df = \sum_{p=1}^P \int_{F_p^*} \frac{S_X^2(f)}{S_{X+\eta}(f)} df.
$$

(iii) The sets $F_1^*,...,F_P^*$ can be constructed as follows: over all frequencies $-\frac{f_s}{2P} \leq f < \frac{f_s}{2P}$, for each $f$ denote by $f_1^*(f),...,f_P^*(f)$ the $P$ frequencies that correspond to the largest values among $\left\{ \frac{S_X^2(f-jk)}{S_{X+\eta}(f-jk)}, k \in \mathbb{Z} \right\}$. Then assign each $f_p^*(f)$ to $F_p^*$. Under this construction, the set $F_p^*$ can be seen as the $p^{th}$ maximal aliasing free set with respect to $f_s/P$ and $\frac{S_X^2(f)}{S_{X+\eta}(f)}$. This is the construction that was used in Fig. 7.

Fig. 9 illustrates $\text{mmse}^*_{X|Y}(P,f_s)$ as a function of $f_s$ for a specific PSD and $P = 1,2$ and 3. As this figure shows, increasing the number of sampling branches does not necessarily increase $\text{mmse}^*_{X|Y}(P,f_s)$ and may even decrease it for some $f_s$. However, we will see below that $\text{mmse}^*_{X|Y}(P,f_s)$ converges to a fixed number as $P$ increases.

---

Note that if $P_1$ and $P_2$ are co-primes, then even if $P_2 > P_1$, there is no way to choose the pre-sampling filters for the system with $P_2$ branches and sampling frequency $f_s/P_2$ at each branch to produce the same output as the system with $P_1$ sampling branches and sampling frequency $f_s/P_1$. 

---

Fig. 9: MMSE under multi-branch sampling and optimal pre-sampling filter-bank, for $P = 1,2$ and 3 and $P$ large enough such that the bound (26) is attained. The upper dashed line corresponds to $P = 1$ and no pre-sampling filtering. The PSD $S_X^2(f)/S_{X+\eta}(f)$ is given in the small frame.
E. Optimal Sampling

The set $F^*$ defined in Theorem 3 to describe $\text{mmse}_{X|Y}^*(f_\epsilon)$ was obtained by imposing two constraints: (1) a measure constraint $\mu(F^*) \leq f_s$, which is associated only with the sampling frequency, and (2) an aliasing-free constraint, imposed by the sampling structure. Theorem 5 says that in the case of multi-branch sampling, the aliasing-free constraint can be relaxed to $F^* = \bigcup_{p=1}^P F_p^*$, where now we only require that each $F_p^*$ is aliasing-free with respect to $f_s/P$. This means that $F^*$ need not be aliasing-free but its Lebesgue measure still does not exceed $f_s$, which implies the following lower bound on $\text{mmse}_{X|Y}(P, f_s)$:

$$\text{mmse}_{X|Y}(P, f_s) \geq \sigma_X^2 - \sup_{\mu(F) \leq f_s} \int F \frac{S_X^2(f)}{S_X + \eta(f)} df$$

(the supremum is taken over all measurable subsets of $\mathbb{R}$ with Lebesgue measure not exceeding $f_s$). Increasing the number of sampling branches $P$ allows more and more freedom in choosing an optimal frequency set $F^* = \bigcup_{p=1}^P F_p^*$, which eventually converges to a set $\mathscr{F}^*$ that achieves the RHS of (26) as shown in Fig. 7. This means that the RHS of (26) can be achieved with an arbitrarily small gap if we increase the number of sampling branches, as stated in the next theorem.

**Theorem 6.** For any $f_s > 0$ and $\epsilon > 0$, there exists $P \in \mathbb{N}$ and a set of LTI filters $H^*_1(f), \ldots, H^*_P(f)$ such that

$$\text{mmse}_{X|Y}(P, f_s) - \epsilon < \sigma_X^2 - \sup_{\mu(F) \leq f_s} \int F \frac{S_X^2(f)}{S_X + \eta(f)} df.$$

**Proof:** Since any interval of length $f_s$ is in $\mathcal{A} F(f_s)$, it is enough to show that the set $\mathscr{F}^*$ of measure $f_s$ that satisfies

$$\int_{\mathscr{F}^*} \frac{S_X^2(f)}{S_X + \eta(f)} df = \sup_{\mu(F) \leq f_s} \int F \frac{S_X^2(f)}{S_X + \eta(f)} df,$$

can be approximated by $P$ intervals of measure $f_s/P$ ($\mathscr{F}^*$ corresponds to the frequencies below the gray are in the case $P \to \infty$ in Fig. 7). By the regularity of the Lebesgue measure, a tight cover of $\mathscr{F}^*$ by a finite number of intervals is possible. These intervals can be split and made arbitrarily small so that after a finite number of steps we can eventually have all of them at approximately the same length. Denote the resulting intervals $I_1, \ldots, I_P$. For $p = 1, \ldots, P$, define the $p$th filter $H^*_p(f)$ to be the indicator function of the $p$th interval: $H^*_p(f) = 1_{I_p}(f) \equiv 1(f \in I_p)$.

By extending the argument above it is possible to show that we can pick each one of the sets $F^*_p$ to be symmetric about the $y$ axis, so that the corresponding filter $H^*_p(f)$ has a real impulse response. In fact, the construction described in Remark (iii) of Theorem 5 yields a symmetric maximal aliasing free sets. \hfill $\blacksquare$

Theorem 6 implies that $\text{mmse}_{X|Y}^*(P, f_s)$ converges to

$$\sigma_X^2 - \sup_{\mu(F) \leq f_s} \int F \frac{S_X^2(f)}{S_X + \eta(f)} df,$$

as the number of sampling branches $P$ goes to infinity.
Fig. 10: Indirect distortion-rate model

**Remark:** If we take the noise process $\eta(\cdot)$ to be zero in (26), we get

\[
\text{mmse}_{X|Y}(P, f_s) \geq \sigma_X^2 - \sup_{\mu(F) \leq f_s} \int S_X(f) df \\
= \inf_{\mu(F) \leq f_s} \int S_X(f) df. \quad (27)
\]

This shows that perfect reconstruction (with probability one) of $X(\cdot)$ is not possible unless the support of $S_X(f)$ is contained in a set of Lebesgue measure not exceeding $f_s$, a fact which agrees with the well-known condition of Landau [21] for a sampling set of an analytic function.

Theorem 6 establishes the optimal MMSE achievable under any uniform multi-branch sampling. In particular, it shows that uniform multi-branch sampling can be used to sample at the Landau rate and achieves an arbitrarily small MMSE by taking enough sampling branches and a careful selection of the pre-sampling filters. Fig. 9 shows the optimal MMSE under uniform multi-branch sampling as a function of the sampling frequency for a specific PSD, which corresponds to the case $P \to \infty$.

**IV. INDIRECT SOURCE CODING**

The main goal of this subsection is to extend the indirect source coding problem solved in [7], in which the source and the observation are two jointly stationary Gaussian processes, to the case where the source and the observation are two jointly Gaussian vector-valued processes. By doing so we introduce concepts and notions which will be useful in the rest of the paper. We begin by reviewing the problem of indirect source coding.

In an indirect or a remote source coding problem, the encoder needs to describe the source by observing a different process, statistically related to the source, rather than the source itself [2, Sec. 4.5.4]. A general model for indirect source coding is described by the diagram in Fig. 10, where $P_{Y|X}$ represents a general ‘test channel’ between the source $X$ and the observable process $Y$. The indirect distortion-rate function (iDRF) describes the trade-off between information rate given on the observed process and the minimum average distortion between the source and its reconstruction. In order to get some insight into the nature of indirect source coding problems in Gaussian settings and the corresponding iDRF, it is instructive to start with a simple example.

**Example 2.** Consider two correlated sequences of i.i.d zero mean and jointly Gaussian random variables $U$ and $V$ with variances $C_U$, $C_V$, and covariance $C_{UV}$. We want to find the iDRF function of the source sequence $U$ given the observed sequence $V$, under a quadratic distortion measure. It follows from [22] that a compression scheme $V \to \hat{V}$
that achieves optimal minimal distortion for a given rate $\hat{R}$ bits per symbol is characterized by a joint Gaussian
distribution of $V$ and $\hat{V}$. Theorem 4.4.5 in [2] then implies a relation of the form
\[
\hat{V} = \alpha V + \zeta,
\] (28)
where $\alpha \in \mathbb{R}$ and $\zeta$ is a standard normal random variable independent of $U$ and $V$. The mutual information between
$V$ and $\hat{V}$ equals
\[
\hat{R} = I(\hat{V};V) = \frac{1}{2} \log \left( 1 + \alpha^2 C_V \right).
\] (29)

Since the distortion is quadratic, the optimal reconstruction is given by the conditional expectation of $U$ given $\hat{V}$,
\[
\hat{U} = \mathbb{E} [U|\hat{V}] = \frac{\alpha C_{UV}}{\alpha^2 C_V + 1} \hat{V}
\]
and the resulting minimal average distortion is
\[
D = \mathbb{E} [(U - \hat{U})^2] = C_U - \frac{\alpha^2 C_{UV}^2}{\alpha^2 C_V + 1}.
\]

Substituting $\alpha^2$ from (29), we obtain
\[
D_{U|V}(\hat{R}) = C_U - \left( 1 - 2^{-2\hat{R}} \right) \frac{C_{UV}^2}{C_V}
\] (30)
\[
= \text{mmse}_{U|V} + 2^{-2\hat{R}} C_{U|V},
\]
where $\text{mmse}_{U|V} = C_U - C_{U|V}$ is the MMSE in estimating $U$ from $V$ and $C_{U|V} \triangleq \frac{C_{UV}^2}{C_V}$ is the variance of the estimator
$\mathbb{E}[U|V]$. The equivalent rate-distortion function is
\[
\tilde{R}_{U|V}(D) = \begin{cases} 
\frac{1}{2} \log \frac{C_{UV}}{\text{mmse}_{U|V}} & C_U > D > \text{mmse}_{U|V}, \\
0 & D \geq C_U.
\end{cases}
\] (31)

Equation (30) can be intuitively interpreted as an extension of the regular DRF of a Gaussian i.i.d source $D_U(\hat{R}) = 2^{-2\hat{R}} C_U$ to the case where the information on the source is not entirely available at the encoder. Since $C_{U|V} \leq C_U$, the slope of $D_{U|V}(\hat{R})$ is more moderate than that of $D_U(\hat{R})$, which confirms the intuition that an increment in the bit-rate when describing noisy measurements is less effective in reducing distortion as the intensity of the noise increases.

If we denote $\theta = D - \text{mmse}_{U|V}$, then (30) and (31) are equivalent to
\[
\tilde{R}(\theta) = \frac{1}{2} \log \left( \frac{C_{UV}}{\theta} \right),
\] (32a)
and
\[
D(\theta) = \text{mmse}_{U|V} + \min \{ C_{U|V}, \theta \}
\] (32b)
\[
= C_U - [C_{U|V} - \theta]^+,
\]
i.e., we express the DRF $D_{U|V}(\hat{R})$, or the equivalent rate-distortion $\tilde{R}_{U|V}(D)$, through a joint dependency of $D$ and $\hat{R}$ on the parameter $\theta$. 

The relation between $V$ and $\hat{V}$ under optimal quantization can also be described by the ‘backward’ Gaussian channel

$$\frac{C_{UV}}{C_V} V = \hat{V} + \xi,$$

(33)

where $\xi$ is a zero-mean normal random variable independent of $\hat{V}$ with variance $\min \{ C_{U|V}, \theta \}$. The random variable $\xi$ can be understood as the part of the observable process that is lost due to lossy compression discounted by the factor $C_{UV}/C_V$. Since $\frac{C_{UV}}{C_V} V = \mathbb{E}[U|V]$, it also suggests that an optimal source code can be achieved by two separate steps:

(i) MMSE estimation of the source given the observed variable $V$.
(ii) Optimal direct source coding for the MMSE estimator $\mathbb{E}[U|V]$.

Although in this example the parametric representation (32) is redundant, we shall see below that this representation, the backward Gaussian channel (33) and the decomposition of distortion into an MMSE part plus the regular DRF of the estimator are repeating motifs of indirect source coding problems in the Gaussian setting.

Next, we consider indirect source coding in the more general case where the source and the observable process are jointly Gaussian and stationary. This problem can be obtained from our general model in Fig. 4 if we assume that $Z(\cdot)$ can be recovered from the samples $Y[\cdot]$ with zero error. In this case, the iDRF of $X(\cdot)$ given $Y[\cdot]$ reduces to the iDRF of $X(\cdot)$ given $Z(\cdot)$, which we denote as $D_{X|Z}(R)$. An expression for $D_{X|Z}(R)$ was first found by Dubroshin and Tsybakov in [7].

**Theorem 7** (Dobrushin and Tsybakov [7]). Let $X(\cdot)$ and $Z(\cdot)$ be two jointly stationary Gaussian stochastic processes
with spectral densities $S_X(f)$, $S_Z(f)$, and joint spectral density $S_{XZ}(f)$. The indirect distortion-rate function of $X(\cdot)$ given $Z(\cdot)$ is given by

$$R(\theta) = \frac{1}{2} \int_{-\infty}^{\infty} \log^+ \left[ S_{XZ}(f) \theta^{-1} \right] df,$$

$$D_{X|Z}(\theta) = \operatorname{mmse}_{X|Z} \left( \int_{-\infty}^{\infty} \min \{ S_{XZ}(f), \theta \} df \right)$$

$$= \int_{-\infty}^{\infty} S_X(f) df - \int_{-\infty}^{\infty} [S_{XZ}(f) - \theta]^+ df,$$

where

$$S_{XZ}(f) \triangleq \frac{|S_{XZ}(f)|^2}{S_Z(f)} = \frac{S_X^2(f) |H(f)|^2}{S_X + \eta(f) |H(f)|^2}$$

is the spectral density of the MMSE estimator of $X(\cdot)$ from $Z(\cdot)$, $[x]^+ = \max \{ x, 0 \}$, and

$$\operatorname{mmse}_{X|Z} = \int_{-\infty}^{\infty} \mathbb{E}(X(t) - \mathbb{E}[X(t)|Z(\cdot)])^2 dt$$

$$= \int_{-\infty}^{\infty} (S_X(f) - S_{X|Z}(f)) df$$

is the MMSE.

Remarks

(i) In (35) and in similar expressions henceforth, we interpret fractions as zero if both numerator and denominator are zero, i.e. (35) can be read as

$$\frac{S_X^2(f) |H(f)|^2}{S_X + \eta(f) |H(f)|^2} = \frac{S_X^2(f)}{S_X + \eta(f)} \mathbf{1}_{\text{supp} H}(f),$$

where $\mathbf{1}_{\text{supp} H}(f)$ is the indicator function of the support of $H(f)$.

(ii) Expressions of the form (34) are still correct if the spectral density $S_{X|Z}(f)$ includes Dirac delta functions. This is because the Lebesgue integral is not affected by infinite values on a set of measure zero. This is in accordance with the fact that periodic components can be determined for all times by specifying only their magnitude and phase, which requires zero information rate.

Equation (34) defines the function $D_{X|Z}(R)$ through a joint dependency of $D_{X|Z}$ and $R$ on the parameter $\theta$. The distortion is the sum of the MMSE in estimating $X(\cdot)$ from $Z(\cdot)$, plus a second term which has a water-filling interpretation. This is illustrated in Fig. [11] This expression generalizes the celebrated Shannon-Kolmogorov-Pinsker (SKP) reverse water filling for a single stationary Gaussian source [20], [23].

We can also describe the relation between the observable process $Z(\cdot)$ to its quantized version $\hat{Z}(\cdot)$ induced by the solution (34) by a backward Gaussian channel similar to (33):

$$\{ q_{X|Z} \ast Z \}(t) = \hat{Z}(t) + \xi(t), \quad t \in \mathbb{R},$$

where $\xi(\cdot)$ is a noise process independent of $\hat{Z}(\cdot)$ with spectral density $S_{\xi}(f) = \min \{ S_{XZ}(f), \theta \}$, and $q_{X|Z}(t)$ is the impulse response of the Wiener filter in estimating $X(\cdot)$ from $Z(\cdot)$ with corresponding frequency response.
\[ Q(f) = \frac{S_X(f)}{S_Z(f)}. \] The spectral counterpart of (36) is
\[ S_{X|Z}(f) = S_Z(f) + \min \{ S_{X|Z}(f), \theta \}. \]
This decomposition of \( S_{X|Z}(f) \) is seen in Fig. 11, where \( S_Z(f) \) corresponds to the preserved part, and \( \min \{ S_{X|Z}(f), \theta \} \) corresponds to the error due to lossy compression.

A. Separation Principle

Example 2 and Theorem 7 suggest a general structure for the solution of indirect source coding problems in Gaussian settings. The idea is that for general Gaussian sources \( U \) and \( V \), which here represent any two jointly Gaussian processes in any dimension, the Markov chain \( U \to V \to \hat{V} \to \hat{U} \) implies
\[
\mathbb{E} d(U, \hat{U}) = \| U - \hat{U} \|^2 = \text{mmse}_{UV} + \mathbb{E} \| \mathbb{E}[U|V] - \hat{U} \|^2 = \text{mmse}_{UV} + \mathbb{E} \| \mathbb{E}[E[U|V]|\hat{V}] - \mathbb{E}[U|V] \|^2 + \mathbb{E} \| \mathbb{E}[E[U|V]|\hat{V}] - \hat{U} \|^2.
\]
(37)
(38)
Taking the infimum over all mappings \( V \to \hat{V} \to \hat{U} \), we see that the map \( \hat{V} \to \hat{U} \) can be chosen such that the term in \( (38) \) vanishes without affecting the other terms. This implies that the iDRF of \( U \) given \( V \) is the sum of the MMSE in estimating \( U \) from the observations \( V \) and the distortion obtained by solving another indirect source coding problem, in which the input is the MMSE estimator \( \mathbb{E}[U|V] \) and the observable process is \( V \). Since \( \mathbb{E}[U|V] \) is a function of \( V \), the optimal solution to the last problem is obtained by first mapping \( V \) to \( \mathbb{E}[U|V] \) which is then compressed into \( \hat{V} \). This says that the new indirect source coding problem is equivalent to a direct source coding problem for the source \( \mathbb{E}[U|V] \).

To summarize, we have the following proposition:

**Proposition 1.** The iDRF of \( U \) given \( V \) under quadratic distortion can be written as
\[
D_{UV}(\bar{R}) = \text{mmse}_{UV} + D_{\mathbb{E}[U|V]}(\bar{R}),
\]
(39)
where \( D_{\mathbb{E}[U|V]}(\bar{R}) \) is the (direct) distortion-rate function of \( \mathbb{E}[U|V] \).

Proposition 1 was already established in [8] in the more general setting where \( U \) and \( V \) are not necessarily Gaussian. This was generalized in some sense to arbitrary distortion measures in [6].

We can now revisit Example 2 and Theorem 7 to observe that both are consequences of Proposition 1.

Going back to our general sampling model of Fig. 4 we can use Proposition 1 to write
\[
D(f_s, R) = \text{mmse}_{X|Y}(f_s) + D_{\tilde{X}}(R),
\]
where the process \( \tilde{X}(\cdot) \) is defined by
\[
\tilde{X}(\cdot) \triangleq \{ \mathbb{E}[X(t)|Y[\cdot]], t \in \mathbb{R} \}.
\]
This shows that the solution to our combined sampling and source coding problem is a sum of two terms. The first term is the MMSE in sub-Nyquist sampling already given by Theorem 4. The second term is the DRF of the process $\hat{X}(\cdot)$. Since $\hat{X}(\cdot)$ is not stationary, we currently do not have the means to find its DRF. This will be derived in Section V below.

### B. Discrete-Time Setting

In this subsection we present a general setting for indirect source coding problems of Gaussian discrete-time processes. A general model for such problems appears in Fig. 12:

[Y[:], X[:]] → Gaussian Channel → [\hat{Y}[:], \hat{X}[:]]

where the infimum is taken over all random mapping $\hat{Y}[:|Y[:]]$ such that $\mathbb{I}(\hat{Y}[:|Y[:]]) \leq \bar{R}$. Since for every given $\hat{Y}[:|Y[:]]$, (40) is minimized by the process $\hat{X}[:|\hat{Y}[:]] = \{\mathbb{E}[X[n]|\hat{Y}[:]], n \in \mathbb{Z}\}$, we can always assume $\hat{X}[n] = \mathbb{E}[X[n]|\hat{Y}[:]]$. The distortion-rate function in this case is defined to be

$$D(\bar{R}) = \inf \mathbb{E}d(X[:], \hat{X}[:]),$$

where $\mathbb{E}d(X[:], \hat{X}[:])$ is defined in general by

$$\mathbb{E}d(X[:], \hat{X}[:]) \triangleq ||X[:]-\hat{X}[:]||^2$$

$$= \limsup_{N \to \infty} \frac{1}{2N} \sum_{n=-N}^{N-1} \frac{1}{M} \sum_{i=1}^{M} \mathbb{E} (X_i[n] - \hat{X}_i[n])^2,$$

where $X_i[:], 1 \leq i \leq M, \text{ is the } i_{th} \text{ component of the vector-valued process } X[:].$ This last term reduces to $\frac{1}{M} \sum_{i=1}^{M} \mathbb{E} (X_i[0] - \hat{X}_i[0])^2$ if $X[:]$ and $\hat{X}[:]$ are stationary.

The distortion-rate function in this case is defined to be

$$D(\bar{R}) = \inf \mathbb{E}d(X[:], \hat{X}[:]),$$

where the infimum is taken over all random mapping $Y[:] \rightarrow \hat{Y}[:] \rightarrow \hat{X}[:]$ such that $\mathbb{I}(\hat{Y}[:|Y[:]]) \leq \bar{R}$. Since for every given $\hat{Y}[:|Y[:]]$, (40) is minimized by the process $\hat{X}[:|\hat{Y}[:]] = \{\mathbb{E}[X[n]|\hat{Y}[:]], n \in \mathbb{Z}\}$, we can always assume $\hat{X}[n] = \mathbb{E}[X[n]|\hat{Y}[:]]$. The distortion-rate function in this case is defined to be

$$D(\bar{R}) = \inf \mathbb{E}d(X[:], \hat{X}[:]),$$

where the infimum is taken over all random mapping $Y[:] \rightarrow \hat{Y}[:] \rightarrow \hat{X}[:]$ such that $\mathbb{I}(\hat{Y}[:|Y[:]]) \leq \bar{R}$. Since for every given $\hat{Y}[:|Y[:]]$, (40) is minimized by the process $\hat{X}[:|\hat{Y}[:]] = \{\mathbb{E}[X[n]|\hat{Y}[:]], n \in \mathbb{Z}\}$, we can always assume $\hat{X}[n] = \mathbb{E}[X[n]|\hat{Y}[:]]$. The distortion-rate function in this case is defined to be

$$D(\bar{R}) = \inf \mathbb{E}d(X[:], \hat{X}[:]),$$

where the infimum is taken over all random mapping $Y[:] \rightarrow \hat{Y}[:] \rightarrow \hat{X}[:]$ such that $\mathbb{I}(\hat{Y}[:|Y[:]]) \leq \bar{R}$. Since for every given $\hat{Y}[:|Y[:]]$, (40) is minimized by the process $\hat{X}[:|\hat{Y}[:]] = \{\mathbb{E}[X[n]|\hat{Y}[:]], n \in \mathbb{Z}\}$, we can always assume $\hat{X}[n] = \mathbb{E}[X[n]|\hat{Y}[:]]$. The distortion-rate function in this case is defined to be
If all processes are one-dimensional and jointly stationary, then the function $D(\tilde{R})$ can be obtained from Theorem [7] by replacing all factors by their discrete-time counterparts, which leads to

$$\tilde{R}(\theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ S_{Y|X} (e^{2\pi i \phi}) \theta^{-1} \right] d\phi,$$

and

$$D(\theta) = \text{mmse}_{X|Y} + \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \theta, S_{X|Y} (e^{2\pi i \phi}) \} d\phi$$

$$= \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ S_{X|Y} (e^{2\pi i \phi}) - \theta \right]^+ d\phi.$$

(C. Vector-valued Sources)

We now derive the counterpart of (42) for vector-valued processes. We recall that for the Gaussian stationary source $X[\cdot]$, the counterpart of the SKP reverse water-filling was given in [24, Eq. (20) and (21)], as

$$\bar{R}(\theta) = \sum_{i=1}^{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2} \log^+ \left[ \lambda_i(S_X) \theta^{-1} \right] d\phi,$$

and

$$D_X(\theta) = \frac{1}{M} \sum_{i=1}^{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \lambda_i(S_X), \theta \} d\phi,$$  

where $\lambda_1(S_X), \ldots, \lambda_M(S_X)$ are the eigenvalues of the spectral density matrix $S_X(e^{2\pi i \phi})$ at frequency $\phi$. Combining (43) with the separation principle allows us to extend Theorem [7] to Gaussian vector processes.

**Theorem 8.** Let $X[\cdot] = (X_1[\cdot], \ldots, X_M[\cdot])$ be an $M$ dimensional vector-valued Gaussian stationary stochastic process, and let $Y[\cdot]$ be another vector valued process such that $X[\cdot]$ and $Y[\cdot]$ are jointly Gaussian and jointly stationary. The distortion-rate function of $X[\cdot]$ given $Y[\cdot]$ under quadratic distortion is given by

$$\bar{R}(\theta) = \sum_{i=1}^{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2} \log^+ \left[ \lambda_i(S_{X|Y}) \theta^{-1} \right] d\phi,$$

$$D(\theta) = \text{mmse}_{X|Y} + \frac{1}{M} \sum_{i=1}^{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \lambda_i(S_{X|Y}), \theta \} d\phi$$

$$= \frac{1}{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{Tr} S_X (e^{2\pi i \phi}) d\phi$$

$$- \frac{1}{M} \sum_{i=1}^{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ \lambda_i(S_{X|Y}) - \theta \right]^+ d\phi,$$

where $\lambda_1(S_{X|Y}), \ldots, \lambda_M(S_{X|Y})$ are the eigenvalues of

$$S_{X|Y} (e^{2\pi i \phi}) \triangleq S_{XY} (e^{2\pi i \phi}) S_Y^{-1} (e^{2\pi i \phi}) S_{XY}^* (e^{2\pi i \phi}),$$
which is the spectral density matrix of the MMSE estimator of $X[\cdot]$ from $Y[\cdot]$. Here $\text{mmse}_{X|Y}$ is given by

$$
\text{mmse}_{X|Y} = \frac{1}{M} \int \frac{1}{2} \text{Tr} \left( S_X(e^{2\pi i \phi}) - S_{X|Y}(e^{2\pi i \phi}) \right) d\phi
$$

$$
= \frac{1}{M} \sum_{i=1}^{M} \text{mmse}_{X|Y}.
$$

**Proof:** This is an immediate consequence of Proposition 1 and Equations (43a) and (43b).

### D. Lower Bound

Throughout this subsection we suppress the time index and allow the processes considered to have either continuous or discrete time indexes. We also use $R$ to represents either bits per time unit or bits per symbol, according to the time index.

In Section [11] we exploited the fact that the polyphase components $X_\Delta[\cdot]$ defined in (8) and the process $Y[\cdot]$ are jointly Gaussian to compute the MMSE of $X(\cdot)$ given $Y[\cdot]$. This was possible since the overall MMSE is given by averaging the MMSE in estimating each one of the polyphase components $X_\Delta[\cdot]$ over $0 \leq \Delta < 1$, as expressed by (9). Unfortunately, the iDRF does not satisfy such an averaging property in general. Instead, we have the following proposition, which holds for any source distribution and distortion measure.

**Proposition 2.** (average distortion bound) Let $X$ and $Y$ be two vector-valued sources. The iDRF of $X$ given $Y$, under distortion measure $\tilde{d}$, satisfies

$$
D_{X|Y}(R) \geq \frac{1}{M} \sum_{i=1}^{M} D_{X|Y_i}(R),
$$

(44)

where we denoted $X = (X_1, \ldots, X_M)$ and $\tilde{d}(X, \hat{X}) = \frac{1}{M} \sum_{i=1}^{M} d(X_i, \hat{X}_i)$.

**Proof:** This is merely an operational statement: for every $i = 1, \ldots, M$, the minimal distortion obtained by introducing an optimal code of rate $R$ to describe $X_i$ is smaller than the distortion at the $i$th coordinate of $X$, by a code of the same rate that was designed to minimize the average over all coordinates.

Note that $D_{X|Y}(\infty) = \text{mmse}_{X|Y} = \frac{1}{M} \sum_{i=1}^{M} D_{X_i|Y}(\infty)$ and $D_{X|Y}(0) = \sigma_X^2 = \frac{1}{M} \sum_{i=1}^{M} D_{X_i|Y}(0)$, i.e. the bound is always tight for $R = 0$ and $R \to \infty$.

The proof of Proposition 2 implies that equality in the bound (44) is achieved when the optimal indirect rate-$R$ code for the vector process $X$ induces an indirect optimal rate-$R$ code on for each one of the coordinates. This is the case if the $M$ optimal indirect rate-$R$ codes for each coordinate are all functions of a single indirect rate-$R$ code. Indeed, this is the case for $R \to \infty$, since then any code essentially describes $E[X(\cdot)|Y[\cdot]]$ which is a sufficient statistic for the MMSE reconstruction problem. Another case of equality is described in the following example.

**Example 3 (i.i.d Vector Source).** Let $U = (U_1, \ldots, U_M)$ and $V = (V_1, \ldots, V_P)$ be two i.i.d jointly Gaussian vector sources with covariance matrices $C_U$, $C_V$, and $C_{UV}$. In order to find the iDRF of $U_m$ given $V$, for $m = 1, \ldots, M$,
we can use Proposition \[ \textbf{7} \] and obtain

\[
D_U|V(\bar{R}) = \text{mmse}_{U|V} + 2^{-2\bar{R}}C_{U|V},
\]

where we used the fact that the distortion-rate function of the Gaussian random i.i.d source \( \mathbb{E}[U_m|V] = 2^{-2\bar{R}}C_{U_m|V}C_V^{-1}C_{U_m|V}^T. \) The bound \[ \textbf{44} \] implies

\[
D_{U|V}(\bar{R}) \geq \frac{1}{M} \sum_{m=1}^{M} \left( \text{mmse}_{U_m|V} + 2^{-2\bar{R}}C_{U_m|V} \right)
= \text{mmse}_{U|V} + \frac{1}{M} 2^{-2\bar{R}} \text{Tr} C_{U|V}
= \text{mmse}_{U|V} + \frac{1}{M} 2^{-2\bar{R}} \sum_{m=1}^{P \wedge M} \lambda_i(C_{U|V}),
\]

\[ (45) \]

where \( P \wedge M = \min \{ P, M \} \) is the maximal rank of the matrix \( C_{U|V}. \) We compare \[ \textbf{45} \] to the true value of the iDRF of \( U \) given \( V, \) which is obtained using Theorem \[ \textbf{8} \]

\[
\bar{R}(\theta) = \frac{1}{2} \sum_{i=1}^{P \wedge M} \log^+ \left( \lambda_i(C_{U|V})^{-1} \right),
\]

\[ D_{U|V}(\theta) = \text{mmse}_{U|V} + \frac{1}{M} \sum_{i=1}^{P \wedge M} \min \left\{ \lambda_i(C_{U|V}), \theta \right\}. \]

\[ (46) \]

From \[ \textbf{45} \] and \[ \textbf{46} \] we conclude that for any \( \bar{R} \geq 0, \)

\[
\sum_{i=1}^{P \wedge M} \min \left\{ \lambda_i(C_{U|V}), \theta \right\} \geq 2^{-2\bar{R}} \sum_{m=1}^{P \wedge M} \lambda_i(C_{U|V}),
\]

\[ (47) \]

where

\[
\bar{R}(\theta) = \frac{1}{2} \sum_{i=1}^{P \wedge M} \log^+ \left( \lambda_i(C_{U|V})^{-1} \right).
\]

If \( P = 1, \) then \( C_{U|V} \) has a single non-zero eigenvalue and equality holds in \[ \textbf{47} \]. As will be seen by the next example, equality in \[ \textbf{44} \] when the observable process is one-dimensional is indeed special to the i.i.d case. The next example will also be later used to prove a lower bound for the combined sampling and source coding problem in Theorem \[ \textbf{9} \]

**Example 4** (vector stationary source). Let \( X[.] = (X_1[.], \ldots, X_M[.]) \) be a Gaussian stationary vector source and let \( Y[.] \) be a one-dimensional process jointly Gaussian and stationary with \( X[.]. \) The iDRF of \( X[.] \) given \( Y[.] \) is given by Theorem \[ \textbf{8} \] to be

\[
D_{X|Y}^{id}(\theta) = \text{mmse}_{X|Y} + \frac{1}{M} \int_{\frac{-1}{2}}^{\frac{1}{2}} \min \left\{ \sum_{m=1}^{M} S_{X_m|Y}(e^{2\pi i \phi}), \theta \right\} d\phi,
\]

\[
R(\theta) = \frac{1}{2} \int_{\frac{-1}{2}}^{\frac{1}{2}} \log^+ \left[ \sum_{m=1}^{M} S_{X_m|Y}(e^{2\pi i \phi}) \theta^{-1} \right] d\phi,
\]

where we used the fact that the rank of \( S_{X|Y}(e^{2\pi i \phi}) \) is at most one, and thus the sum of the eigenvalues of \( S_{X|Y}(e^{2\pi i \phi}) \) equals its trace, which is given by \( \sum_{m=1}^{M} S_{X_m|Y}(e^{2\pi i \phi}). \) Considering the \( m \)\textsuperscript{th} coordinate of \( X[.]. \) separately, the iDRF
of $X_m[\cdot]$ given $Y[\cdot]$ is

$$D_{X|Y}(\theta) = \text{mmse}_{X_{m}|Y} + \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ S_{X_{m}|Y}(e^{2\pi i \phi}), \theta_m \} d\phi,$$

where

$$R(\theta_m) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [S_{X_{m}|Y}(e^{2\pi i \phi}) \theta_m^{-1}] d\phi.$$  

Since $\text{mmse}_{X|Y} = \frac{1}{M} \sum_{m=1}^{M} \text{mmse}_{X_{m}|Y}$, the bound (44) implies

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=1}^{M} \min \{ S_{X_{m}|Y}(e^{2\pi i \phi}), \theta_m(R) \} d\phi$$

$$\leq \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ \sum_{m=1}^{M} S_{X_{m}|Y}(e^{2\pi i \phi}), \theta(R) \right\} d\phi.$$  

V. DISTORTION-RATE FUNCTION AT A GIVEN SAMPLING FREQUENCY

In this section we solve our main source coding problem for the case of a single branch sampling. We will derive a closed form expression for the function $D(f_s, R)$ defined in (4) and for its minimal value over all pre-sampling filters $H(f)$. From the definition of $D(f_s, R)$ in Section II we can already deduce the following facts about $D(f_s, R)$:

Proposition 3.

(i) For all $f_s > 0$ and $R \geq 0$,

$$D(f_s, R) \geq D_{X|Z}(R),$$

where $D_{X|Z}(R)$ was defined in (54). In addition,

$$D(f_s, R) \geq \text{mmse}_{X|Y}(f_s),$$

where $\text{mmse}_{X|Y}(f_s)$ is the MMSE in reconstructing $X(\cdot)$ from the uniform samples $Y[\cdot]$ given in Theorem I.

(ii) If the process $Z(\cdot)$ has almost surely Riemann integrable realizations, then the reconstruction error of $Z(\cdot)$ from $Y[\cdot]$ can be made arbitrarily small by sampling at a high enough frequency. It follows that as $f_s$ goes to infinity, $D(f_s, R)$ converges to $D_{X|Z}(R)$. In particular, if $Z(\cdot)$ is band-limited, then $D(f_s, R) = D_{X|Z}(R)$ for any $f_s$ above the Nyquist frequency of $Z(\cdot)$.

(iii) For a fixed $f_s > 0$, $D(f_s, R)$ is a monotone non-increasing function of $R$ which converges to $\text{mmse}_{X|Y}(f_s)$ as $R$ goes to infinity. It is not necessarily non-increasing in $f_s$ since $\text{mmse}_{X|Y}(f_s)$ is not necessarily non-increasing in $f_s$.

Note that $Z(\cdot)$ does not need to be band-limited. The only assumption on $Z(\cdot)$ is finite variance, i.e. $S_Z(f)$ is in $L_1$. 


A. Lower Bound

Note that (i) in Proposition 3 implies that the manifold defined by $D(f_s, R)$ in the three dimensional space $(f_s, R, D)$ is bounded from below by the two cylinders $\text{mmse}_{X|Y}(f_s)$ and $D_{X|Z}(R)$ (and from above by the plane $D = \sigma_Y^2$). A tighter lower bound is obtained by the bound in Proposition 3.

**Theorem 9.** We have the following bound:

$$D(f_s, R) \geq \text{mmse}_{X|Y}(f_s)$$

$$+ \int_0^\Delta \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ S_{X|Y} \left( e^{2\pi i \phi} \right), \theta_\Delta \right\} d\phi d\Delta,$$

where

$$S_{X|Y} \left( e^{2\pi i \phi} \right) = \frac{\sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} S_{XZ}(f_s(\phi - k)) S_{Y}(f_s(\phi - l)) e^{2\pi i (k-l)\Delta}}{\sum_{k \in \mathbb{Z}} S_{Z}(f_s(\phi - k))^{-1}},$$

and for each $0 \leq \Delta \leq 1$, $\theta_\Delta$ satisfies

$$\bar{R} = R/f_s = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ S_{X|Y} \left( e^{2\pi i \phi} \right), \theta_\Delta^{-1} \right] d\phi.$$

**Proof:** For a given finite set of points $\Delta_1, \ldots, \Delta_M$ in $[0, 1)$ define the vector values process

$$X^M[n] = (X_{\Delta_1}[n], \ldots, X_{\Delta_M}[n]), \quad n \in \mathbb{Z},$$

where for $m = 1, \ldots, M$, the discrete time process $X_{\Delta_m}[:]$ is defined in (9). By Proposition 3 we have

$$D_{X^M|Y}(\bar{R}) \geq \frac{1}{1} \sum_{m=1}^{M} D_{X_{\Delta_m}|Y}(\bar{R}).$$

It follows from the proof of Theorem 2 that for all $m = 1, \ldots, M$, $X_{\Delta_m}[:]$ and $Y[:]$ are jointly Gaussian and stationary, with $S_{X_{\Delta_m}|Y} \left( e^{2\pi i \phi} \right)$ given by (12). Applying the discrete-time version of Theorem 2, i.e. (42), we get that the iDRF of $X_{\Delta_m}[:]$ given $Y[:]$ is

$$D_{X_{\Delta_m}|Y}(\bar{R}) = \text{mmse}_{X|Y} + \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ S_{X_{\Delta_m}|Y} \left( e^{2\pi i \phi} \right), \theta_{\Delta_m} \right\} d\phi,$$

where for a fixed $\bar{R}$, $\theta_{\Delta_m}$ satisfies

$$\bar{R}(\theta_{\Delta_m}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ S_{X_{\Delta_m}|Y} \left( e^{2\pi i \phi} \right), \theta_{\Delta_m} \right] d\phi.$$

Since (12) is a continuous function of $\Delta$, it follows that $D_{X_{\Delta}}(\bar{R})$ is a continuous function of $\Delta$ and hence integrable. As the number of points $M$ goes to infinity with vanishing division parameter $\max_{m_1 \neq m_2} |\Delta_{m_1} - \Delta_{m_2}|$, the LHS of (49) is equivalent to $D_{X|Y}(\bar{R}) = D(f_s, R)$ and its RHS converges to the integral of (50) over the interval $(0, 1)$. Using (9), (48) follows.
B. The Discrete-time Case

In order to solve our main source coding problem, we first solve its discrete-time counterpart. Here the underlying process is $X[\cdot]$ and we observe a factor $M$ down-sampled version of the discrete time process $Z[\cdot]$, which is jointly Gaussian and jointly stationary with $X[\cdot]$. Note that unlike what was discussed in Section IV, the source process and the observable process are no longer jointly stationary.

**Theorem 10** (single branch decimation). Let $X[\cdot]$ and $Z[\cdot]$ be two jointly Gaussian stationary processes. Given $M \in \mathbb{N}$, define the process $Y[\cdot]$ by $Y[n] = Z[Mn]$, for all $n \in \mathbb{Z}$. The indirect distortion-rate function of $X[\cdot]$ given $Y[\cdot]$, under the quadratic distortion (40), is given by

$$\bar{R}(M, \theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left( J_M(e^{2\pi i \phi}) \theta^{-1} \right) d\phi,$$

$$D(M, \theta) = \text{mmse}_{X|Y}(M) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ J_M(e^{2\pi i \phi}), \theta \right\} d\phi$$

where

$$J_M(e^{2\pi i \phi}) \doteq \frac{1}{M} \sum_{m=0}^{M-1} \left| \mathbb{S}_{XZ} \left( e^{2\pi i \phi \frac{m}{M}} \right) \right|^2,$$

and $\text{mmse}_{X|Y}(M)$ is defined by

$$\lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E} (X[n] - \mathbb{E}(X[n]|Y[\cdot]))^2$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \mathbb{E} (X[n] - \mathbb{E}(X[n]|Y[\cdot]))^2.$$

**Proof:** While the details can be found in Appendix XI, an outline of the proof is as follows: given $M \in \mathbb{N}$, define the vector-valued process $X^M[\cdot]$ by

$$X^M[n] \doteq (X[Mn], X[Mn+1], \ldots, X[Mn+M-1]), \quad n \in \mathbb{Z}.$$

Both processes $X[\cdot]$ and $X^M[\cdot]$ consists of the same data with the same entropy rate hence share the same indirect distortion-rate function given $Y[\cdot]$. Since $X[\cdot]$ and $Y[\cdot]$ are jointly Gaussian and stationary, the result follows by applying Theorem 8.

C. Main Result: Single Branch Sampling

We are now ready to solve our combined sampling and source coding problem introduced in Section II. Note that here we go back to the model of Fig. 4 with the single branch sampler of Fig. 5(a).

**Theorem 11** (single branch sampling). Let $X(\cdot)$ and $Z(\cdot)$ be two jointly Gaussian stationary stochastic processes with almost surely Riemann integrable realizations and $L_1$ PSD’s $S_X(f)$, $S_Z(f)$ and $S_{XZ}(f)$. Let $Y[\cdot]$ be the discrete
The time process defined by $Y[n] = Z(n/f_s)$, where $f_s > 0$. The indirect distortion-rate function of $X(\cdot)$ given $Y[\cdot]$, is given by

$$R(f_s, \theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ [J(f) \theta^{-1}] df;$$

(51a)

$$D(f_s, \theta) = \text{mmse}_{X|Y}(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min \{J(f), \theta\} df$$

(51b)

where

$$J(f) = \frac{\sum_{k \in Z} |S_{XZ}(f - f_s k)|^2}{\sum_{k \in Z} S_Z(f - f_s k)},$$

and

$$\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} J(f) df.$$

Proof: see Appendix XII. The basic idea of the proof is to approximate the continuous time processes $X(\cdot)$ and $Z(\cdot)$ by discrete time processes, and take the limit in the solution to the discrete problem given by Theorem 10.

D. Discussion

We see that for a given sampling frequency $f_s$, the optimal solution has a similar form as in the stationary case (34) and Theorem 7, where the function $J(f)$ takes the role of $S_{X|Z}(f)$. That is, the minimal distortion is obtained by a MMSE term plus a term determined by reverse water-filling over the function $J(f)$, which can be described by Fig. 11 if we replace $S_{X|Y}(f)$ by $J(f)$. By comparing equations (51b) and (39), we have the following interpretation of the second term in (51b):

**Proposition 4.** The (direct) distortion-rate function of the non-stationary process $\tilde{X}(\cdot) = \{E[X(t)|Y[\cdot]], t \in \mathbb{R}\}$ is given by

$$R(\theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ [J(f) \theta^{-1}] df;$$

$$D_{\tilde{X}}(\theta) = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min \{J(f), \theta\} df.$$
Note that the function \( J(f) \) is periodic, and in (14) it has been expressed as the average of the spectral densities \( S_{X|Z}(e^{2\pi i \theta}) \) over \( 0 \leq \Delta \leq 1 \). It follows that \( J(e^{2\pi i \theta}) \triangleq f_s J(\phi(f_s)) \) is the spectral density of a discrete-time stationary process which reflects some time averaging over the non stationary process \( \hat{X}(\cdot) \), in the sense that both processes share the same distortion-rate function.

The function \( J(f) \) depends on the sampling frequency, the filter \( H(f) \) and the spectral densities \( S_X(f) \) and \( S_\eta(f) \), but is independent of \( R \). If we fix \( R \) and consider a change in \( J(f) \) such that \( \int_{-\frac{R}{2}}^{\frac{R}{2}} J(f) df \) is increased, then from (51a) we see that \( \theta \) also increases to maintain the same fixed rate \( R \). On the other hand, the expression for \( D(f_s,R) \) in (51b) exhibits a negative linear dependency on \( f_s \int_{-\frac{R}{2}}^{\frac{R}{2}} J(f) df \). In this game between the two terms in (51b), the negative linear dependency in \( J(f) \) is stronger then a logarithmic dependency of \( \theta \) in \( \int_{-\frac{R}{2}}^{\frac{R}{2}} J(f) df \) and the distortion reduces with an increment in \( \int_{-\frac{R}{2}}^{\frac{R}{2}} J(f) df \). The exact behavior is obtained by taking the functional derivative of \( D(f_s,R) \) with respect to \( J \) at the point \( f \in \left( -\frac{R}{2}, \frac{R}{2} \right) \), which is non-positive. A simple analogue for that dependency can be seen in Example 2 where the distortion in (30) is a non-increasing function of \( C_{U|Y} \). We summarize the above in the following proposition:

**Proposition 5.** For a fixed \( R \geq 0 \), minimizing \( D(f_s,R) \) is equivalent to maximizing \( \int_{-\frac{R}{2}}^{\frac{R}{2}} J(f) df \).

This says that a larger \( J(f) \) accounts for more information available on the source through the samples \( Y[\cdot] \), and motivates us to bound \( J(f) \). Since \( J(f) \) can be written as

\[
J(f) = \sum_{k \in \mathbb{Z}} S_{X|Z}(f - f_s k) S_Z(f - f_s k) \sum_{k \in \mathbb{Z}} S_Z(f - f_s k),
\]

the following holds for almost every \( f \in \left( -\frac{R}{2}, \frac{R}{2} \right) \),

\[
J(f) \leq \sup_k S_{X|Z}(f - f_s k)
\]

\[
= \sup_k \frac{S^2_X(f - f_s k)}{S_X(f - f_s k)H(f - f_s k)^2}
\]

\[
= \sup_k \frac{S_X^2(f - f_s k)}{S_{X+\eta}(f - f_s k)},
\]

with equality if and only if for each \( k \in \mathbb{Z} \), either \( S_{X|Z}(f - f_s k) = \sup_k S_{X|Z}(f - f_s k) \) or \( S_Z(f - f_s k) = 0 \). Thus, we have the following proposition.

**Proposition 6.** For all \( f_s > 0 \) and \( R \geq 0 \),

\[
D(f_s,R) \geq D^*(f_s,R),
\]

where \( D^*(f_s,R) \) is the distortion-rate function of the Gaussian stationary continuous time process with spectrum

\[
J^*(f) = \begin{cases} 
\sup_k \frac{S^2_X(f - f_s k)}{S_X(f - f_s k)H(f - f_s k)^2}, & f \in \left( -\frac{R}{2}, \frac{R}{2} \right), \\
0, & \text{otherwise}.
\end{cases}
\]
Note that the last expression is independent of the pre-sampling filter $H(f)$, thus Proposition 6 describes a lower bound which depends only on the statistics of the source and the noise. We will see in Theorem 14 below that $D^*(f_s, R)$ is attainable for any given $f_s$, if we are allowed to choose the pre-sampling filter $H(f)$.

It is interesting to observe how Theorem 11 agrees with the properties of $D(f_s, R)$ in the two special cases illustrated in Fig. 2:

(i) For $f_s$ above the Nyquist frequency of $Z(\cdot)$, $S_Z(f-f,k) = 0$ for any $k \neq 0$. In this case the conditions for equality in \(54\) hold and

$$J(f) = \sup_k \frac{S_X^2(f-f,k)}{S_X+\eta(f-f,k)} = \frac{S_X^2(f)}{S_X+\eta(f)},$$

which means that \(51\) is equivalent to \(34\).

(ii) If we take $R$ to infinity, then $\theta$ goes to zero and $51b$ reduces to $10$.

In view of the above we see that Theorem 11 subsumes the two classical problems of finding $\text{mmse}_{X|Y}(f_s)$ and finding $D_{X|Z}(R)$.

E. Examples

In Examples 5 and 6 below we find a single letter expression for the function $D(f_s, R)$ under a given PSD $S_X(f)$, zero noise $S_\eta(f) \equiv 0$ and unit pre-sampling filter $|H(f)| \equiv 1$, i.e. when $S_X(f) = S_Z(f)$.

**Example 5** (Rectangular spectrum).

Let the spectrum of the source $X(\cdot)$ be

$$S_X(f) = \begin{cases} \frac{\sigma^2}{2W} & |f| \leq W, \\ 0 & \text{otherwise}, \end{cases}$$

for some $W > 0$ and $\sigma > 0$. For all frequencies $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$,

$$J(f) = \frac{\sum_{k \in \mathbb{Z}} S_X^2(f-f,k)}{\sum_{k \in \mathbb{Z}} S_X(f-f,k)} = \frac{\sigma^2}{2W} \begin{cases} 1 & |f| < W, \\ 0 & |f| \geq W, \end{cases}$$

thus

$$\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} J(f) \, df = \sigma^2 \min \left\{ \frac{f_s}{2W}, 1 \right\}.$$  

By Theorem 11 we have

$$R(f_s, \theta) = \begin{cases} \frac{f_s}{2} \log \left( \frac{f_s \sigma^2}{2W} \right) & 0 \leq \frac{\theta}{\sigma^2} \leq \frac{f_s}{2W} \leq 1, \\ \frac{f_s}{2} \log \left( \frac{\sigma^2}{W} \right) & 0 \leq \frac{\theta}{\sigma^2} < 1 \leq \frac{f_s}{2W}, \\ 0 & \text{otherwise}, \end{cases}$$

and
Fig. 13: Distortion as a function of sampling frequency $f_s$ and compression rate $R = 1 \text{[bit/sec]}$, for a process with spectral density function depicted in the small frame. Here the noise is taken to be zero and $|H(f)| \equiv 1$. The dashed line represents the DRF of the source, which coincides with $D(f_s, R)$ for $f_s$ above the Nyquist frequency.

This can be written in a single expression as

$$D(f_s, R) = \begin{cases} \frac{\sigma^2}{2} \left( 1 - \frac{f_s}{2W} \right)^2 + \frac{\theta}{\sigma^2} & \frac{\theta}{\sigma^2} \leq \min \left\{ \frac{f_s}{2W}, 1 \right\} \\ \frac{\sigma^2}{2} - \frac{R}{W} & \frac{f_s}{2W} \geq 1. \end{cases}$$

(55)

Expression (55) has a very intuitive structure: for frequencies below the Nyquist frequency of the signal, the distortion as a function of the rate increases by a constant factor due to the error as a result of non-optimal sampling. This factor completely vanishes for $f_s$ greater than the Nyquist frequency of the signal, in which case $D(f_s, R)$ equals the (direct) distortion-rate function of the process $X(\cdot)$ given in this case by $D_X(R) = \sigma^2 2^{-R/W}$. This is depicted in Fig. 13.

Example 6.

The following example shows that the distortion-rate function is not necessarily monotone in the sampling frequency. Here $S_X(f)$ has the band-pass structure

$$S_X(f) = \begin{cases} \frac{\sigma^2}{2} & \frac{1}{2} \leq |f| \leq \frac{3}{2}, \\ 0 & \text{otherwise}, \end{cases}$$

(56)
and we still assume $S_X(f) = S_Z(f)$. We again obtain that for any $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$, $J(f)$ is either $\frac{\sigma^2}{2}$ or 0. Thus, in order to find $D(f_s, R)$, all we need to know at a given sampling frequency $f_s$ are for which values of $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$ the function $J(f)$ vanishes. This leads to

$$D(f_s, R) = \sigma^2 \left\{ \begin{array}{ll}
2^{-R} & 3 \leq f_s, \\
1 - \frac{f_s - 1}{2} \left(1 - 2^{-\frac{2\pi}{f_s}}\right) & 2 \leq f_s < 3, \\
1 - \frac{3 - f_s}{2} \left(1 - 2^{-\frac{2\pi}{f_s}}\right) & 1.5 \leq f_s < 2, \\
1 - \frac{f_s}{2} \left(1 - 2^{-\frac{2\pi}{f_s}}\right) & 0 \leq f_s < 1.5,
\end{array} \right.$$ 

which is depicted in Fig. 14.

**F. Optimal pre-Sampling Filter**

An optimization similar to the one carried out in Subsection V-B over the pre-sampling filter $H(f)$ can be carried out over the function $J(f)$ in order to minimize the function $D(f_s, R)$. By Proposition 5 minimizing distortion for a given $f_s$ and $R$ is equivalent to maximizing $J(f)$ for every $f \in \left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$ separately. But recall that the optimal pre-sampling filter $H^*(f)$ that maximizes $J(f)$ was already given in Theorem 3 in terms of the maximal aliasing free set associated with $\frac{S_X(f)}{S_{X+\eta}(f)}$, which leads us to the following conclusion:

**Theorem 12.** Given $f_s > 0$, the optimal pre-sampling filter $H^*(f)$ that minimizes $D(f_s, R)$, for all $R \geq 0$, is given by

$$H^*(f) = \begin{cases}
1 & f \in F^*, \\
0 & \text{otherwise},
\end{cases}$$
where $F^*$ is in $\mathcal{A}(f_s)$ and satisfies
\[
\int_{F^*} \frac{S_X^2(f)}{S_{X+\eta}(f)} \, df = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sup_k \frac{S_X^2(f-f,k)}{S_{X+\eta}(f-f,k)} \, df.
\]
The maximal value of $J(f)$ obtained this way is
\[
J^*(f) = \sup_k \frac{S_X^2(f-f,k)}{S_{X+\eta}(f-f,k)},
\]
and the distortion-rate function at a given sampling frequency is given by
\[
R^*(f_s, \theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ J^*(f) \theta^{-1} \right] \, df \tag{57a}
\]
\[
= \frac{1}{2} \int_{F^*} \log^+ \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} \theta^{-1} \right] \, df,
\]
\[
D^*(f_s, \theta) = \sigma_X^2 - \int_{\frac{1}{2}^+} \left[ J^*(f) - \theta \right]^+ \, df \tag{57b}
\]
\[
= \sigma_X^2 - \int_{F^*} \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} - \theta \right]^+ \, df.
\]

Proof: From Theorem 3 we conclude that the filter $H^*(f)$ that maximizes $J(f)$ is given by the indicator function of the maximal aliasing free set $F^*$. Moreover, with this optimal filter, (51) reduces to (57).

We emphasize that even in the absence of noise, the filter $H^*(f)$ still plays a crucial role in reducing distortion by preventing aliasing as described in Subsection III-B. Fig. 15 illustrates the effect of optimal pre-sampling filtering on the function $D(f_s, R)$.

VI. MULTI-BRANCH SAMPLING

We now generalize our analysis to the case where the sampling operation can be described by a multi-branch sampler as given in Fig. 5(b). Similar to the case of single branch sampling, we first prove the discrete-time counterpart and use it to prove our main result.

A. Multi-Branch Decimation

In the discrete-time counterpart of our source coding problem with multi-branch sampling, the source is the discrete-time process $X[\cdot]$ and the sampling operation at each branch is replaced by decimation by a factor $PM$, where $P \in \mathbb{N}$ is the number of sampling branches and $M \in \mathbb{N}$ is the average number of time units at which $Y[\cdot]$ samples $X[\cdot]$.

Theorem 13 (Discrete Multi-Branch Sampling). For $M \in \mathbb{N}$ and $p = 1, \ldots, P$, let $Y_p[\cdot]$ be an $PM$ factor decimation of the process $Z_p[\cdot]$, namely,
\[
Y[n] = (Z_1[PMn], \ldots, Z_P[PMn]),
\]
Fig. 15: $D^*(f_s, R)$ and $D(f_s, R)$ at two fixed values of $R$. These are obtained using the optimal pre-sampling filter ($H(f) = H^*(f)$) and without ($|H(f)| \equiv 1$) for the same source statistic with $S_\eta(f) \equiv 0$ and $S_X(f)$ as given in the small frame.

Fig. 16: $R(f_s, D)$ for the process PSD given in the small frame at a given value of $D$. The cutoffs correspond to values of $f_s$ at which $\text{mmse}_{X|Y}(f_s) > D$, i.e., outside the domain of $R(f_s, D)$. The difference between each curve to $R_{X|Z}(D)$ is said to represent the amount of information lost due to the sampling system that corresponds to that curve.
where \( X[\cdot] \) and \( Z_p[\cdot] \) are jointly Gaussian stationary processes with spectral densities

\[
S_{Z_p}(e^{2\pi i \phi}) = S_{X+\eta}(e^{2\pi i \phi}) |H_p(e^{2\pi i \phi})|^2,
\]

and

\[
S_{XZ_p}(e^{2\pi i \phi}) = S_X(e^{2\pi i \phi}) H_p^*(e^{2\pi i \phi}).
\]

The indirect distortion-rate function of the process \( X[\cdot] \) given \( Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot]) \), is

\[
R(P,M,\theta) = \frac{1}{2} \sum_{p=1}^P \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \lambda_p \left( J_M(e^{2\pi i \phi}) \right) \theta^{-1} \right] d\phi
\]

\[
D(P,M,\theta) = \text{mmse}_{X|Y}
\]

\[
= \sigma_X^2 - \sum_{p=1}^P \int_{-\frac{1}{2}}^{\frac{1}{2}} \lambda_p \left( J_M(e^{2\pi i \phi}) \right) - \theta \right] d\phi,
\]

where \( \lambda_1 \left( J_M(e^{2\pi i \phi}) \right) \leq \ldots \leq \lambda_P \left( J_M(e^{2\pi i \phi}) \right) \) are the eigenvalues of the \( P \times P \) matrix

\[
J_M(e^{2\pi i \phi}) = S_Y^{-\frac{1}{2}} \cdot e^{2\pi i \phi} \cdot K_M(e^{2\pi i \phi}) S_Y^{-\frac{1}{2}}(e^{2\pi i \phi}).
\]  

(58)

\( S_Y(e^{2\pi i \phi}) \) is the PSD matrix of the process \( Y[\cdot] \) and is given by

\[
(S_Y(e^{2\pi i \phi}))_{i,j} \triangleq \frac{1}{MP} \sum_{r=0}^{MP-1} S_{XZ,r} \left( e^{2\pi i \frac{r \phi}{MP}} \right)
\]

\[
= \frac{1}{MP} \sum_{r=0}^{MP-1} \left( S_X + \eta H_r H_j^\dagger \right) \left( e^{2\pi i \frac{r \phi}{MP}} \right),
\]

and \( S_Y^{\frac{1}{2}}(e^{2\pi i \phi}) \) is such that \( S_Y(e^{2\pi i \phi}) = S_Y^{\frac{1}{2}}(e^{2\pi i \phi}) S_Y^{\frac{1}{2}}(e^{2\pi i \phi}) \). The \((i,j)\)'th entry of the \( P \times P \) matrix \( K_M(e^{2\pi i \phi}) \) is given by

\[
(K_M)_{i,j}(e^{2\pi i \phi}) \triangleq \frac{1}{(MP)^2} \sum_{r=0}^{MP-1} \left( S_X H_r H_j^\dagger \right) \left( e^{2\pi i \frac{r \phi}{MP}} \right).
\]

Remark: The case where the matrix \( S_Y(e^{2\pi i \phi}) \) is not invertible for some \( \phi \in (-\frac{1}{2}, \frac{1}{2}) \) corresponds to linear dependency between the spectral components of the vector \( Y[\cdot] \). In this case, we can apply the theorem to the process \( Y'[\cdot] \) which is obtained from \( Y[\cdot] \) by removing linearly dependent components.

Proof: The proof is a multi-dimensional extension of the proof of Theorem\,[10] The details are in Appendix \[XIII\].
B. Main Result: Multi-Branch Sampling

Theorem 14 (filter-bank sampling). For each \( p = 1, \ldots, P \), let \( Z_p(\cdot) \) be the process obtained by passing a Gaussian stationary source \( X(\cdot) \) corrupted by a Gaussian stationary noise \( \eta(\cdot) \) through an LTI system \( H_p \). Let \( Y_p[\cdot] \), be the samples of the process \( Z_p(\cdot) \) at frequency \( f_s/P \), namely

\[
Y_p[n] = Z_p(Pn/f_s) = q_p \ast (X + \eta)(Pn/f_s), \quad p = 1, \ldots, P.
\]

The indirect distortion-rate function of \( X(\cdot) \) given \( Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot]) \), is given by

\[
R(P, f_s, \theta) = \frac{1}{2} \sum_{p=1}^{P} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ [\lambda_p(J(f)) - \theta] df
\]

(59a)

\[
D(P, f_s, \theta) = \text{mmse}_{X|Y}(f_s, P) + \sum_{p=1}^{P} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min \{ \lambda_p(J(f)) : \theta \} df,
\]

(59b)

where \( \lambda_1(J(f)) \leq \cdots \leq \lambda_P(J(f)) \) are the eigenvalues of the \( P \times P \) matrix

\[
J(f) = \tilde{S}_Y^{-\frac{1}{2}}(f)K(f)\tilde{S}_Y^{-\frac{1}{2}}(f),
\]

and the \( (i, j) \)th entry of the matrices \( \tilde{S}_Y(f), K(f) \in \mathbb{C}^{P \times P} \) are given by

\[
(\tilde{S}_Y)_{i,j}(f) = \sum_{k \in \mathbb{Z}} \{ S_{X+\eta H_i H_j^\ast} \}(f - f_s k),
\]

and

\[
K_{i,j}(f) = \sum_{k \in \mathbb{Z}} \{ S_{X+\eta H_i H_j^\ast} \}(f - f_s k).
\]

Proof: A full proof can be found in Appendix [V]. The idea is similar to the proof of Theorem [II] approximate the continuous time processes \( X(\cdot) \) and \( Z(\cdot) \) by discrete time processes, and then take the limit in the discrete counterpart of the problem given by Theorem [I]3

C. Optimal pre-Sampling Filter-Bank

A similar analysis as in the case of single branch sampling will show that for a fixed \( R \), the distortion is a non-increasing function of the eigenvalues of \( J(f) \). This implies that the optimal pre-sampling filters \( H_1^\ast(f), \ldots, H_P^\ast(f) \) that minimize the distortion for a given \( R \) and \( f_s \) are the same filters that minimize the MMSE in the estimation of \( X(\cdot) \) from the samples \( Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot]) \) at sampling frequency \( f_s \), given in Theorem [V]. Therefore, the following theorem applies:

Theorem 15. Given \( f_s > 0 \), the optimal pre-sampling filters \( H_1^\ast(f), \ldots, H_P^\ast(f) \) that minimize \( D(P, f_s, R) \), for all \( R \geq 0 \), are given by

\[
H_p^\ast(f) = \begin{cases} 
1 & f \in F_p^\ast, \quad p = 1, \ldots, P, \\
0 & f \notin F_p^\ast,
\end{cases}
\]

(60)
where $F_1^\star, \ldots, F_p^\star$ satisfy conditions (i) and (ii) in Theorem 5. The minimal distortion-rate function obtained this way is given by

\[
R^\star(P, f_s, \theta) = \frac{1}{2} \sum_{p=1}^{P} \int_{F_p^\star} \log^+ \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} - \theta \right] df \tag{61a}
\]

\[
D^\star(P, f_s, \theta) = \text{mmse}_{X|Y}(f_s) + \sum_{p=1}^{P} \int_{F_p^\star} \min \left\{ \frac{S_X^2(f)}{S_{X+\eta}(f)}, \theta \right\} df,
\]

\[
= \sigma_X^2 - \sum_{p=1}^{P} \int_{F_p^\star} \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} - \theta \right]^+ df. \tag{61b}
\]

**Proof:** The filters $H_1^\star(f), \ldots, H_P^\star(f)$ given by Theorem 5 maximize the eigenvalues of the matrix $J(f)$ for every $f \in (- \frac{1}{2}, \frac{1}{2})$. Since $D(P, f_s, R)$ is monotone non-increasing in these eigenvalues, $H_1^\star(f), \ldots, H_P^\star(f)$ also minimize $D(P, f_s, R)$. For this choice of $H_1(f), \ldots, H_P(f)$, (59) reduces to (61).

\[\blacksquare\]

**D. Optimal Sampling**

We have seen in Theorem 5 that minimizing the MMSE in sub-Nyquist sampling at frequency $f_s$ is equivalent to choosing a set of frequencies $\mathcal{F}^\star$ with $\mu(\mathcal{F}^\star) \leq f_s$ such that

\[
\int_{\mathcal{F}^\star} \frac{S_X^2(f)}{S_{X+\eta}(f)} df = \sup_{\mu(F) \leq f_s} \int_{F} \frac{S_X^2(f)}{S_{X+\eta}(f)} df. \tag{62}
\]

As in the case of Subsection III-E we see that for a given $R$ and $f_s$, by multi-branch uniform sampling we cannot achieve distortion lower than

\[
D_l(f_s, R(\theta)) \triangleq \sigma_X^2 - \int_{\mathcal{F}^\star} \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} - \theta \right]^+ df,
\]

where $\theta$ is determined by

\[
R = \int_{\mathcal{F}^\star} \log^+ \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} \theta^{-1} \right] df. \tag{64}
\]

This is because Proposition 5 asserts that in a parametric reverse water-filling representation of the form (61), an increment in

\[
\int_{\mathcal{F}^\star \cup \mathcal{F}_s} \frac{S_X^2(f)}{S_{X+\eta}(f)} df
\]

reduces distortion. But for any $P$, $\mu(\bigcup_{p=1}^{P} F_p^\star) \leq f_s$ so we conclude that $D_l(f_s, R) \leq D^\star(P, f_s, R)$. Moreover, $D_l(f_s, R)$ can be achieved as we extend Theorem 5 as follows:

**Theorem 16.** For any $f_s > 0$ and $\varepsilon > 0$, there exists $P \in \mathbb{N}$ and a set of LTI filters $H_1^\star(f), \ldots, H_P^\star(f)$ such that using $P$ uniform sampling branches we have

\[
D^\star(P, f_s, R) - \varepsilon < \sigma_X^2 - \int_{\mathcal{F}^\star} \left[ \frac{S_X^2(f)}{S_{X+\eta}(f)} - \theta \right]^+ df. \tag{65a}
\]
where $\theta$ is determined by

$$ R = \int_{\mathcal{F}} \log^+ \left[ \frac{S_X(f)}{S_{X+\eta}(f)} \theta^{-1} \right] df, $$

(65b)

and $\mathcal{F}^*$ is defined by (62).

**Proof:** In Theorem $6$ we found a set of pre-sampling filters $H_1^*(f), \ldots, H_P^*(f)$ such that

$$ \text{mmse}^*_{X|Y}(f_s) - \varepsilon < \sigma^2_X - \sum_{p=1}^P H_p^*(f) S_X^2(f) S_X + \eta(f) d f. $$

Since

$$ \text{mmse}^*_{X|Y}(f_s) = \sigma^2_X - \sum_{p=1}^P \int_{\mathcal{F}_p^*} \frac{S_X^2(f)}{S_{X+\eta}(f)} df, $$

where for $p = 1, \ldots, P, H_p^*(f) = 1_{F_p^*}(f)$, we conclude that

$$ \int_{\bigcup_{p=1}^P F_p^*} \frac{S_X^2(f)}{S_{X+\eta}(f)} df + \varepsilon > \int_{\mathcal{F}} \frac{S_X^2(f)}{S_{X+\eta}(f)} df. $$

By Proposition $5$ maximizing $\sum_{p=1}^P \frac{S_X^2(f)}{S_{X+\eta}(f)}$ minimizes the distortion, so the distortion $D^*(P, f_s, R)$ obtained by using $H_1^*(f), \ldots, H_P^*(f)$ is arbitrary close to $D_l(f_s, R)$. $\blacksquare$

A corollary of Theorem $16$ and Theorem $15$ is

$$ \lim_{P \to \infty} D^*(P, f_s, R) = D_l(f_s, R). $$

$D_l(f_s, R)$ is plotted in Fig. $17$ as a function of $f_s$ for two values of $R$. The equivalent rate-distortion function $R_l(f_s, D)$ at average sampling frequency $f_s$ as the number of sampling branches goes to infinity is given in Fig. $16$.

**E. Discussion**

The function $D_l(f_s, R)$ is monotone in $f_s$ by its definition (63), which is in contrast to $D(P, f_s, R)$ and $D^*(P, f_s, R)$ that are not guaranteed to be monotone in $f_s$ as the example in Fig. $17$ shows.

Fig. $17$ also suggests that multi-branch sampling can significantly reduce distortion for a given sampling frequency $f_s$ and source coding rate $R$ over single-branch sampling. Theorem $16$ shows it can achieve the bound $D_l(f_s, R)$ with the price of increasing the number of sampling branches. Since having fewer branches is more appealing from a practical point of view, it is sometimes desired to use alternative sampling techniques yielding the same performance as uniform multi-branch sampling with less sampling branches. For example, it was noted in $[17]$ that a system with a large number of uniform sampling branches can be replaced by a system with fewer a number of branches with a different sampling frequency at each branch, or by a single branch sampler with modulation.

Fig. $17$ also raises the possibility of reducing the sampling frequency without significantly affecting performance, as the function $D^*(P, f_s, R)$ for $P > 1$ approximately achieves the asymptotic value of $D^*(f_{Nyquist}, R)$ at $f_s \approx f_{Nyquist}/3$. 
We have considered a combined sampling and source coding problem, and derived an expression for the indirect distortion-rate function of an analog stationary Gaussian process corrupted by noise, given the uniform samples of this process obtained by single branch sampling and multi-branch sampling. By doing so we have generalized and unified the Shannon-Nyquist-Whittaker sampling theorem and Shannon’s rate-distortion theory for the important case of Gaussian stationary processes. An optimal design of the sampling structure that minimizes the distortion for a given sampling frequency $f_s$ and any source coding rate $R$ is shown to be the same as the sampling structure that minimizes the MMSE of signal reconstruction under sub-Nyquist sampling. This optimal sampling structure extracts the frequency components with the highest value of the source power spectral density. The function $D^*(f_s, R)$ associated with the optimal sampling structure is expressed only in terms of the spectral density of the source and the noise. It is therefore describes a fundamental trade-off in information theory and signal processing associated with any Gaussian stationary source.

Since the optimal design of the sampling structure that leads to $D^*(f_s, R)$ is tailored for a specific source statistic, it would be interesting to obtain a more universal sampling system which gives optimal performance in the case where the source statistic is unknown and taken from a family of possible distributions. For example, one may consider a ‘minmax’ distortion approach which can be seen as the source coding dual of the channel coding problem considered in [26].

The functions $D(f_s, R)$ and $D^*(f_s, R)$ fully describe the amount of information lost in uniform sampling of an analog stationary Gaussian process, in the sense that any sampling and quantization scheme with the same constraining parameters must result in a worse distortion in reconstruction. A comparison between the distortion
obtained by existing analog to digital conversion (ADC) techniques and the information theoretic bound $D^*(f_s, R)$ can motivate the search for new ADC schemes or establish the optimality of existing ones. For example, Sigma-Delta encoding [27] yields a representation of the analog source by a stream of bits at rate $R = f_s$ bits per time unit, and the distortion in the recovery of a given stationary Gaussian process under this representation can be compared to $D^*(f_s, f_s)$.

More generally, the combined source coding and sampling problem considered in this work can be seen as a source coding problem in which a constraint on the code to be a function of samples of the analog source at frequency $f_s$ is imposed. In practical ADC implementation other restrictions such as limited memory at the encoder and causality may apply. In order to understand the information theoretic bounds on such systems, it would be beneficial to extend our model to incorporate such restrictions. In particular, it is interesting to understand which restrictions lead to a non-trivial trade-off between the average number of bits per second used to represent the process and the sampling frequency of the system.

APPENDIX

VIII. PROOF OF THEOREM 3: OPTIMAL PRE-SAMPLING FILTER IN SINGLE BRANCH SAMPLING

Since $J(f) \geq 0$, we can maximize the integral over $J(f)$ by maximizing it for every $f$ in $\left( -\frac{f_s}{2}, \frac{f_s}{2} \right)$. For a given $f$, denote $h_k = |H(f - f_s k)|^2$, $x_k = S_X^2 (f - f_s k)$ and $y_k = S_{X+\eta} (f - f_s k) = S_X (f - f_s k) + S_\eta (f - f_s k)$. We arrive at the following optimization problem

\[
\begin{aligned}
\text{maximize} & \quad \frac{\sum_{k \in \mathbb{Z}} x_k h_k}{\sum_{k \in \mathbb{Z}} y_k h_k} \\
\text{subject to} & \quad h_k \geq 0, \quad k \in \mathbb{Z}
\end{aligned}
\]

Because the objective function is homogeneous in $h = (..., h_{-1}, h_0, h_1, ...)$, the last problem is equivalent to

\[
\begin{aligned}
\text{maximize} & \quad \sum_{k \in \mathbb{Z}} x_k h_k \\
\text{subject to} & \quad h_k \geq 0, \quad k \in \mathbb{Z}, \\
& \quad \sum_{k \in \mathbb{Z}} y_k h_k = 1.
\end{aligned}
\]

The optimal value of this problem is $\max_k \frac{x_k}{y_k}$, i.e. the maximal ratio over all pairs $x_k$ and $y_k$. The optimal $h$ is the indicator for the optimal ratio:

\[
h^*_k = \begin{cases} 
1 & \text{ if } k \in \arg\max_k \frac{x_k}{y_k}, \\
0 & \text{ otherwise.}
\end{cases}
\]

If there is more than one $k$ that maximizes $\frac{x_k}{y_k}$, then we can arbitrarily decide on one of them.

Going back to our standard notations, we see that for almost every $f \in \left( -\frac{f_s}{2}, \frac{f_s}{2} \right)$, the optimal $J(f)$ is given by

\[
J^*(f) = \max_{k \in \mathbb{Z}} \frac{S_X^2 (f - f_s k)}{S_{X+\eta} (f - f_s k)},
\]

and the optimal $H(f)$ is such that $|H(f - f_s k)|^2$ is non-zero for the particular $k$ that achieves this maximum. This also implies that $F^*$, the support of $H^*(f)$, satisfies properties (i) and (ii) in Definition 1.
IX. Proof of Theorem 4: MMSE in Sub-Nyquist Multi-Branch Sampling

The result is obtained by evaluating (9):

\[ \text{mmse}_{\Delta X}(f) = \sigma_X^2 - \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_0^1 S_{\Delta X}(e^{2i\pi \phi}) \, d\Delta \, \text{d}\phi. \] (66)

Since

\[ C_{\Delta XY}[k] = \mathbb{E}[X[n+k]Y^*[n]] \]

\[ = (C_{\Delta XY}[k], \ldots, C_{\Delta XY}[k]), \]

we get

\[ S_{\Delta XY}(e^{2i\pi \phi}) = (S_{\Delta XY}(e^{2i\pi \phi}), \ldots, S_{\Delta XY}(e^{2i\pi \phi})). \]

Using \( X[n] = X \left( \frac{n+\Delta}{f_s} \right) \), for each \( p = 1, \ldots, P \), we have

\[ S_{\Delta XY}[p](e^{2i\pi \phi}) = \sum_{k \in \mathbb{Z}} \mathbb{E} \left[ X \left( \frac{n+l+\Delta}{f_s} \right) Z_{p} \left( \frac{n}{f_s} \right) e^{-2\pi i \phi} \right] \]

\[ = \sum_{k \in \mathbb{Z}} S_X(f_s(\phi - k)) H_p^*(f_s(\phi - k)) e^{2\pi i k \Delta}. \]

In addition, the \((p, r)\)th entry of the \( P \times P \) matrix \( S_{Y}(e^{2i\pi \phi}) \) is given by

\[ \{ S_Y(e^{2i\pi \phi}) \}_{p,r} = \sum_{k \in \mathbb{Z}} \{ S_X + H_r H_p \} (f_s(\phi - k)), \]

where we have used the shortened notation

\[ \{ S_1 S_2 \} (x) \triangleq S_1(x) S_2(x) \]

for two functions \( S_1 \) and \( S_2 \) with the same domain. It follows that

\[ S_{\Delta XY}(e^{2i\pi \phi}) = \{ S_{\Delta XY} S_Y^{-1} S_{\Delta XY} \} (e^{2i\pi \phi}) \]

can also be written as

\[ S_{\Delta XY}(e^{2i\pi \phi}) = \text{Tr} \left\{ S_Y^{-\frac{1}{2}} S_{\Delta XY} S_Y^{-\frac{1}{2}} \right\} (e^{2i\pi \phi}), \] (67)

where \( S_Y^{-\frac{1}{2}} (e^{2i\pi \phi}) \) is the \( P \times P \) matrix satisfying \( S_Y^{-\frac{1}{2}} (e^{2i\pi \phi}) S_Y^{-\frac{1}{2}} (e^{2i\pi \phi}) = S_Y^{-1} (e^{2i\pi \phi}) \). The \((p, r)\)th entry of \( S_{\Delta XY}(e^{2i\pi \phi}) S_{\Delta XY}(e^{2i\pi \phi}) \) is given by

\[ \{ S_{X} S_{X} \}_{p,r} (e^{2i\pi \phi}) = \sum_{k \in \mathbb{Z}} \{ S_X H_p \} (f_s(\phi - k)) e^{2\pi i k \Delta} \]

\[ \times \sum_{l \in \mathbb{Z}} \{ S_X H_r \} (f_s(\phi - l)) e^{-2\pi i l \Delta} \]

\[ = \sum_{k, l \in \mathbb{Z}} \left[ \{ S_X H_p \} (f_s(\phi - k)) \{ S_X H_r \} (f_s(\phi - l)) e^{2\pi i (k-l)} \right], \]

which leads to

\[ \int_0^1 \{ S_{X} S_{X} \}_{p,r} (e^{2i\pi \phi}) \, d\Delta = \sum_{k \in \mathbb{Z}} \{ S_X^2 H_p^* H_r \} (f_s(\phi - k)). \]
From this we conclude that integrating \( \{67\} \) with respect to \( \Delta \) from 0 to 1 results in
\[
\text{Tr} \left\{ S_Y^{-1/2} K S_Y^{-1/2} \right\} (e^{2\pi i\theta}),
\]
where \( \tilde{K}(e^{2\pi i\theta}) \) is the \( P \times P \) matrix given by
\[
\tilde{K}_{p,r}(e^{2\pi i\theta}) = \sum_{k \in \mathbb{Z}} \{ S_{X} \mathcal{H}_{p}^{*} \mathcal{H}_{r} \} (f_\theta - k).
\]
The proof is completed by changing the integration variable in \( \{66\} \) from \( \theta \) to \( f = \theta f_s \), so \( S_Y(e^{2\pi i\theta}) \) and \( \tilde{K}(e^{2\pi i\theta}) \) are replaced by \( \hat{S}_{Y}(f) \) and \( \tilde{K}(f) \), respectively.

X. PROOF OF THEOREM 5: OPTIMAL FILTER-BANK IN MULTI-BRANCH SAMPLING

Let \( \mathbf{H}(f) \in \mathbb{C}^{2 \times P} \) be the matrix with \( P \) columns of infinite length defined by
\[
\mathbf{H}(f) = \begin{pmatrix}
\vdots & \vdots & \cdots & \vdots \\
H_1(f - 2f_s) & H_2(f - 2f_s) & \cdots & H_p(f - 2f_s) \\
H_1(f - f_s) & H_2(f - f_s) & \cdots & H_p(f - f_s) \\
H_1(f) & H_2(f) & \cdots & H_p(f) \\
H_1(f + f_s) & H_2(f + f_s) & \cdots & H_p(f + f_s) \\
H_1(f + 2f_s) & H_2(f + 2f_s) & \cdots & H_p(f + 2f_s) \\
\vdots & \vdots & \cdots & \vdots
\end{pmatrix}
\]
In addition, denote by \( \mathbf{S}(f) \in \mathbb{R}^{2 \times \mathbb{Z}} \) and \( \mathbf{S}_n(f) \in \mathbb{R}^{2 \times \mathbb{Z}} \) the infinite diagonal matrices with diagonal elements \( \{ S_X(f - f_k), k \in \mathbb{Z} \} \) and \( \{ S_X(f - f_k), k \in \mathbb{Z} \} \), respectively. With this notation we can write
\[
\mathbf{J}(f) = (\mathbf{H}^* \mathbf{S}_n \mathbf{H})^{-1/2} \mathbf{H}^* \mathbf{S}^2 \mathbf{H}(\mathbf{H}^* \mathbf{S}_n \mathbf{H})^{-1/2},
\]
where we suppressed the dependency on \( f \). Denote by \( \mathbf{H}^*(f) \) the matrix \( \mathbf{H}(f) \) that corresponds to the filters \( H^*(f), \ldots, H^*(f) \) that satisfy conditions (i) and (ii) in Theorem 5. By part (iii) of the remark at the end of Theorem 5, the structure of \( \mathbf{H}^*(f) \) can be described as follows: each column has a single non-zero entry, such that the first column indicates the largest among \( \frac{S_X(f - f_k)}{S_X(f - f_k)}, k \in \mathbb{Z} \), which is the diagonal of \( \mathbf{S}(f) \mathbf{S}_n^{-1}(f) \mathbf{S}(f) \). The second column corresponds to the second largest entry of \( \frac{S_X(f - f_k)}{S_X(f - f_k)}, k \in \mathbb{Z} \), and so on for all \( P \) columns of \( \mathbf{H}^*(f) \). This means that \( \mathbf{J}^*(f) \) is a \( P \times P \) diagonal matrix whose non-zero entries are the \( P \) largest values among \( \frac{S_X(f - f_k)}{S_X(f - f_k)}, k \in \mathbb{Z} \), i.e., \( \lambda_p(\mathbf{J}^*(f)) = J^*_p(f) \), for all \( p = 1, \ldots, P \).

It is left to establish the optimality of this choice of pre-sampling filters. Since the rank of \( \mathbf{J}(f) \) is at most \( P \), in order to complete the proof it is enough to show that for any \( \mathbf{H}(f) \), the \( P \) eigenvalues of the corresponding \( \mathbf{J}(f) \) are smaller than the \( P \) largest eigenvalues of \( \mathbf{S}(f) \mathbf{S}_n^{-1}(f) \mathbf{S}(f) \) compared by their respective order. Since the matrix entries of the diagonal matrices \( \mathbf{S}(f) \) and \( \mathbf{S}_n(f) \) are positive, the eigenvalues of \( \mathbf{J}(f) \) are identical to the \( P \) non-zero eigenvalues of the matrix
\[
\mathbf{S} \mathbf{H}(\mathbf{H}^* \mathbf{S}_n \mathbf{H})^{-1} \mathbf{H}^* \mathbf{S}.
\]
It is enough to prove that the matrix

\[ S_{n}^{-1} S - S (H^* S_n H)^{-1} H^* S, \]

is positive. This is equivalent to

\[ a^* S_{n}^{-1} S a - a^* S (H^* S_n H)^{-1} H^* S a \geq 0, \]  

for any sequence \( a \in \ell_2 (\mathbb{C}) \). By factoring out \( S_{n}^{-1} \) from both sides, (68) reduces to

\[ a^* a - a^* S_n^{\frac{1}{2}} H (H^* S_n H)^{-1} H^* S_n^{\frac{1}{2}} a \geq 0. \]  

(69)

The Cauchy-Schwartz inequality implies

\[ \left( a^* S_n^{\frac{1}{2}} H (H^* S_n H)^{-1} H^* S_n^{\frac{1}{2}} a \right)^2 \leq a^* a \times a^* S_n^{\frac{1}{2}} H (H^* S_n H)^{-1} H^* S_n^{\frac{1}{2}} a \]

\[ = a^* a \left( a^* S_n^{\frac{1}{2}} H (H^* S_n H)^{-1} H^* S_n^{\frac{1}{2}} a \right). \]

(70)

Dividing (70) by \( a^* S_n^{\frac{1}{2}} H (H^* S_n H)^{-1} H^* S_n^{\frac{1}{2}} a \) leads to (69).

XI. PROOF OF THEOREM 10: DISTORTION-RATE FUNCTION OF A DECIMATED PROCESS

Note that \( X[\cdot] \) and \( Y[\cdot] \) are in general not jointly stationary for \( M > 1 \), and we cannot use the discrete-time version of Theorem 7 in [42] as is. Instead we proceed as follows: For a given \( M \in \mathbb{N} \), define the vector-valued process \( X^M[\cdot] \) by

\[ X^M[n] = \langle X[Mn], X[Mn + 1], \ldots, X[Mn + M - 1] \rangle, \]

and denote by \( X^M_m[\cdot] \) its \( m^{th} \) coordinate, \( m = 0, \ldots, M - 1 \). For each \( r, m = 0, \ldots, M - 1 \) and \( n, k \in \mathbb{Z} \), the covariance between \( X^M_m[n] \) and \( X^M_k[k] \) is given by

\[ C_{X^M_m X^M_k}[k] = \mathbb{E} \left[ X^M_m[n + k] X^M_r[n]^* \right] = C_X[Mk + m - r]. \]

This shows that \( Y[\cdot] = h[\cdot] \ast X^M_0[\cdot] \) is jointly stationary with the processes \( X^M[\cdot] \). By properties of multi-rate signal processing (see for example [28]),

\[ S_Y(e^{2\pi i \phi}) = \frac{1}{M} \sum_{m=0}^{M-1} S_X \left( e^{2\pi i \frac{\phi - m}{M}} \right), \]

\[ S_{X^M X^M}(e^{2\pi i \phi}) = S_{X^M_0 X^M} \left( e^{2\pi i \phi} \right) \]

\[ = \sum_{k \in \mathbb{Z}} C_X[Mk + m - r] e^{-2\pi i \phi} \]

\[ = \frac{1}{M} \sum_{m=0}^{M-1} e^{2\pi i \frac{\phi - m}{M}} S_X \left( e^{2\pi i \frac{\phi - m}{M}} \right), \]

In the sense that it defines a positive linear operator on the Hilbert space \( \ell_2 (\mathbb{C}) \). The linear algebra notation we use here is consistent with the theory of positive operators on a Hilbert space.
and

\[ S_{X'Y'} (e^{2\pi i \phi}) = \sum_{k \in \mathbb{Z}} C_{X'Y'} [k] e^{-2\pi i k \phi} \]
\[ = \sum_{k \in \mathbb{Z}} C_{XZ} [Mk + r] e^{-2\pi i k \phi} \]
\[ = \frac{1}{M} \sum_{m=0}^{M-1} e^{2\pi i r \frac{m - \pi}{M}} S_{XZ} \left( e^{2\pi i \frac{m - \pi}{M}} \right), \]

from which we can form the \( M \times 1 \) matrix

\[ S_{X'Y'} (e^{2\pi i \phi}) = \begin{pmatrix} S_{X'Y'} (e^{2\pi i \phi}) \\ \vdots \\ S_{X'M-1} (e^{2\pi i \phi}) \end{pmatrix}. \]

The spectral density of the MMSE estimator of \( X \) from \( Y \), which is given by the trace:

\[ J_M (e^{2\pi i \phi}) = \text{Tr} S_{X'Y'} S_Y^{-1} S_{X'Y'} (e^{2\pi i \phi}) \]
\[ = \frac{1}{S_Y (e^{2\pi i \phi})} \sum_{r=0}^{M-1} \left| S_{X'Y'} (e^{2\pi i \phi}) \right|^2 \]
\[ = \frac{1}{M} \sum_{r=0}^{M-1} \sum_{m=0}^{M-1} S_{XZ} \left( e^{2\pi i \frac{m - \pi}{M}} \right)^2 \]
\[ = \frac{1}{\sum_{m=0}^{M-1} S_{XZ} \left( e^{2\pi i \frac{m - \pi}{M}} \right)} \sum_{m=0}^{M-1} S_{XZ} \left( e^{2\pi i \frac{m - \pi}{M}} \right)^2. \]

By Theorem 8, the iDRF of \( X' \) given \( Y \) is

\[ R(\theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ J_M (e^{2\pi i \phi}) \theta^{-1} \right] d\phi, \quad (71a) \]

\[ D_{X'|Y} (\theta) = \text{mmse}_{X'|Y} + \frac{1}{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ J_M (e^{2\pi i \phi}), \theta \right\} d\phi. \quad (71b) \]

Note that

\[ \text{mmse}_{X'|Y} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E} \left( X [n] - \mathbb{E} [X [n] | Y [:]] \right)^2 \]
\[ = \frac{1}{M} \sum_{m=0}^{M} \mathbb{E} \left( X_m [n] - \mathbb{E} [X_m [n] | Y [:]] \right)^2 \]
\[ = \frac{1}{M} \sum_{m=0}^{M-1} \text{mmse}_{X_m|Y} \]
\[ = \text{mmse}_{X'|Y}. \]

Since \( X' \) is a stacked version of \( X \), both processes share the same indirect rate-distortion function given \( Y \). Thus, the result is obtained substituting \( \text{mmse}_{X'|Y} \) and \( J_M (e^{2\pi i \phi}) \) in (71).


XII. PROOF OF THEOREM II

Main Result I: Single Branch Sampler

For each \( M = 1, 2, \ldots \) define \( X^M [\cdot] \) and \( Z^M [\cdot] \) to be the processes obtained by uniformly sampling \( X (\cdot) \) and \( Z (\cdot) \) at frequency \( f_s M \), i.e. \( X^M [n] = X \left( \frac{n}{f_s M} \right) \) and \( Z^M [n] = Z \left( \frac{n}{f_s M} \right) \). The spectral density of \( X^M [\cdot] \) is

\[
S_{X^M} (e^{2\pi i \phi}) = M f_s \sum_{k \in \mathbb{Z}} S_X (M f_s (\phi - k)).
\]

Using similar considerations as in the proof of Theorem 10, we see that \( X^M [\cdot] \) and \( Z^M [\cdot] \) are jointly stationary processes with cross correlation function

\[
C_{X^M Z^M} [k] = C_{XZ} \left( \frac{k}{M f_s} \right),
\]

and cross spectral density

\[
S_{X^M Z^M} (e^{2\pi i \phi}) = M f_s \sum_{k \in \mathbb{Z}} S_{XZ} (M f_s (\phi - k)).
\]

Note that \( Y [\cdot] \) is a factor-\( M \) down-sampled version of \( Z^M [\cdot] \), and the indirect rate-distortion function of \( X^M [\cdot] \) given \( Y [\cdot] \) is obtained by Theorem 10 as follows:

\[
\tilde{R}_{X^M | Y} (\theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left[ \frac{1}{M} \int J_M (e^{2\pi i \phi}) \theta^{-1} \right] d\phi,
\]

\[
D_{X^M | Y} (\theta) = \text{mmse}_{X^M | Y} (M) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ J_M (e^{2\pi i \phi}), \theta \} d\phi
\]

\[
= \sigma_{X^M}^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ J_M (e^{2\pi i \phi}) - \theta \right]^+ d\phi.
\]

Since the sampling operation preserves the \( L_2 \) norm of the signal, we have \( \sigma_{X^M}^2 = \sigma_X^2 \). In our case \( J_M (e^{2\pi i \phi}) \) is obtained by substituting the spectral densities \( S_{X^M Z^M} (e^{2\pi i \phi}) \) and \( S_{Z^M} (e^{2\pi i \phi}) \),

\[
J_M (e^{2\pi i \phi}) = \frac{1}{M} \sum_{m=0}^{M-1} \left[ S_{X^M Z^M} \left( e^{2\pi i \frac{k-m}{M}} \right) \right]^2
\]

\[
= f_s \sum_{m=0}^{M-1} \sum_{k \in \mathbb{Z}} S_{XZ} (f_s M (\phi - m - Mk)) | \frac{f_s M (\phi - m - Mk)}{M f_s}|^2
\]

\[
\text{The ideal now is to take the limit } M \to \infty \text{ in (72) and (73). Under the assumption of Riemann integrability, the distortion between almost any sample path of } X (\cdot) \text{ and any reasonable reconstruction of it from } X^M [\cdot] \text{ will converge to zero. It follows that the distortion in reconstructing } X^M [\cdot] \text{ form } \hat{Y} [\cdot] \text{ must also converge to the distortion in reconstructing } X (\cdot) \text{ from } \hat{Y} [\cdot], \text{ and the indirect distortion-rate function of } X (\cdot) \text{ given } Y [\cdot] \text{ is obtained by this limit. Thus, all that remains is to show that}
\]

\[
\lim_{M \to \infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} J_M (e^{2\pi i \phi}) d\phi = f_s \int_{-\frac{1}{2}}^{\frac{1}{2}} J (f, \phi) d\phi = \int_{-\frac{1}{2}}^{\frac{1}{2}} J (f) df.
\]

Denote

\[
g (f) \triangleq \sum_{n \in \mathbb{Z}} |S_{XZ} (f - f_n)|^2,
\]

\[
h (f) \triangleq \sum_{n \in \mathbb{Z}} S_{Z} (f - f_n),
\]
and
\[
g_M(f) \triangleq \sum_{m=0}^{M-1} \left| \sum_{k \in \mathbb{Z}} S_{XZ}(f - f_s(m-Mk)) \right|^2
\]
\[
= \sum_{m=0}^{M-1} \sum_{k \in \mathbb{Z}} S_{XZ}(f - f_s(m-Mk)) \sum_{l \in \mathbb{Z}} S_{XZ}(f - f_s(m-Mk)).
\]

Since the denominator in (74) reduces to \(\sum_{n \in \mathbb{Z}} S_Z(f_i(\phi - n))\), (75) can be written as
\[
\lim_{M \to \infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{g_M(f_i,\phi)}{h(f_s,\phi)} d\phi = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{g(f_i,\phi)}{h(f_s,\phi)} d\phi.
\]
Since the function \(h(f)\) is periodic with period \(f_s\), we can write the RHS of (76) as
\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{g_M(f_i,\phi)}{h(f_s,\phi)} d\phi = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=0}^{M-1} \left\{ \sum_{k \in \mathbb{Z}} S_{XZ}(f_s(\phi - m+Mk)) \right. \\
\times \left. \sum_{l \in \mathbb{Z}} S_{XZ}(f_s(\phi - m+Mk)) \right\} d\phi.
\]
Denoting
\[
f_1(\phi) = \frac{S_{XZ}(f_s(\phi))}{\sqrt{h(f_s,\phi)}},
\]
and \(f_2(\phi) = f_1(\phi)\), (76) follows from the following lemma:

**Lemma 1.** Let \(f_1(\phi)\) and \(f_2(\phi)\) be two complex valued bounded functions such that \(\int_{-\infty}^{\infty} |f_i(\phi)|^2 d\phi < \infty, i = 1, 2\). Then for any \(f_s > 0\),
\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=0}^{M-1} \sum_{k \in \mathbb{Z}} f_1(\phi + m + kM) \sum_{l \in \mathbb{Z}} f_2(\phi + m + lM) d\phi
\]
converges to
\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{n \in \mathbb{Z}} f_1(\phi - n) f_2(\phi - n) d\phi,
\]
as \(M\) goes to infinity.

Equation (78) can be written as
\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=0}^{M-1} \sum_{k \in \mathbb{Z}} f_1(\phi + m + kM) f_2(\phi + m + kM) d\phi
\]
\[
+ \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=0}^{M-1} \sum_{k \neq l} f_1(\phi + m + kM) f_2(\phi + m + lM) d\phi.
\]
Since the term (80) is identical to (79), all that is left is to show that (81) vanishes as \(M \to \infty\). Take \(M\) large enough such that
\[
\int_{|\phi| \leq \frac{M+1}{1+M}} |f_i(\phi)|^2 d\phi < \varepsilon^2, \quad i = 1, 2.
\]
W.L.G, take this \(M\) to be even. By a change of variables (81) can be written as
\[
\sum_{k \neq l} \int_{-\frac{M+1}{2}}^{\frac{M+1}{2}} f_1(\phi + \frac{M}{2} + kM) f_2(\phi + \frac{M}{2} + lM) d\phi.
\]
We split the indexes in the last sum into three disjoint sets:
1) $\mathcal{I} = \{k, l \in \mathbb{Z} \setminus \{0, -1\}, k \neq l\}$,

$$\sum_{y} \int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1 \left( \varphi + \frac{M}{2} + kM \right) f_2 \left( \varphi + \frac{M}{2} + lM \right) d\varphi \leq \sum_{y} \int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} \left| f_1 \left( \varphi + \frac{M}{2} + kM \right) \right|^2 d\varphi$$

$$+ \sum_{y} \int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} \left| f_2 \left( \varphi + \frac{M}{2} + lM \right) \right|^2 d\varphi$$

$$\leq \int_{\mathbb{R} \setminus [-\frac{M+1}{M+2}, \frac{M+1}{M+2}]} \left| f_1 (\varphi) \right|^2 d\varphi$$

$$+ \int_{\mathbb{R} \setminus [-\frac{M+1}{M+2}, \frac{M+1}{M+2}]} \left| f_2 (\varphi) \right|^2 d\varphi \leq 2\varepsilon^2,$$

(83)

where (a) is due to the triangle inequality and since for any two complex numbers $a, b$, $|ab| \leq \frac{|a|^2 + |b|^2}{2} \leq |a|^2 + |b|^2$.

2) $k = 0, l = -1$,

$$\int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1 \left( \varphi + \frac{M}{2} \right) f_2 \left( \varphi - \frac{M}{2} \right) d\varphi = \int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1 \left( \varphi + \frac{M}{2} \right) f_2 \left( \varphi - \frac{M}{2} \right) d\varphi$$

$$+ \int_{0}^{M} f_1 \left( \varphi + \frac{M}{2} \right) f_2 \left( \varphi - \frac{M}{2} \right) d\varphi$$

$$\leq \sqrt{\int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1^2 \left( \varphi + \frac{M}{2} \right) d\varphi} \sqrt{\int_{0}^{M} f_2^2 \left( \varphi - \frac{M}{2} \right) d\varphi}$$

$$+ \sqrt{\int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1^2 \left( \varphi + \frac{M}{2} \right) d\varphi} \sqrt{\int_{0}^{M} f_2^2 \left( \varphi - \frac{M}{2} \right) d\varphi}$$

$$\leq \varepsilon \|f_1\|_2 + \varepsilon \|f_2\|_2,$$

(84)

where (a) follows from the Cauchy-Schwarz inequality.

3) $k = -1, l = 0$, using the same arguments as in the previous case,

$$\int_{-\frac{M+1}{M+2}}^{\frac{M+1}{M+2}} f_1 \left( \varphi + \frac{M}{2} \right) f_2 \left( \varphi - \frac{M}{2} \right) d\varphi \leq \varepsilon (\|f_1\|_2 + \|f_2\|_2).$$

(85)

From (83), (84) and (85), the sum (82) can be bounded by

$$2\varepsilon (\|f_1\|_2 + \|f_2\|_2) + 2\varepsilon^2,$$

which can be made as close to zero as required. Since (76) follows from (77) and Lemma [1] the proof is complete.
XIII. Proof of Theorem 13

Similar to the proof of Theorem 10, the iDRF of \( X[\cdot] \) given \( Y[\cdot] \) coincides with the iDRF of the vector-valued process \( X^{PM}[\cdot] \) defined by

\[
X^{MP}[n] = (X[PMn], X[PMn+1], \ldots, X[PMn+PM-1]) .
\]

For a given \( M \in \mathbb{N} \), \( X^{MP}[\cdot] \) is a stationary Gaussian process with PSD matrix

\[
(S_X)_{r,s} (e^{2\pi i \phi}) = \frac{1}{MP} \sum_{m=0}^{MP-1} e^{2\pi i (r-s) \frac{\phi}{PM}} S_X \left( e^{2\pi i \frac{m}{MP}} \right) .
\]

The processes \( Y[\cdot] \) and \( X^{PM}[\cdot] \) are jointly Gaussian and stationary with a \( PM \times P \) cross PSD whose \((m+1,p)\)th entry is given by

\[
(S_{X^{PM}Y})_{m,p} (e^{2\pi i \phi}) = S_{X^{MP}Y} (e^{2\pi i \phi})
\]

\[
= \sum_{k \in \mathbb{Z}} \mathbb{E} [X[PMk+m]Z_p(0)] e^{-2\pi i \phi k}
\]

\[
= \frac{1}{PM} \sum_{r=0}^{PM-1} e^{2\pi im \frac{\phi}{PM}} S_{XZ_r} \left( e^{2\pi i \frac{r}{PM}} \right),
\]

where we denoted by \( X^{PM}_m \) the \( m \)th coordinate of \( X^{PM}[\cdot] \). The PSD of the MMSE estimator of \( X^{PM}[\cdot] \) from \( Y[\cdot] \) is given by

\[
S_{X^{PM}Y} (e^{2\pi i \phi}) = \left\{ S_{X^{MP}Y} S_YS_2^{-1} \right\} (e^{2\pi i \phi}),
\]

(86)

Since only the non-zero eigenvalues of \( S_{X^{PM}Y} (e^{2\pi i \phi}) \) contribute to the distortion in (43b), we are interested in the non-zero eigenvalues of (86). These are identical to the non-zero eigenvalues of

\[
\left\{ S_2^{-1/2} S_{X^{PM}Y} S_YS_2^{-1/2} \right\} (e^{2\pi i \phi}),
\]

(87)

where \( \left\{ S_2^{-1/2} S_2^{-1/2} \right\} (e^{2\pi i \phi}) = S_2^{-1} (e^{2\pi i \phi}) \). The \((p,q)\)th entry of the \( P \times P \) matrix \( \left\{ S_{X^{PM}Y} S_{X^{PM}Y} \right\} (e^{2\pi i \phi}) \) is given by

\[
\frac{1}{(PM)^2} \sum_{l=0}^{PM-1} \sum_{r=0}^{PM-1} e^{-2\pi i l \frac{\phi}{PM}} S_{XZ_r} \left( e^{2\pi i \frac{l}{PM}} \right)
\]

\[
\times \sum_{k=0}^{PM-1} e^{2\pi i k \frac{\phi}{PM}} S_{XZ_k} \left( e^{2\pi i \frac{k}{PM}} \right)
\]

\[
= \frac{1}{(PM)^2} \sum_{r=0}^{PM-1} S_{XZ_r} \left( e^{2\pi i \frac{r}{PM}} \right) S_{XZ_r} \left( e^{2\pi i \frac{r}{PM}} \right)
\]

\[
= \frac{1}{(PM)^2} \sum_{r=0}^{PM-1} \left\{ S_2^{-1} H_r H_r \right\} \left( e^{2\pi i \frac{r}{PM}} \right),
\]

which is the matrix \( K_M (e^{2\pi i \phi}) \) defined in Theorem 13. Applying Theorem 8 with the eigenvalues of (87) completes the proof.
XIV. Proof of Theorem 14: Main Result II: Multi-Branch Sampling

For $M \in \mathbb{N}$, define $X^M[\cdot]$ and $Z^M_p[\cdot]$, $p = 1, \ldots, P$ to be the processes obtained by uniformly sampling $X(\cdot)$ and $Z_p(\cdot)$ at frequency $f_pMP$, i.e. $X^M[n] = X\left(\frac{n}{f_pMP}\right)$ and $Z^M_p[n] = Z\left(\frac{n}{f_pMP}\right)$. We have

$$S_{X^M}(e^{2\pi f}) = M P f_s \sum_{k \in \mathbb{Z}} S_X(M P f_s (\phi - k)),$$

and

$$S_{Z^M_p}(e^{2\pi f}) = M P f_s \sum_{k \in \mathbb{Z}} \left\{ S_X |H_p|^2 \right\} (M P f_s (\phi - k)).$$

In addition, $X^M[\cdot]$ and $Z^M_p[\cdot]$ are jointly stationary processes with cross spectral densities

$$S_{X^M Z^M_p}(e^{2\pi f}) = M P f_s \sum_{m=0}^{MP-1} \sum_{k \in \mathbb{Z}} \left[ S_{X_m Z}\left(\frac{\phi - m}{MP} - k\right) \right],$$

for all $p, r = 1, \ldots, P$, and

$$S_{X^M Z^M_p}(e^{2\pi f}) = M P f_s \sum_{m=0}^{MP-1} \sum_{k \in \mathbb{Z}} S_{X_m Z}(f_s \left(\frac{\phi - m}{MP} - k\right)).$$

Since $Y_p[\cdot]$ is a factor $M$ down-sampled version of $Z^M_p[\cdot]$, the indirect distortion-rate function of $X^M[\cdot]$ given $Y[\cdot] = (Y_1[\cdot], \ldots, Y_P[\cdot])$ was found in Theorem 13 to be

$$\bar{R}_{X^M|Y}(P, M, \theta) = \frac{1}{2} \sum_{p=1}^{P} \int_{-1/2}^{1/2} \log^+ \left[ \lambda_p \left(J_M(e^{2\pi f})\right) \theta^{-1} \right] d\phi,$$

(88)

$$D_{X^M|Y}(P, M, \theta) = \sigma^2_{X^M} - \sum_{p=1}^{P} \int_{-1/2}^{1/2} \left[ \lambda_p \left(J_M(e^{2\pi f})\right) - \theta^+ \right] d\phi,$$

(89)

where

$$J_M(e^{2\pi f}) = S^\frac{1}{2} K_M S^\frac{1}{2} (e^{2\pi f}),$$

$S_{Y}(e^{2\pi f})$ is the spectral density matrix of the process $Y[\cdot]$ with $(p, r)^{th}$ entry

$$\left(S_{Y}(e^{2\pi f})\right)_{p, r} = \frac{1}{MP} \sum_{m=0}^{MP-1} S_{Z^M_p Z^M_r}(e^{2\pi f} e^{-m})$$

$$= f_s \sum_{m=0}^{MP-1} \sum_{k \in \mathbb{Z}} S_{Z^M_p Z^M_r} (f_s (\phi - m - MPk))$$

$$= f_s \sum_{n \in \mathbb{Z}} S_{Z^M_p Z^M_r} (f_s (\phi - n))$$

$$= f_s \sum_{n \in \mathbb{Z}} \left\{ S_{X+n H_p H_p^*} \right\} (f_s \phi - f_s n)$$

$$= (S_{Y})_{p, r} (f_s f_s f_s),$$

(90)
and

$$(K_M)_{p,r} = \frac{1}{(MP)^2} \sum_{m=0}^{MP-1} \left\{ S_{X_H}^2 H_p H_r \right\} \left( e^{2\pi i \phi/m} \right)$$

$$= \frac{1}{(MP)^2} \sum_{m=0}^{MP-1} \left\{ S_{X}^2 \phi S_{X}^2 \phi \right\} \left( e^{2\pi i \phi/m} \right)$$

$$= f_s^2 \sum_{m=0}^{MP-1} \sum_{k \in Z} S_{X_k} \left( f_s (\phi - m - kMP) \right) \times \sum_{l \in Z} S_{X_l} \left( f_s (\phi - m - lMP) \right) \right\} \right).$$

The idea now is that under the assumption of Riemann integrability, the distortion between almost any sample path of $X(\cdot)$ and any reasonable reconstruction of it from $X^M(\cdot)$ will converge to zero as $M \to \infty$. It follows that the distortion in reconstructing $X^M(\cdot)$ from $\hat{Y}(\cdot)$ must also converge to the distortion in reconstructing $X(\cdot)$ from $\hat{Y}(\cdot)$, and the indirect distortion-rate function of $X(\cdot)$ given $Y(\cdot)$ is obtained by this limit. That is, we are looking to evaluate $\lim_{M \to \infty}$ in the limit $M \to \infty$.

First note that

$$\sigma^2_{\phi} = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} X^M \left( e^{2\pi i \phi} \right) d\phi$$

$$= \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} MP \sum_{k \in Z} X_k \left( f_s (\phi - k) \right) d\phi$$

$$= \int_{-\infty}^{\infty} X(f) df = \sigma^2_{\phi}.$$

In addition, by a change of the integration variable from $f$ to $\phi = f/f_s$, we can write $\lim_{M \to \infty}$ as

$$R(P, f_s, \theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \lambda_p \left( J(e^{2\pi i \phi}) / \theta \right) \right] d\phi$$

$$D(P, f_s, \theta) = \sigma^2_{\phi} - \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \lambda_p \left( J(e^{2\pi i \phi}) / \theta \right) \right] d\phi,$$

where in $\phi \in (-\frac{1}{2}, \frac{1}{2})$, the matrix $J(e^{2\pi i \phi})$ is given by

$$J(e^{2\pi i \phi}) = S_Y^{-\frac{1}{2}} (e^{2\pi i \phi}) K(e^{2\pi i \phi}) S_Y^{-\frac{1}{2}} (e^{2\pi i \phi}),$$

and $K(e^{2\pi i \phi}) = f_s^2 K(f_s \phi)$. It follows that in order to complete the proof, it is enough to show that the eigenvalues of $J_M(e^{2\pi i \phi})$ seen as $L_1 \left( \left( -\frac{1}{2}, \frac{1}{2} \right) \right)$ functions in $\phi$ converge to the eigenvalues of $J(e^{2\pi i \phi})$. Since

$$\|S_Y^{-\frac{1}{2}} K_M S_Y^{-\frac{1}{2}} \|_2 < \|S_Y^{-\frac{1}{2}} K S_Y^{-\frac{1}{2}} \|_2 \leq \|S_Y^{-1}\|_2 \|f_s K_M - K\|_2,$$

it is enough to prove convergence in $L_1 \left( \left( -\frac{1}{2}, \frac{1}{2} \right) \right)$ for each entry, i.e. that

$$\lim_{M \to \infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ (K_M)_{p,r} (e^{2\pi i \phi}) - (K)_{p,r} (e^{2\pi i \phi}) \right\} d\phi = 0$$

(93)
for all $p, r = 1, \ldots, P$. Since

$$
\left(\hat{K}\right)_{p,r} \left(\epsilon^{2\pi \phi}\right) = f_s^2 \left(\hat{K}\right)_{p,r} \left(f_s \phi\right)
$$

$$
= f_s^2 \sum_{k \in \mathbb{Z}} \left\{S_k^2 H_i H_j^*\right\} \left(f_s (\phi - k)\right),
$$

(93) follows by applying Lemma 1 to (91) with

$$
f_1(\phi) = \frac{S_{XZ_p}(f_s \phi)}{\sqrt{\|SY(e^{2\pi i \phi})\|_2}},
$$

$$
f_2(\phi) = \frac{S_{XZ_r}(f_s \phi)}{\sqrt{\|SY(e^{2\pi i \phi})\|_2}}.
$$

ACKNOWLEDGMENT

We thank I. E. Aguerri and D. Gündüz for valuable discussions regarding the problem formulation, and S. Rini, Y. Chen, A. Manolakos, N. Soltani and M. Chowdhury for helpful remarks and discussions. This work was supported in part by the NSF Center for Science of Information (CSoI) under grant CCF-0939370, the BSF Transformative Science Grant 2010505 and the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI).

REFERENCES

[1] T. M. Cover and J. A. Thomas, Elements of information theory (2. ed.). Wiley, 2006.
[2] T. Berger, Rate-Distortion Theory. Wiley Online Library, 1971.
[3] M. Mishali and Y. C. Eldar, “Sub-nyquist sampling,” Signal Processing Magazine, IEEE, vol. 28, no. 6, pp. 98–124, 2011.
[4] C. E. Shannon, “A mathematical theory of communication,” Bell System Tech. J., vol. 27, pp. 379–423, 623–656, 1948.
[5] M. Matthews, “On the linear minimum-mean-squared-error estimation of an undersampled wide-sense stationary random process,” Signal Processing, IEEE Transactions on, vol. 48, no. 1, pp. 272–275, 2000.
[6] H. Witsenhausen, “Indirect rate distortion problems,” Information Theory, IEEE Transactions on, vol. 26, no. 5, pp. 518–521, 1980.
[7] R. Dobrushin and B. Tsybakov, “Information transmission with additional noise,” Information Theory, IRE Transactions on, vol. 8, no. 5, pp. 293–304, 1962.
[8] J. Wolf and J. Ziv, “Transmission of noisy information to a noisy receiver with minimum distortion,” Information Theory, IEEE Transactions on, vol. 16, no. 4, pp. 406–411, 1970.
[9] D. Chan and R. Donaldson, “Optimum pre- and postfiltering of sampled signals with application to pulse modulation and data compression systems,” Communication Technology, IEEE Transactions on, vol. 19, no. 2, pp. 141–157, April 1971.
[10] T. Michaeli and Y. C. Eldar, “High-rate interpolation of random signals from nonideal samples,” Signal Processing, IEEE Transactions on, vol. 57, no. 3, pp. 977–992, 2009.
[11] 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.
[12] 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
[13] Y.-P. Lin and P. Vaidyanathan, “Periodically nonuniform sampling of bandpass signals,” Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on, vol. 45, no. 3, pp. 340–351, Mar 1998.
[14] M. Unser and J. Zerubia, “A generalized sampling theory without band-limiting constraints,” Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on, vol. 45, no. 8, pp. 959–969, Aug 1998.
[15] A. Papoulis, “Generalized sampling expansion,” Circuits and Systems, IEEE Transactions on, vol. 24, no. 11, pp. 652–654, Nov 1977.
[16] Y. Chen, Y. Eldar, and A. Goldsmith, “Shannon meets nyquist: Capacity of sampled gaussian channels,” Information Theory, IEEE Transactions on, vol. 59, no. 8, pp. 4889–4914, 2013.
[17] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, “Channel Capacity under Sub-Nyquist Nonuniform Sampling,” ArXiv e-prints, Apr. 2012.
[18] R. Gallager, Information theory and reliable communication. Wiley, 1968.
[19] M. Bellanger, G. Bonnerot, and M. Coudreuse, “Digital filtering by polyphase network: application to sample-rate alteration and filter banks,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 2, pp. 109–114, Apr 1976.
[20] T. Berger and J. Gibson, “Lossy source coding,” Information Theory, IEEE Transactions on, vol. 44, no. 6, pp. 2693–2723, 1998.
[21] E. Landau, “Ueber einen Bieberbachschen Satz,” Rend. Circ. Mat. Palermo (2), vol. 46, pp. 456–462, 1922.
[22] S. Pinsker, Information and information stability of random variables and processes, ser. Holden-Day series in time series analysis. Holden-Day, 1964.
[23] M. S. Pinsker, “Computation of the message rate of a stationary random process and the capacity of a stationary channel,” Dokl. Akad. Nauk. USSR, vol. 111, no. 4, 1956.
[24] V. Kafedziski, “Rate distortion of stationary and nonstationary vector gaussian sources,” in Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on, 2005, pp. 1054–1059.
[25] A. Kipnis and A. Goldsmith, “Rate-distortion function of cyclostationary gaussian processes,” in preparation.
[26] Y. Chen, A. Goldsmith, and Y. C. Eldar, “Minimax capacity loss under sub-nyquist universal sampling,” CoRR, vol. abs/1304.7751, 2013.
[27] I. Daubechies and R. DeVore, “Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order,” Annals of Mathematics, pp. 679–710, 2003.
[28] P. Vaidyanathan, “Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial,” Proceedings of the IEEE, vol. 78, no. 1, pp. 56–93, 1990.