Source Coding with Feed-Forward

Ramji Venkataramanan
University of Michigan
rvenkata@eecs.umich.edu

S. Sandeep Pradhan [1]
University of Michigan
pradhanv@eecs.umich.edu

Abstract — **In this work, we consider a source coding model with feed-forward. We analyze a system with a noiseless feed-forward link where the decoder has knowledge of all previous source samples while reconstructing the present sample. The rate-distortion function for an arbitrary source with feed-forward is derived in terms of directed information, a variant of mutual information. The special case of Gaussian sources with feed-forward is further examined. We also derive an error exponent which is used to bound the probability of decoding error for a source code (with feed- forward) of finite block length. Source coding with feed-forward may be considered the dual problem of channel coding with feedback.**

I. Introduction

With the recent emergence of applications involving sensor networks [1], the problem of source coding with side-information at the decoder gained special significance [2]. Here, the encoder represents the source with an index based on the knowledge that the decoder has access to some correlated side-information. In a typical setting, at each instant of time, the source produces a symbol X and a sample of the side-information Y appears at the decoder. We are interested in considering a variant of this problem, where there is a delay in the side-information available at the decoder. For instance, if the delay is 3 time units, the sequence of events at the (encoder, decoder) would be $(X_1,-),(X_2,-),(X_3,-),(X_4,Y_1),(X_5,Y_2)$ and so on. We would like to analyze this problem of source coding with delayed side-information.

Frequently the side information Y is a noisy version of X. Thus, we would expect that Y_1 be strongly correlated with X_1, Y_2 with X_2 and so on. Such a model would be relevant in applications involving estimation of an information field (e.g a seismic/acoustic signal) in a sensor network. A node may have to estimate (compressed) signals received from other nodes and process these signals in real-time. However, the signal to be estimated might be available at the node in a delayed and perhaps, noisy form, i.e., there is a feed-forward path from the source to the decoder. Thus an efficient decoder must take into account all the information available while decoding a particular sample. In this work, we consider an idealized version of this problem called source coding with feed-forward [3]. In this model, we assume that noiseless source samples are available with a delay at the decoder, i.e. $Y = X$.

[1]This work was supported by NSF ITR Grants CCR-0329715 and CCR-0219735

Related Work: The problem of source coding with noiseless feed-forward arose in the context of competitive prediction in [4], where it was shown that for IID discrete sources feedforward does not reduce the optimal rate-distortion function and the optimal error-exponent with block coding. Around the same time, the model of source coding with feed-forward was defined in [3] as a variant of the problem of source coding with side information [2] at the decoder, and a simple and deterministic block-coding scheme to achieve the optimal rate-distortion bound for arbitrary rates for an IID Gaussian sources with feed-forward was described. At the time of writing this paper, we also became aware of another work [5] which gives a variable-length coding strategy to achieve the rate-distortion bound for any finite-alphabet, IID source with feed-forward. The problem of source coding with feed-forward is also related to source coding with a delay-dependent distortion function [6] and causal source coding [7].

The main results of the present paper can be summarized as follows:

1. The optimal rate-distortion function for a general discrete source with general distortion measures and with noiseless feed-forward, $R_{ff}(D)$, is given by the minimum of the directed information function [8] between the source and the reconstruction. $R_{ff}(D) \leq R(D)$, where $R(D)$ denotes the optimal Shannon rate-distortion function for the source without feed-forward.

2. The performance of the best possible source code (with feed-forward) of rate R, distortion D and block length N is characterized by an error exponent $E_{N-ff}(R,D)$. $E_{N-ff}(R,D)$ is greater than or equal to the error exponent without feed-forward.

3. Feed-forward does not decrease the rate-distortion function of general discrete memoryless sources with memoryless distortion measures.

II. The Source Coding Model

The model is shown in Figure 1. Consider a discrete source X with Nth order probability distribution P_{X^N}, alphabet $\mathcal{X}$ and reconstruction alphabet $\hat{\mathcal{X}}$. There is an associated distortion measure $d_N : \mathcal{X}^N \times \hat{\mathcal{X}}^N \rightarrow \mathbb{R}^+$ for pairs of sequences of length N. We assume that $d_N(\cdot,\cdot)$ is normalized with respect to N and is uniformly bounded in N. The distortion measure is said to be memoryless if $\forall x^N \in \mathcal{X}^N$ and $\hat{x}^N \in \hat{\mathcal{X}}^N$,

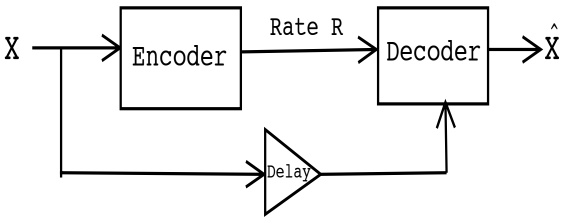

Figure 1: Source coding system with feed-forward.

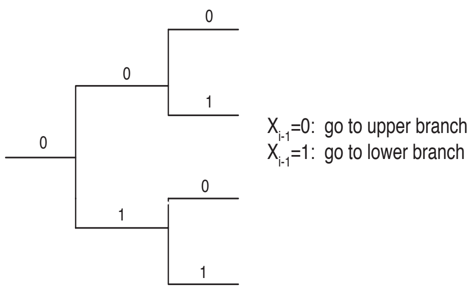

Figure 2: Code function represented as a tree. The reconstruction $\hat{X}$ is represented on the branches of the tree.

$d_N(x^N, \hat{x}^N) = \frac{1}{N}\sum_{i=1}^{N} d_i'(x_i, \hat{x}_i)$ for some $d_i' : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$, $\forall i$.

For a source code of block length N and rate R, the encoder is a mapping to an index set: $e : \mathcal{X}^N \to \{1, \ldots, 2^{NR}\}$. The decoder receives the index transmitted by the encoder, and to reconstruct the ith sample, it has access to all the past $(i-1)$ samples of the source. In other words, the decoder is a sequence of mappings $g_i : \{1, \ldots, 2^{NR}\} \times \mathcal{X}^{i-1} \to \widehat{\mathcal{X}}, \quad i = 1, \ldots, N$. Let $\hat{x}^N$ denote the reconstruction of the source sequence x^N. We want to minimize the distortion for a given rate R. For any D, let $R_{ff}(D)$ denote the infimum of R over all encoder decoder pairs for any block length N such that the distortion is less than D. It is worthwhile noting that source coding with feed-forward can be considered the dual problem [9, 10] of channel coding with feedback.

We describe the set-up in Section III. Section IV contains the heart of this work- rate-distortion functions for sources with feed-forward. Section V deals with the performance of the best possible source codes for a finite block length. Due to constraints of space, most theorems are stated without proof. However, we attempt to give a broad idea of the proof wherever possible.

III. Source, Encoder and Decoder Set-up

In this section, we describe the apparatus we will use for proving coding theorems for sources with feed-forward. We introduce code-functions, which map the feed-forward information to a source reconstruction symbol $\hat{X}$. The idea of code-functions was introduced by Shannon in 1961 [11]. We first give a formal definition of a code-function and then see how it is useful in analyzing systems with feed-forward.

Definition 1. *A source code-function f^N is a set of N functions $\{f_n\}_{n=1}^{N}$ such that $f_n : \mathcal{X}^{n-1} \to \widehat{\mathcal{X}}$ maps each source sequence $x^{n-1} \in \mathcal{X}^{n-1}$ to a reconstruction symbol $\hat{x}_n \in \widehat{\mathcal{X}}$. Denote the space of all code-functions by $\mathcal{F}^N = \mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \mathcal{F}_N \triangleq \{f^N : f^N$ is a code function$\}$.*

Definition 2. *A $(N, 2^{NR})$ source codebook of rate R and block length N is a set of 2^{NR} code-functions. Denote them by $f^N[w], \quad w = 1, \ldots, 2^{NR}$.*

For each source sequence of length N, the encoder sends an index to the decoder. Using the code-function corresponding to this index, the decoder maps the information fed forward from the source to produce an estimate $\hat{X}$. A code-function can be represented as a tree. Figure 2 shows a code-function for a binary source with a binary reconstruction alphabet. Using the code-function shown in the figure, a source sequence (001) would be reconstructed as (000) and (101) would be reconstructed as (010). In a system without feed forward, a code-function generates the reconstruction independent of the past source samples. In this case, the code-function reduces to a codeword. In other words, for a system without feed-forward, a source codeword is a source code-function $f^N = \{f_1, \ldots, f^N\}$ where for each $n \in \{1, \ldots, N\}$, the function f_n is a constant mapping.

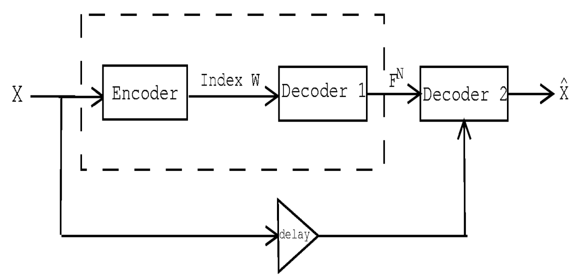

Figure 3: Representation of a source coding scheme with feed-forward.

A source code with feed-forward can be thought of as having two components. The first is a usual source coding problem with F^N as the reconstruction for the source sequence X^N. In other words, for each source sequence x^N, the encoder chooses the best code-function among $f^N[i], \quad i \in \{1, \ldots, 2^{NR}\}$ and sends the index of the chosen code function. This is the part inside the dashed box in Figure 3. If we denote the chosen code-function by f^N, the second component (decoder 2 in Fig. 3) produces the reconstruction given by

$$\hat{X}_i = f_i(X^{i-1}), \qquad i = 1, \ldots, N, \tag{1}$$

IV. Coding Theorems

A. Discrete Memoryless Sources: We start with the simplest kind of source, viz. a discrete memoryless source. The optimal

Shannon rate distortion function for an IID source without feed-forward is given by

$$R_{DM}(D) = \min_{P(\hat{x}|x): \sum_{(x,\hat{x})} P(x)P(\hat{x}|x)d(x,\hat{x}) \leq D} I(X;\hat{X}) \quad (2)$$

We state the following theorem without proof for a discrete memoryless source with feed-forward with expected distortion constraint.

Theorem 1. *Feed-forward does not decrease the optimal rate-distortion function of a general discrete memoryless source with memoryless distortion measures.*

It parallels the well known result that feedback does not increase the rate-distortion function of a discrete memoryless channel [12].

B. Arbitrary Sources: This section contains the main contribution of this paper- the optimal rate-distortion function for an arbitrary source with feed-forward. For a source without feed-forward, the rate-distortion function is characterized by the mutual information between X and $\hat{X}$. It turns out that for sources with feed-forward, the rate-distortion function is characterized by directed information, a variant of mutual information.

B.1 Directed Information

The directed information function was introduced by Massey [8] and has been used to characterize the capacity of channels with feedback [13] [14].

Definition 3. *The directed information flowing from a sequence A^N to a sequence B^N is defined as*

$$I(A^N \to B^N) = \sum_{n=1}^{N} I(A^n; B_n | B^{n-1}). \quad (3)$$

Note that the definition is similar to that of mutual information $I(A^N; B^N)$ except that the mutual information has A^N instead of A^n in the sum on the right.

The directed information has a nice interpretation in the context of our problem. The directed information flowing from $\hat{X}^N$ to X^N can be written as

$$I(\hat{X}^N \to X^N) = \sum_{i=1}^{N} I(\hat{X}^i; X_i | X^{i-1})$$

$$= I(\hat{X}^N; X^N) - \sum_{i=1}^{N} I(X^{i-1}; \hat{X}_i | \hat{X}^{i-1}) \quad (4)$$

We know that for the usual source coding problem (without feed-forward), the mutual information $I(X^N; \hat{X}^N)$ represents the minimum number of bits needed to represent X^N by $\hat{X}^N$. With feed-forward, the decoder knows the symbols X^{i-1} to reconstruct $\hat{X}_i$. This is reflected in the terms subtracted from $I(X^N; \hat{X}^N)$ in (4). (4) says that since the information $I(X^{i-1}; \hat{X}_i | \hat{X}^{i-1})$ is already known through the feed-forward link, we need not spend bits to code this information. Consequently, it is reasonable to expect that the directed information characterizes the rate-distortion function for sources with feed-forward.

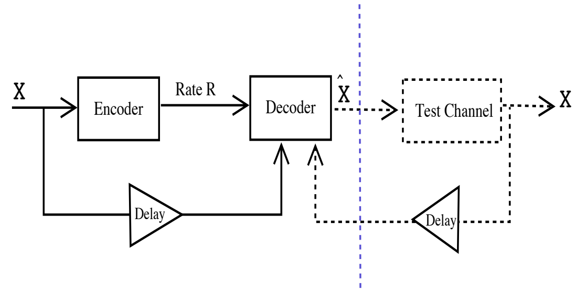

Figure 4: Source coding system with feed-forward.

We can also interpret the directed information in terms of the backward test-channel $\hat{X}^N - X^N$. A source code with feed-forward can be thought of as having feedback in the test-channel and the directed information gives the information flow through the channel with feedback.

B.2 Rate Distortion function for sources with Feed-Forward

We now give the rate-distortion function for arbitrary sources with feed-forward. Before stating the general result, we need the following definitions of a few quantities (see [15],[14]).

Definition 4. *The* limsup in probability *of a sequence of random variables $\{X_n\}$ is defined as the smallest extended real number α such that* $\forall \epsilon > 0$

$$\lim_{n\to\infty} Pr[X_n \geq \alpha + \epsilon] = 0.$$

The liminf in probability *of a sequence of random variables $\{X_n\}$ is defined as the largest extended real number β such that* $\forall \epsilon > 0$

$$\lim_{n\to\infty} Pr[X_n \leq \beta - \epsilon] = 0.$$

Definition 5. *For any sequence of joint distributions $\{P_{X^N,\hat{X}^N}\}_{n=1}^{\infty}$, define $\forall x^N \in \mathcal{X}^N, \hat{x}^N \in \hat{\mathcal{X}}^N$*

$$\vec{P}_{\hat{X}^N|X^N}(\hat{x}^N|x^N) \triangleq \prod_{i=1}^{N} P_{\hat{X}_i|\hat{X}^{i-1},X^{i-1}}(\hat{x}_i|\hat{x}^{i-1}, x^{i-1}),$$

$$\vec{P}_{X^N|\hat{X}^N}(x^N|\hat{x}^N) \triangleq \prod_{i=1}^{N} P_{X_i|\hat{X}^i,X^{i-1}}(x_i|\hat{x}^i, x^{i-1}).$$

$$\overline{I}(\hat{X} \to X) \triangleq \limsup_{inprob} \frac{1}{N} \log \frac{P_{X^N,\hat{X}^N}(x^N,\hat{x}^N)}{\vec{P}_{\hat{X}^N|X^N}(\hat{x}^N|x^N)P_{X^N}(x^N)}$$

$$\underline{I}(\hat{X} \to X) \triangleq \liminf_{inprob} \frac{1}{N} \log \frac{P_{X^N,\hat{X}^N}(x^N,\hat{x}^N)}{\vec{P}_{\hat{X}^N|X^N}(\hat{x}^N|x^N)P_{X^N}(x^N)}$$

As pointed out in [14], the directed information rate, defined by $\lim_{n\to\infty}\frac{1}{n}\log I(\hat{X}^n \to X^n)$ may not exist for an arbitrary random process which may not be stationary. But the sup-directed information rate $\overline{I}(\hat{X}\to X)$ and the inf-directed information rate $\underline{I}(\hat{X}\to X)$ always exist. Tatikonda and Mitter [14] showed that for arbitrary channels with feedback, the capacity is an optimization of $\underline{I}(X\to Y)$, the inf-directed information rate. Our result is that the rate distortion function for an arbitrary source with feed-forward is an optimization of $\overline{I}(\hat{X}\to X)$, the sup-directed information rate.

The source distribution, defined by a sequence of finite-dimensional distributions [16], is denoted by

$$\mathbf{P_X} \triangleq \{P_{X^n}\}_{n=1}^{\infty}. \tag{5}$$

Similarly, a conditional distribution is denoted by

$$\mathbf{P_{\hat{X}|X}} \triangleq \{P_{\hat{X}^n|X^n}\}_{n=1}^{\infty}. \tag{6}$$

Theorem 2. *For an arbitrary source X characterized by a distribution $\mathbf{P_X}$, the rate-distortion function with feed-forward, the infimum of all achievable rates at a distortion D, is given by*

$$R_{ff}(D) = \inf_{\mathbf{P_{\hat{X}|X}}:\rho(\mathbf{P_{\hat{X}|X}})\le D} \overline{I}(\hat{X}\to X), \tag{7}$$

where

$$\rho(\mathbf{P_{\hat{X}|X}}) \triangleq \limsup_{inprob} d_n(x^n,\hat{x}^n)$$
$$= \inf\left\{h: \lim_{n\to\infty} P_{X^n}P_{\hat{X}^n|X^n}\left((x^n,\hat{x}^n): d_n(x^n,\hat{x}^n) > h\right) = 0\right\}.$$

From [14], we have the following result. For any sequence of joint distributions $\{P_{X^N,\hat{X}^N}\}_{n=1}^{\infty}$, we have

$$\begin{aligned}\underline{I}(\hat{X}\to X) &\le \liminf_{N\to\infty}\frac{1}{N}I(\hat{X}^N\to X^N)\\ &\le \limsup_{N\to\infty}\frac{1}{N}I(\hat{X}^N\to X^N) \le \overline{I}(\hat{X}\to X).\end{aligned} \tag{8}$$

If

$$\underline{I}(\hat{X}\to X) = \overline{I}(\hat{X}\to X)$$

we say that the process $\{P_{X^n,\hat{X}^n}\}_{n=1}^{\infty}$ is information stable [17], and all four quantities in (8) are equal. Note that if the joint process $\left\{X_n,\hat{X}_n\right\}_{n=1}^{\infty}$ is information stable, the rate-distortion function becomes

$$R_{ff}(D) = \inf_{\mathbf{P_{\hat{X}|X}}:\rho(\mathbf{P_{\hat{X}|X}})\le D}\lim_{N\to\infty}\frac{1}{N}I(\hat{X}^N\to X^N). \tag{9}$$

We do not give the detailed proofs of the direct and converse parts of Theorem 2. Instead, we give a brief idea of the direct part here. For the sake of intuition, assume information stability. We want to show the achievability of all rates greater than the $R_{ff}(D)$ in (9).

Let $P^*(\hat{X}^N|X^N)$ be the distribution that maximizes $I(\hat{X}^N \to X^N)$, subject to the constraint. Our goal is to construct a joint distribution over $X^N, \hat{X}^N$ and F^N, say $Q_{F^N,X^N,\hat{X}^N}$, such that the marginal over X^N and $\hat{X}^N$ satisfies

$$Q_{X^N,\hat{X}^N} = P_{X^N}P^*_{\hat{X}^N|X^N}. \tag{10}$$

We also impose certain additional constraints [1] on $Q_{F^N,X^N,\hat{X}^N}$ so that

$$I_Q(F^N;X^N) = I_Q(\hat{X}^N\to X^N). \tag{11}$$

Using (10) in the above equation, we get

$$I_Q(F^N;X^N) = I_{P_{X^N}P^*_{\hat{X}^N|X^N}}(\hat{X}^N\to X^N). \tag{12}$$

Using the usual techniques for source coding without feed-forward, it can be shown that all rates greater than $\frac{1}{N}I_Q(F^N;X^N)$ can be achieved. From (12), it follows that all rates greater than $I_{P_{X^N}P^*_{\hat{X}^N|X^N}}$. The bulk of the proof lies in constructing a suitable joint distribution Q.

C. Gaussian Sources with feed-forward: In this section, we study the rate-distortion function for the special case of Gaussian sources with feed-forward. A source X is Gaussian if the random process $\{X_n\}_{n=1}^{\infty}$ is jointly Gaussian. A Gaussian source is continuous valued unlike the sources hitherto discussed. However, it is straightforward to extend the results derived earlier for discrete sources to continuous sources. In particular, feed-forward does not decrease the rate-distortion function of a memoryless Gaussian source. Interestingly though, feed-forward in a memoryless Gaussian source enables us to achieve rates arbitrarily close to the rate-distortion function with a low complexity coding scheme involving just scalar quantization [3]. We have the following result for a general Gaussian source.

Theorem 3. *Gaussian conditional distributions achieve the rate-distortion function for Gaussian sources with feed-forward and with expected quadratic distortion constraint.*

We give a sketch of the proof. Let X be a Gaussian source with distribution $\mathbf{P_X}$ and let $\mathbf{P_{\hat{X}|X}}$ be any conditional distribution. We show that there exists a jointly Gaussian conditional distribution $G_{\hat{X}^N|X^N}$ such that $G_{X^N,\hat{X}^N} = P_{X^N}\cdot G_{\hat{X}^N|X^N}$ is a jointly Gaussian distribution that has the same second order properties as $P_{X^N,\hat{X}^N} = P_{X^N}\cdot P_{\hat{X}^N|X^N}$ and the following hold.

1. $I_G(\hat{X}^N\to X^N) \le I_P(\hat{X}^N\to X^N)$
2. The average distortion is the same under both distributions, i.e.,

$$E_P[d_N(X^N,\hat{X}^N)] = E_G[d_N(X^N,\hat{X}^N)]. \tag{13}$$

[1] For clarity, wherever necessary, we will indicate the distribution used to calculate the information quantity as a subscript.

This means we can restrict our attention to Gaussian conditional distributions to evaluate the rate-distortion function of a Gaussian source.

V. Error Exponents

We now consider error exponents for sources with feed-forward and show that the feed-forward error exponent is no smaller than the exponent for the same source without feed-forward.

A. Upper bound on the probability of error: The error-exponent for a source code of block-length N for a discrete memoryless source was derived by Blahut[18] and by Marton in [19]. A procedure identical to the proof of Theorem 6.5.1 in [18] yields the error exponent for an arbitrary source (without feed-forward). Therefore, we have the following fact for discrete sources without feed-forward.

Given a source with N-th order distribution P_{X^N}, there exists a $(N, 2^{NR})$ source code (without feed-forward) such that the probability that a source sequence of length N cannot be encoded with distortion $\leq D$ satisfies

$$P_e \leq e^{-NE_N(R,D)+o(N)}, \tag{14}$$

where $E_N(R,D)$ is the error exponent for the source (without feed-forward) and is given by

$$E_N(R,D) = \max_{s\geq 0} \min_{t\leq 0} \max_{q_{\hat{X}^N}} \Bigg[sR - stD - \frac{1}{N}\log_2 \sum_{x^N} P_{X^N}(x^N) \left(\sum_{\hat{x}^N} q_{\hat{X}^N}(\hat{x}^N) e^{td(x^N,\hat{x}^N)} \right)^{-s} \Bigg], \tag{15}$$

and for large enough N, $o(N) = 0$.

The proof of this in [18] involves choosing random codewords with distribution $q_{\hat{X}^N}$. For a source code with feed-forward, the decoder knows x^{i-1} to decode $\hat{x}_i$. So we can choose codewords with distribution

$$\vec{q}_{\hat{X}^N|x^N} = \prod_i q_{\hat{X}_i|\hat{X}^{i-1},x^{i-1}}$$

By randomly picking codewords with the above distribution, we can derive the error exponent for a source with feed-forward.

Theorem 4. *Given a source with N-th order distribution P_{X^N}, there exists a $(N, 2^{NR})$ source code with feed-forward so that the probability that a source sequence of length N cannot be encoded with distortion $\leq D$ satisfies*

$$P_e \leq e^{-NE_{ff-N}(R,D)+o(N)}, \tag{16}$$

where $E_{ff-N}(R,D)$ is the error exponent for the source (with feed-forward) and is given by

$$E_{ff-N}(R,D) = \max_{s\geq 0} \min_{t\leq 0} \max_{\vec{q}_{\hat{X}^N|X^N}} \Bigg[sR - stD - \frac{1}{N}\log_2 \sum_{x^N} P_{X^N}(x^N) \left(\sum_{\hat{x}^N} \vec{q}_{\hat{X}^N|X^N}(\hat{x}^N|x^N) e^{td(x^N,\hat{x}^N)} \right)^{-s} \Bigg], \tag{17}$$

where

$$\vec{q}_{\hat{X}^N|X^N}(\hat{x}^N|x^N) = \prod_{i=1}^N q_{\hat{X}_i|X^{i-1},\hat{X}^{i-1}}(\hat{x}_i|x^{i-1},\hat{x}^{i-1}).$$

We now compare the error exponents for a source with and without feed-forward given by Eqs.(17) and (15), respectively. Denote the space of all distributions of the form $q_{\hat{X}^N}$ by $\mathcal{S}_q$ and the space of all distributions of the form $\vec{q}_{\hat{X}^N|X^N}$ by $\mathcal{S}_{\vec{q}}$. The only difference between the expressions for the error exponents with and without feed-forward is that the former involves a maximization over distributions in $\mathcal{S}_q$, while the latter involves a maximization over $\mathcal{S}_{\vec{q}}$.

Now, every distribution $q_{\hat{X}^N} = \prod_{i=1}^N q_{\hat{X}_i|\hat{X}^{i-1}}$ belongs to the space of distributions of the form $\vec{q}_{\hat{X}^N|X^N} = \prod_{i=1}^N q_{\hat{X}_i|\hat{X}^{i-1},X^{i-1}}$. Therefore,

$$\mathcal{S}_q \subset \mathcal{S}_{\vec{q}}.$$

Thus in the no feed-forward case, we are maximizing over a subset of the distributions available to us in the feed-forward case. Equivalently, we have proved the following theorem.

Theorem 5. *For any source X, the error exponent with feed-forward is at least as large as the error exponent without feed-forward.*

Equation (16) guarantees an exponentially small probability of error only when $E_{ff-N}(R,D)$ is positive. An alternate definition of the error exponent is better suited to determine the values of R for which $E_{ff-N}(R,D)$ is positive. We first have the following definition.

Definition 6.

$$B_N(\hat{p}_{X^N}, D) \triangleq \min_{q_{\hat{X}^N|X^N}: \sum_{x^N,\hat{x}^N} E_{\hat{p}q}[d(x^N,\hat{x}^N) \leq D]} \frac{1}{N} I_{\hat{p}q}(\hat{X}^N \to X^N)$$

where the subscript denotes the joint distribution used to calculate the directed information.

We state the following theorem without proof.

Theorem 6. *An equivalent representation of $E_{ff-N}(R,D)$ is*

$$E_{ff-N}(R,D) = \min_{\hat{p}_{X^N}\in\mathcal{P}} \frac{1}{N} \sum_{x^N} \hat{p}_{X^N}(x^N) \log \frac{\hat{p}_{X^N}(x^N)}{P_{X^N}(x^N)}, \tag{18}$$

where

$$\mathcal{P} = \{ \hat{p}_{X^N} : B_N(\hat{p}_{X^N}, D) \geq R \}. \tag{19}$$

The quantity on the right hand side of (18) is a discrimination. It is 0 iff the source distribution $P_{X^N} \in \mathcal{P}$ and positive otherwise. From the definition of $\mathcal{P}$, it follows that $P_{X^N} \in \mathcal{P}$ if $R \leq B_N(P_{X^N}, D)$. Thus we have the following theorem.

Theorem 7. *$E_{ff-N}(R,D)$ is strictly positive only for rates R such that*

$$R > B_N(P_{X^N}, D).$$

REFERENCES

[1] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Mobile networking for smart dust," *ACM/IEEE International Conference on Mobile Computing*, Seattle, WA, August 1999.

[2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, January 1976.

[3] S. S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proc. of IEEE International Symposium on Information Theory*, (Chicago), 2004.

[4] T. Weissman and N. Merhav, "On competitive prediction and its relation to rate-distortion theory," *IEEE Transactions on Information theory*, vol. IT-49, pp. 3185–3194, December 2003.

[5] E. Martinian and G. W. Wornell, "Source Coding with Fixed Lag Side Information," *To appear in Proceedings of the 42nd Annual Allerton Conference (Monticello, IL)*, 2004.

[6] S. I. Krich, "Coding for a Delay-Dependent Fidelity Criterion," *IEEE Transactions on Information Theory*, pp. 77–85, January 1974.

[7] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Trans. Inform. Theory*, pp. 701–713, Sept. 1982.

[8] J. Massey, "Causality, Feedback and Directed Information," *Proceedings of the 1990 Symposium on Information Theory and its Applications (ISITA-90)*, pp. 303–305, 2004.

[9] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side-information case," *IEEE Transactions on Information Theory*, vol. 49, pp. 1181–1203, May 2003.

[10] R. J. Barron, B. Chen, and G. W. Wornell, "The Duality between Information Embedding and Source Coding with Side Information and Some Applications," *IEEE Transactions on Information Theory*, vol. 49, pp. 1159–1180, May 2003.

[11] C. E. Shannon, "Two-Way Communication Channels," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA*, vol. 1, pp. 611–644, 1961.

[12] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. IT-2, pp. 8–19, 1956.

[13] G. Kramer, *Directed Information for channels with Feedback*. PhD thesis, Swiss Federal Institute of Technology, Zurich, 1998.

[14] S. Tatikonda and S. Mitter, "The Capacity of Channels with Feedback-Part 1: The General Case," submitted to *IEEE Transactions on Information Theory*, October 2001.

[15] S. Verdú and T.Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, pp. 1147–1157, July 1994.

[16] T. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 40, pp. 1147–1157, July 1994.

[17] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA:Holden-Day, 1964. Translated by A.Feinstein.

[18] R. E. Blahut, *Principles and Practice of Information Theory*. Massachusetts: Addison Wesley, 1988.

[19] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 197–199, March 1974.