Some Results on the Information Loss in Dynamical Systems

Bernhard C. Geiger*, Gernot Kubin*
*Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria
{geiger,gernot.kubin}@tugraz.at

***Abstract*—In this work we investigate the information loss in (nonlinear) dynamical input-output systems and provide some general results. In particular, we present an upper bound on the information loss rate, defined as the (non-negative) difference between the entropy rates of the jointly stationary stochastic processes at the input and output of the system.**

We further introduce a family of systems with vanishing information loss rate. It is shown that not only linear filters belong to that family, but – under certain circumstances – also finite-precision implementations of the latter, which typically consist of nonlinear elements.

I. Introduction

Transmission and processing of information is the primary concern in many fields of communications, signal processing, and machine learning. The typical impairments considered in these contexts are noise and interference, incomplete data sets, and coarse observations, eliciting both information-theoretic and energy-centered analyses. In contrary, the effect of deterministic input-output systems on the information content, i.e., the entropy rate, of a signal has not yet been thoroughly analyzed. Still, nonlinear dynamical systems – capable of changing information content – are omnipresent in communication systems in the roles of high-power amplifiers or frequency mixers. Another example is the energy detector (e.g., [1]), a low-complexity receiver architecture for wireless communications, which is typically modeled as a Hammerstein system, cf. Section VI-C. To obtain a better understanding of the effects of these system components, an information-theoretic treatment is essential.

In this paper, we establish a framework for analyzing the effects of discrete-time dynamical systems with a finite-dimensional state vector on the entropy rate of a signal. While the analysis of continuous-valued stochastic processes will be left for future work, here we focus on (jointly) stationary input and output processes taking values from countable alphabets.

The data processing inequality (DPI, [2, pp. 35]) states that the entropy of a discrete random variable (RV) cannot increase by passing the RV through a static nonlinearity. It was shown that the same result holds for entropy rates of jointly stationary stochastic processes on finite alphabets, both for static nonlinearities [3] and general dynamical systems [4]. Continuous-valued processes passing through linear filters were already analyzed by Shannon in terms of differential entropy rates [5], [6], which in our opinion are not adequate measures of information loss, cf. Section V. The conditional entropy, used to characterize the information lost by passing a continuous RV through a static nonlinearity [7] or by multiplying two integers [8], appears to be more appropriate.

We start by defining the information loss rate in Section II and show that this quantity is equal to the difference between the entropy rates of the input and output processes. This choice establishes the DPI for dynamical systems in Section III, stating that the information loss rate is non-negative. This result is then complemented by an upper bound that can be evaluated easily. In Section IV we introduce a family of dynamical systems for which we show that the information loss rate vanishes. This family not only comprises a large class of stable linear filters (see Section V), but also their finite-precision counterparts, commonly used in digital signal processing. Aside from the latter, Section VI discusses some other examples illustrating our theoretical results.

An extended version of this paper is available in [9].

II. Problem Statement & Preliminaries

We consider a discrete-time regular two-sided stationary stochastic process $\mathbf{X}$ taking values from a countable set $\mathcal{X}$. Let X_n denote the RV of the n-th sample and let $X_k^n = (X_k, X_{k+1}, \dots, X_n)$, thus $\mathbf{X} = X_{-\infty}^{\infty}$. For the actual value of X_n we write x_n. We further consider another countable set $\mathcal{Y}$ which needs not be identical to $\mathcal{X}$. Let $H(X_n)$ denote the zeroth-order entropy of X_n and let $\bar{H}(\mathbf{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n)$ denote the entropy rate of $\mathbf{X}$. The restriction to countable sets ensures that entropies and entropy rates are well-defined.

The following class of dynamical systems is treated in this work:

Definition 1 (Finite-Dimensional Dynamical System)**.** Let $Y_n = f(X_{n-N}^n, Y_{n-M}^{n-1})$, $0 \leq M, N < \infty$, be the RV of the n-th output sample of a dynamical system with a finite-dimensional state vector subject to the input process $\mathbf{X}$. Here, $f\colon \mathcal{X}^{N+1} \times \mathcal{Y}^M \to \mathcal{Y}$ is a function such that the sequence of output samples, Y_n, constitutes a two-sided stochastic process $\mathbf{Y}$ jointly stationary with $\mathbf{X}$.

Definition 2 (Information Loss Rate)**.** Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly stationary processes on countable sets related as in Definition 1. The average information lost per sample is given by the conditional entropy rate

$$\bar{H}(\mathbf{X}|\mathbf{Y}) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n|Y_1^n). \tag{1}$$

Characterizing the information loss as a conditional entropy rate is quite intuitive: The conditional entropy rate denotes the average number of bits per sample unknown about the input sequence after observing the output sequence; i.e., the average information lost per sample by passing the sequence through the system in question.

Before proceeding with the analysis, we will introduce two Lemmas:

Lemma 1. *For any set of discrete RVs Z_1^n and any function $f(Z_k, Z_l, \dots)$, $1 \leq k, l, \cdots \leq n$, the following holds:*

$$H(Z_1^n, f(Z_k, Z_l, \dots)) = H(Z_1^n) \tag{2}$$

Proof: See [2, Prob. 2.4]. ■

Lemma 2. *Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly stationary stochastic processes on countable sets. Then, for $M < \infty$,*

$$\bar{H}(\mathbf{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n | Y_1^M) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n, Y_1^M).$$

Proof: See [9]. ■

III. Information Loss Rate in Dynamical Systems

In this Section, which comprises the main contribution of this work, we present some general results on the information loss rate induced by a system satisfying Definition 1. We will start by proving a Theorem which essentially states that the information loss rate is identical to the difference of entropy rates:

Theorem 1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly stationary processes on countable sets related as in Definition 1. Then, the information loss rate is given by the difference of entropy rates:*

$$\bar{H}(\mathbf{X}|\mathbf{Y}) = \bar{H}(\mathbf{X}) - \bar{H}(\mathbf{Y}) \tag{3}$$

Proof: While the proof for static functions (i.e., $M = N = 0$) is relatively simple [3], for dynamical systems we have to show that

$$\bar{H}(\mathbf{X}|\mathbf{Y}) = \lim_{n\to\infty} \frac{1}{n} \left(H(X_1^n, Y_1^n) - H(Y_1^n) \right) \tag{4}$$

$$= \lim_{n\to\infty} \frac{1}{n} H(X_1^n) - \lim_{n\to\infty} \frac{1}{n} H(Y_1^n) \tag{5}$$

i.e., that

$$\lim_{n\to\infty} \frac{1}{n} H(X_1^n, Y_1^n) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n). \tag{6}$$

Consider that, for $n > \max\{M, N\}$

$$H(X_1^n, Y_1^n) = H(Y_n, X_1^n, Y_1^{n-1}) \tag{7}$$

$$= H(f(X_{n-N}^n, Y_{n-M}^{n-1}), X_1^n, Y_1^{n-1}) \tag{8}$$

$$\overset{(a)}{=} H(X_1^n, Y_1^{n-1}) \tag{9}$$

where (a) is due to Lemma 1. By repeated application,

$$H(X_1^n, Y_1^n) = H(X_1^n, Y_1^{\max\{M,N\}}). \tag{10}$$

Since this holds for all $n > \max\{M, N\}$, it also holds in the limit and with Lemma 2 we obtain

$$\lim_{n\to\infty} \frac{1}{n} H(X_1^n, Y_1^{\max\{M,N\}}) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n) \tag{11}$$

and thus

$$\bar{H}(\mathbf{X}|\mathbf{Y}) = \bar{H}(\mathbf{X}) - \bar{H}(\mathbf{Y}). \tag{12}$$

This completes the proof. ■

The significance of this Theorem lies in the fact that the information loss can be inferred by comparing the entropy rates of the input and output processes. Note that the same does not hold for differential entropy rates, as we will argue in Section V.

By the non-negativity of the conditional entropy rate the following Corollary to Theorem 1 shows that the entropy rate of the system output cannot be larger than the entropy rate of the system input. This result, originally stated in [4] for finite alphabets, further justifies our intuitive definition of information loss:

Corollary 1 (DPI for Dynamical Systems)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly stationary processes on countable sets related as in Definition 1. Then, the entropy rate of the output process $\mathbf{Y}$ cannot be larger than the entropy rate of the input process $\mathbf{X}$, i.e.,*

$$\bar{H}(\mathbf{Y}) \leq \bar{H}(\mathbf{X}). \tag{13}$$

Generally, the computation of entropy rates is a non-trivial problem, where closed-form solutions exist only for simple processes (e.g., Markov chains). Since functions of stochastic processes rarely allow such a simplified treatment, the availability of bounds is of vital importance. We will thus present an upper bound on the information loss rate, which is simple to evaluate:

Theorem 2 (Upper Bound)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly stationary processes on countable sets related as in Definition 1. Then, the information loss rate is bounded by*

$$\bar{H}(\mathbf{X}|\mathbf{Y}) \leq \max_{(x,\theta)\in\mathcal{X}\times\mathcal{T}} \log |f_\theta^{-1}[f_\theta(x)]| \tag{14}$$

where $\mathcal{T} = \mathcal{X}^N \times \mathcal{Y}^M$, $\theta \in \mathcal{T}$ are the possible values of the RV $\Theta_n = \{X_{n-N}^{n-1}, Y_{n-M}^{n-1}\}$, and $f_\theta^{-1}[\cdot]$ denotes the preimage under f_θ, an instantiation of the function $f_{\Theta_n}(\cdot) = f(\cdot, \Theta_n)$.

Proof:

$$\bar{H}(\mathbf{X}|\mathbf{Y}) \overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_1^{i-1}, Y_1^n) \tag{15}$$

$$\overset{(b)}{\leq} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_1^{i-1}, Y_1^i) \tag{16}$$

where (a) is due to the chain rule of entropy and (b) is due to the fact that conditioning reduces entropy. The expression under the sum in (16) is a non-negative decreasing sequence in i and thus has a limit. We use the Cesáro mean [2, Thm. 4.2.3] and obtain

$$\bar{H}(\mathbf{X}|\mathbf{Y}) \leq \lim_{n\to\infty} H(X_n | X_1^{n-1}, Y_1^n) \tag{17}$$

$$\leq H(X_n | X_{n-N}^{n-1}, Y_{n-M}^n) \tag{18}$$

$$= H(X_n | Y_n, \Theta_n). \tag{19}$$

We now replace $Y_n = f(X_n, \Theta_n) = f_{\Theta_n}(X_n)$, where we treat the collection of all previous RVs influencing Y_n as a (random) parameter Θ_n of the function. This approach lets us interpret the dynamical system as a parameterized static system $f_{\Theta_n}\colon \mathcal{X} \to \mathcal{Y}$, where we let Θ_n take values θ from $\mathcal{T} = \mathcal{X}^N \times \mathcal{Y}^M$. We thus continue

$$\begin{aligned}\bar{H}(\mathbf{X}|\mathbf{Y}) &\leq H(X_n|f_{\Theta_n}(X_n), \Theta_n) && (20)\\ &= \sum_{(x,\theta)\in\mathcal{X}\times\mathcal{T}} H(X_n|f_\theta(x),\theta)\Pr(X_n = x, \Theta_n = \theta)\\ &\overset{(c)}{\leq} \sum_{(x,\theta)\in\mathcal{X}\times\mathcal{T}} \log|f_\theta^{-1}[f_\theta(x)]|\Pr(X_n = x, \Theta_n = \theta)\\ &\leq \max_{(x,\theta)\in\mathcal{X}\times\mathcal{T}} \log|f_\theta^{-1}[f_\theta(x)]| && (21)\end{aligned}$$

where (c) is due to conditioning and the maximum entropy property of the uniform distribution over an alphabet size equal to the cardinality of the preimage under f_θ. Maximizing over all possible x and parameter values θ completes the proof. ■

This result can be interpreted as relating the information loss rate of a dynamical system to the information loss rate induced by a static function. In particular, we let the static function be parameterized by previous input and output values taking effect on Y_n and upper bound the information loss rate by the maximum cardinality of the preimage under f_θ. While this upper bound may be rather conservative, it is particularly simple to evaluate if the system function from Definition 1 is available. We will illustrate the use of this result in Section VI-C.

Finally, we present a result about the cascade of systems:

Theorem 3 (Cascading Systems)**.** *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{Z}$ *be jointly stationary stochastic processes on countable sets* $\mathcal{X}$*,* $\mathcal{Y}$*, and* $\mathcal{Z}$*, respectively, where* $\mathbf{Y}$ *is generated by passing* $\mathbf{X}$ *through a system satisfying Definition 1, and* $\mathbf{Z}$ *is generated by passing* $\mathbf{Y}$ *through another such system. Then, the information loss rate induced by the cascade is*

$$\bar{H}(\mathbf{X}|\mathbf{Z}) = \bar{H}(\mathbf{X}|\mathbf{Y}) + \bar{H}(\mathbf{Y}|\mathbf{Z}). \quad (22)$$

Proof: By using Theorem 1 $\bar{H}(\mathbf{X}|\mathbf{Z})$ can be written as

$$\begin{aligned}\bar{H}(\mathbf{X}|\mathbf{Z}) &= \bar{H}(\mathbf{X}) - \bar{H}(\mathbf{Z}) && (23)\\ &= \bar{H}(\mathbf{X}) - \bar{H}(\mathbf{Y}) + \bar{H}(\mathbf{Y}) - \bar{H}(\mathbf{Z}) && (24)\\ &= \bar{H}(\mathbf{X}|\mathbf{Y}) + \bar{H}(\mathbf{Y}|\mathbf{Z}). && (25)\end{aligned}$$

■

IV. Partially Invertible Systems

We now impose an additional restriction on the system function in Definition 1. This additional restriction defines a family of systems for which the information loss rate can be shown to vanish.

Definition 3 (Partially invertible system)**.** A system satisfying Definition 1 is *partially invertible* if there exists a function $f_{\text{inv}}\colon \mathcal{X}^N \times \mathcal{Y}^{M+1} \to \mathcal{X}$ such that

$$X_n = f_{\text{inv}}(X_{n-N}^{n-1}, Y_{n-M}^{n}) = f_{\text{inv}}(Y_n, \Theta_n) = f_{\Theta_n}^{-1}(Y_n). \quad (26)$$

In other words, a system is partially invertible if its parameterized static function f_{Θ_n} is invertible for all possible parameter values $\theta \in \mathcal{T}$.

We will now argue that for this class of systems the information loss rate vanishes. We start by showing that the total information loss for a finite-length input sequence X_1^K after observing an output sequence Y_1^K of the same length remains bounded independently of the sequence length:

Theorem 4. *Let* X_1^K *and* Y_1^K*,* $K > \max\{M, N\}$*, be two finite-length sequences of jointly stationary processes* $\mathbf{X}$ *and* $\mathbf{Y}$ *on countable sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively, where* $\mathbf{Y}$ *is generated by passing* $\mathbf{X}$ *through a partially invertible system. Then, the information loss becomes*

$$H(X_1^K|Y_1^K) = H(X_1^{\max\{M,N\}}|Y_1^K). \quad (27)$$

Proof: We start by noticing that $H(X_1^K|Y_1^K) = H(X_1^K, Y_1^K) - H(Y_1^K)$ and

$$\begin{aligned}H(X_1^K, Y_1^K) &= H(X_K, X_1^{K-1}, Y_1^K) && (28)\\ &= H(f_{\text{inv}}(X_{K-N}^{K-1}, Y_{K-M}^K), X_1^{K-1}, Y_1^K)\\ &\overset{(a)}{=} H(X_1^{K-1}, Y_1^K) && (29)\end{aligned}$$

where (a) is due to Lemma 1. Repeating this step a number of times yields

$$H(X_1^K, Y_1^K) = H(X_1^{\max\{M,N\}}, Y_1^K). \quad (30)$$

Subtracting $H(Y_1^K)$ completes the proof. ■

It is important to note that $H(X_1^K|Y_1^K) \neq H(X_1^K) - H(Y_1^K)$. While the information loss rate is equal to the difference of entropy rates (cf. Theorem 1), it does not hold generally that the difference of joint entropies is equal to the joint conditional entropy.

We will now make use of this result in proving that partially invertible systems have a vanishing information loss rate:

Corollary 2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be jointly stationary processes on countable sets related as in Theorem 4. Then, the information loss rate induced by passing the process* $\mathbf{X}$ *through the system vanishes, i.e.,*

$$\bar{H}(\mathbf{X}|\mathbf{Y}) = 0. \quad (31)$$

Proof: We provide two proofs for this Corollary. For the first, note that irrespective of θ the inverse function f_θ^{-1} always exists by Definition 3. With Theorem 2 this immediately leads to $\bar{H}(\mathbf{X}|\mathbf{Y}) = 0$.

For the second proof we note that Theorem 4 holds for all K, thus also in the limit. With Definition 2 we can therefore write the information loss rate as:

$$\begin{aligned}\bar{H}(\mathbf{X}|\mathbf{Y}) &= \lim_{n\to\infty}\frac{1}{n}H(X_1^n|Y_1^n) && (32)\\ &= \lim_{n\to\infty}\frac{1}{n}H(X_1^{\max\{M,N\}}|Y_1^n) && (33)\\ &\leq \lim_{n\to\infty}\frac{1}{n}H(X_1^{\max\{M,N\}}) = 0 && (34)\end{aligned}$$

by similar arguments as in the proof of Lemma 2. ■

An immediate consequence of this important Corollary is that, except for the initial samples $X_1^{\max\{M,N\}}$ after starting the observation of $\mathbf{Y}$ (cf. Theorem 4), the remaining information of the input process can be recovered by observing the output process.

V. The Case of Linear Filters

It is interesting to note that an important subclass of discrete-time stable causal linear filters falls in the category of partially invertible systems, as long as the input and output alphabets are countable. This subclass comprises filters with a finite-dimensional state vector described by constant-coefficient difference equations:

$$Y_n = \sum_{k=0}^{N} b_k X_{n-k} + \sum_{l=1}^{M} a_l Y_{n-l} \tag{35}$$

By rearranging the terms in (35) it can be verified that this subclass of linear systems satisfies the definition of partially invertible systems and, thus, has a vanishing information loss rate.

It is noteworthy that this property is independent of the minimum-phase property (cf. [10, pp. 280]) of linear filters, which ensures that the filter has a stable and causal inverse. Indeed, for filters which are not minimum-phase, the partial inverse function f_{inv} used in Definition 3 describes a causal, but unstable linear filter. As a consequence, to an arbitrary stationary stochastic input process, the inverse filter described by f_{inv} may respond with a non-stationary output process; however, the response to $\mathbf{Y}$ will be $\mathbf{X}$.

This complements a result already introduced by Shannon [5], which states that the change in *differential* entropy rate caused by stable, causal linear filtering of continuous-valued stationary processes is independent of the process statistics. In particular, for a linear filter with frequency response $G(\mathrm{e}^{\jmath\theta})$ the differential entropy rate of the output is given by [6, pp. 663]

$$\bar{h}\left(\mathbf{Y}\right) = \bar{h}\left(\mathbf{X}\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|G(\mathrm{e}^{\jmath\theta})| d\theta. \tag{36}$$

It can be shown (see, e.g., [11]) that the integral above evaluates to $\ln|b_0| + \sum_{i:|z_i|>1} \ln|z_i|$, where z_i are the zeros of the transfer function $G(z)$. For causal minimum-phase systems ($|z_i| < 1\ \forall i$) with $b_0 = 1$, the differential entropy rates for the input and output process are equal. Scaling the transfer function of such a filter such that $b_0 \neq 1$ leads to $\bar{h}\left(\mathbf{X}\right) \neq \bar{h}\left(\mathbf{Y}\right)$, despite the fact that by scaling no information is lost. Conversely, it is easily possible that $\bar{h}\left(\mathbf{X}\right) = \bar{h}\left(\mathbf{Y}\right)$ for systems which destroy information. Therefore, we believe that differential entropies and differential entropy rates are not adequate measures for information loss. Future investigations will show if alternative descriptions for continuous-valued processes will yield more appropriate characterizations.

VI. Other Examples

While the case of linear filters is a particularly interesting one, the restriction to countable input and output alphabets suggests further examples illustrating the application of our theoretical results.

A. Example 1: Finite-Precision Linear Filters

The first example considers an extension to the subclass of discrete-time linear filters discussed in Section V. In many practical applications in digital signal processing linear filters are implemented with finite-precision number representations only. We thus assume that both input process and filter coefficients take values from a finite set. For example, $\mathcal{X}$ may be a finite subset of the rational numbers $\mathbb{Q}$, closed under modulo-addition. Multiplying two values from that set, e.g., by multiplying an input sample with a filter coefficient, typically yields a result not representable in $\mathcal{X}$. As a consequence, after every multiplication a quantizer is necessary, essentially truncating the additional bits resulting from multiplication. Let the quantizer be described by a function $Q\colon \mathbb{R} \to \mathcal{X}$ with $Q\left(a + X_n\right) = Q\left(a\right) \oplus X_n$ if $X_n \in \mathcal{X}$, where $\oplus$ denotes modulo-addition (e.g., [10, pp. 373]). With this (35) changes to

$$Y_n = \bigoplus_{k=0}^{N} Q\left(b_k X_{n-k}\right) \oplus \bigoplus_{l=1}^{M} Q\left(a_l Y_{n-l}\right) \tag{37}$$

where quantization is performed after multiplication (the extension to systems which perform quantization after truncation can be found in [9]). Note that due to modulo-addition the result Y_n remains in $\mathcal{X}$.

We will now focus on filters with $b_0 = 1$. For filters with infinite precision this can be done without loss of generality by considering a constant gain factor b_0 and by normalizing all b_k coefficients. However, this gain normalization poses a restriction in the finite-precision case since b_k/b_0 is not necessarily an element of $\mathcal{X}$. With $b_0 = 1$ (37) changes to

$$Y_n = X_n \oplus \left(\bigoplus_{k=1}^{N} Q\left(b_k X_{n-k}\right) \oplus \bigoplus_{l=1}^{M} Q\left(a_l Y_{n-l}\right) \right) \tag{38}$$

by the property of the quantizer. Since the term in parentheses in (38) is in $\mathcal{X}$, and modulo-addition has an inverse element, it can be seen that this implementation is partially invertible. Consequently, even filters with nonlinear elements can be shown to preserve information under certain circumstances despite the fact that the quantizer function is non-injective.

B. Example 2: Multiplying Consecutive Inputs

Another nonlinear system satisfying Definition 3 is given by the following input-output relationship:

$$Y_n = X_n X_{n-1} \tag{39}$$

The partial inverse in this case would be $X_n = \frac{Y_n}{X_{n-1}}$ if $X_{n-1} \neq 0$, while for $X_{n-1} = 0$ no such inverse exists. Therefore, this example represents a class of systems whose partial invertibility depends on the alphabet $\mathcal{X}$ of the stochastic process. If the process $\mathbf{X}$ is such that $\mathcal{X}$ does not contain the element 0, the partial inverse exists and we obtain for X_n, $n > 1$:

$$X_n = \begin{cases} X_1 \prod_{k=1}^{\frac{n-1}{2}} \frac{Y_{2k+1}}{Y_{2k}}, & \text{for odd } n \\ \frac{Y_n}{X_1} \prod_{k=1}^{\frac{n}{2}-1} \frac{Y_{2k}}{Y_{2k+1}}, & \text{for even } n \end{cases} \tag{40}$$

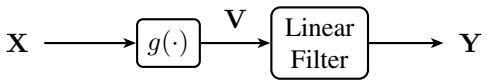

Fig. 1. Discrete-Time Hammerstein System.

Indeed, since all X_n, $n > 1$, can be computed from X_1 and Y_1^n, we obtain $H(X_1^n|Y_1^n) = H(X_1|Y_1^n)$ which is in perfect accordance with Theorem 4. Reconstruction of $\mathbf{X}$ is thus possible up to an unknown X_1. Note, however, that this unknown sample influences the whole reconstructed sequence as shown in (40).

C. Example 3: Hammerstein Systems

A final example considers a simple special case of a nonlinear dynamical system, namely, a cascade of a static nonlinearity and a linear filter [12]. Such a cascade, usually referred to as Hammerstein system, is depicted in Fig. 1. In the discrete-time case the input-output relationship is given by

$$Y_n = \sum_{k=0}^{N} b_k g(X_{n-k}) + \sum_{l=1}^{M} a_l Y_{n-l}. \tag{41}$$

As it is easily seen, this system is partially invertible if and only if the function g has an inverse. If g is not invertible, we obtain in the light of Theorem 2:

$$Y_n = f_{\Theta_n}(X_n) = b_0 g(X_n) + C_{\Theta_n} \tag{42}$$

where C_{Θ_n} is a constant depending on the random parameter Θ_n. With this and $f_\theta^{-1}[f_\theta(x)] = g^{-1}[g(x)]$ for all $x \in \mathcal{X}$, $\theta \in \mathcal{T}$ we obtain an upper bound on the information loss rate:

$$\bar{H}(\mathbf{X}|\mathbf{Y}) \leq \max_{x\in\mathcal{X}} \log |g^{-1}[g(x)]| \tag{43}$$

As mentioned in the introduction, the energy detector can be modeled as a Hammerstein system where the static nonlinearity $g(\cdot) = (\cdot)^2$ is followed by a moving-average filter (the downsampling operation is omitted here). For an input alphabet $\mathcal{X}$ with positive and negative elements of the same magnitude, the upper bound in (43) evaluates to 1 bit.

Interestingly, the structure of a Hammerstein system allows a simplified analysis: Since the information loss rate of a cascade of systems is equal to the sum of individual information loss rates (cf. Theorem 3) we can analyze both constituent systems separately. The linear filter was already shown to preserve full information, so any information loss will be caused by the static nonlinearity, i.e., $\bar{H}(\mathbf{X}|\mathbf{Y}) = \bar{H}(\mathbf{X}|\mathbf{V})$. This is in accordance with the observation that the Hammerstein system is partially invertible if the static nonlinearity is invertible.

For static nonlinearities the analytic treatment of information loss is simple compared to dynamical systems. In particular, for an independent, identically distributed (iid) input process $\mathbf{X}$ the information loss rate can be shown to be equal to the zeroth-order conditional entropy, $H(X|V)$, while for a general stationary process this quantity acts as an upper bound [3]. The upper bound from Theorem 2 turns out to be even more general, since it also provides an upper bound on $H(X|V)$ in the case of an iid input process (cf. Theorem 4 in [7]). An in-depth analysis of the interplay between these bounds is the object of future work.

VII. Conclusion

In this work we have presented general results on the information loss of dynamical systems for stationary stochastic input and output processes on countable alphabets. Furthermore, we have extended the proof of the data processing inequality stating that the entropy rate at the output of the system cannot be larger than the entropy rate at the input and have derived an upper bound on the information loss rate. The additivity of information loss rates for cascaded systems could be shown, too.

We have further identified a family of systems for which this upper bound is zero, i.e., for which the information loss rate vanishes. Not only linear filters belong to that family, but also their nonlinear counterparts common in finite-precision signal processing.

Future research will extend these results to the case of continuous-valued stochastic processes and the application to common nonlinear systems, e.g., Volterra models.

Acknowledgments

The authors gratefully acknowledge discussions with Sebastian Tschiatschek concerning mathematical notation.

References

[1] K. Witrisal, G. Leus, G. J. Janssen, M. Pausini, F. Troesch, T. Zasowski, and J. Romme, "Noncoherent ultra-wideband systems," *IEEE Signal Process. Mag.*, vol. 26, no. 4, pp. 48–66, July 2009.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley Interscience, 2006.

[3] S. Watanabe and C. T. Abraham, "Loss and recovery of information by coarse observation of stochastic chain," *Information and Control*, vol. 3, no. 3, pp. 248–278, Sep. 1960.

[4] M. S. Pinsker, *Information und Informationsstabilität zufälliger Grössen und Prozesse*, ser. Arbeiten zur Informationstheorie, H. Grell, Ed. Berlin: Deutscher Verlag der Wissenschaften, 1963, vol. V, (Information and information stability of random variables and processes, German translation of the original Russian text).

[5] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, Oct. 1948.

[6] A. Papoulis and U. S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY: McGraw Hill, 2002.

[7] B. C. Geiger, C. Feldbauer, and G. Kubin, "Information loss in static nonlinearities," in *Proc. IEEE Int. Sym. Wireless Communication Systems (ISWSC)*, Aachen, Nov. 2011, accepted; preprint available: `arXiv:1102.4794 [cs.IT]`.

[8] N. Pippenger, "The average amount of information lost in multiplication," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 684–687, Feb. 2005.

[9] B. C. Geiger and G. Kubin, "Some results on the information loss in dynamical systems," Jun. 2011, `arXiv:1106.2404 [cs.IT]`.

[10] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc., 1999.

[11] S. Yu and P. G. Mehta, "The Kullback-Leibler rate pseudo-metric for comparing dynamical systems," *IEEE Trans. Autom. Control*, vol. 55, no. 7, pp. 1585–1598, Jul. 2010.

[12] Y. S. Shmaliy, *Continuous-Time Systems*. Dordrecht: Springer, 2007.