

# On Directed Information and Gambling

Haim H. Permuter  
Stanford University  
Stanford, CA, USA  
haim1@stanford.edu

Young-Han Kim  
University of California, San Diego  
La Jolla, CA, USA  
yhk@ucsd.edu

Tsachy Weissman  
Stanford University/Technion  
Stanford, CA, USA/Haifa, Israel  
tsachy@stanford.edu

**Abstract**—We study the problem of gambling in horse races with causal side information and show that Massey’s directed information characterizes the increment in the maximum achievable capital growth rate due to the availability of side information. This result gives a natural interpretation of directed information  $I(Y^n \rightarrow X^n)$  as the amount of information that  $Y^n$  causally provides about  $X^n$ . Extensions to stock market portfolio strategies and data compression with causal side information are also discussed.

## I. INTRODUCTION

Mutual information arises as the canonical answer to a variety of problems. Most notably, Shannon [1] showed that the capacity  $C$ , the maximum data rate for reliable communication over a discrete memoryless channel  $p(y|x)$  with input  $X$  and output  $Y$ , is given by

$$C = \max_{p(x)} I(X; Y), \quad (1)$$

which leads naturally to the operational interpretation of mutual information  $I(X; Y) = H(X) - H(X|Y)$  as the amount of uncertainty about  $X$  that can be reduced by observation  $Y$ , or equivalently, the amount of information  $Y$  can provide about  $X$ . Indeed, mutual information  $I(X; Y)$  plays the central role in Shannon’s random coding argument because the probability that independently drawn  $X^n$  and  $Y^n$  sequences “look” as if they were drawn jointly decays exponentially with exponent  $I(X; Y)$ . Shannon also proved a dual result [2] showing that the minimum compression rate  $R$  to satisfy a certain fidelity criterion  $D$  between the source  $X$  and its reconstruction  $\hat{X}$  is given by  $R(D) = \min_{p(\hat{x}|x)} I(X; \hat{X})$ . In another duality result (Lagrange duality this time) to (1), Gallager [3] proved the minimax redundancy theorem, connecting the redundancy of the universal lossless source code to the capacity of the channel with conditional distribution described by the set of possible source distributions.

Later on, it was shown that mutual information also has an important role in problems that are not necessarily related to describing sources or transferring information through channels. Perhaps the most lucrative example is the use of mutual information in gambling.

Kelly showed in [4] that if each horse race outcome can be represented as an independent and identically distributed (i.i.d.) copy of a random variable  $X$ , and the gambler has some side information  $Y$  relevant to the outcome  $X$  of the race, then under some conditions on the odds, the mutual information  $I(X; Y)$  captures the difference between the growth

rates of the optimal gambler’s wealth with and without side information  $Y$ . Thus, Kelly’s result gives an interpretation that mutual information  $I(X; Y)$  is the value of side information  $Y$  for the horse race  $X$ .

In order to tackle problems arising in information systems with causally dependent components, Massey [5] introduced the notion of directed information as

$$I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}),$$

and showed that the maximum directed information upper bounds the capacity of channels with feedback. Subsequently, it was shown that Massey’s directed information and its variants indeed characterize the capacity of feedback and two-way channels [6]–[13] and the rate distortion function with feedforward [14].

The main contribution of this paper is showing that directed information  $I(Y^n \rightarrow X^n)$  has a natural interpretation in gambling as the gain in growth rates due to causal side information. As a special case, if the horse race outcome and the corresponding side information sequences are i.i.d., then the (normalized) directed information becomes a single letter mutual information  $I(X; Y)$ , which coincides with Kelly’s result.

The rest of the paper is organized as follows. We describe the notation of directed information and causal conditioning in Section II. In Section III, we formulate the horse-race gambling problem, in which side information is revealed causally to the gambler. We present the main result in Section IV and an analytically solved example in Section V. Finally, Section VI concludes the paper and states two possible extensions of this work to stock market and data compression with causal side information.

## II. DIRECTED INFORMATION AND CAUSAL CONDITIONING

Throughout this paper, we use the *causal conditioning* notation  $(\cdot|\cdot)$  developed by Kramer [6]. We denote as  $p(x^n|y^{n-d})$  the probability mass function (pmf) of  $X^n = (X_1, \dots, X_n)$  causally conditioned on  $Y^{n-d}$ , for some integer  $d \geq 0$ , which is defined as

$$p(x^n|y^{n-d}) \triangleq \prod_{i=1}^n p(x_i|x^{i-1}, y^{i-d}).$$

(By convention, if  $i - d \leq 0$ , then  $x^{i-d}$  is set to null.) In particular, we use extensively the cases  $d = 0, 1$ :

$$p(x^n||y^n) \triangleq \prod_{i=1}^n p(x_i|x^{i-1}, y^i),$$

$$p(x^n||y^{n-1}) \triangleq \prod_{i=1}^n p(x_i|x^{i-1}, y^{i-1}).$$

Using the chain rule, we can easily verify that

$$p(x^n, y^n) = p(x^n||y^n)p(y^n||x^{n-1}).$$

The *causally conditional entropy*  $H(X^n||Y^n)$  is defined as

$$H(X^n||Y^n) \triangleq \mathbb{E}[\log p(X^n||Y^n)]$$

$$= \sum_{i=1}^n H(X_i|X^{i-1}, Y^i).$$

Under this notation, directed information can be written as

$$I(Y^n \rightarrow X^n) = \sum_{i=1}^n I(X_i; Y^i|X^{i-1})$$

$$= H(X^n) - H(X^n||Y^n),$$

which hints, in a rough analogy to mutual information, of a possible interpretation of directed information  $I(Y^n \rightarrow X^n)$  as the amount of information causally available side information  $Y^n$  can provide about  $X^n$ .

Note that the channel capacity results involve the term  $I(X^n \rightarrow Y^n)$ , which measures the information in the forward link  $X^n \rightarrow Y^n$ . In contrast, in gambling the gain in growth rate is due to the side information (backward link), and therefore the expression  $I(Y^n \rightarrow X^n)$  appears.

### III. GAMBLING IN HORSE RACES WITH CAUSAL SIDE INFORMATION

Suppose that there are  $m$  racing horses in an infinite sequence of horse races, and let  $X_i \in \mathcal{X} \triangleq \{1, 2, \dots, m\}$ ,  $i = 1, 2, \dots$ , denote the horse that wins at time  $i$ . Before betting in the  $i$ -th horse race, the gambler knows some side information  $Y_i \in \mathcal{Y}$ . We assume that the gambler invests all his capital in the horse race as a function of the information that he knows at time  $i$ , i.e., the previous horse race outcomes  $X^{i-1}$  and side information  $Y^i$  up to time  $i$ . Let  $b(x_i|x^{i-1}, y^i)$  be the proportion of wealth that the gambler bets on horse  $x_i$  given  $X^{i-1} = x^{i-1}$  and  $Y^i = y^i$ . The betting scheme should satisfy  $b(x_i|x^{i-1}, y^i) \geq 0$  (no short) and  $\sum_{x_i} b(x_i|x^{i-1}, y^i) = 1$  for any history  $x^{i-1}, y^i$ . Let  $o(x_i|x^{i-1})$  denote the odds of a horse  $x_i$  given the previous outcomes  $x^{i-1}$ , which is the amount of capital that the gambler gets for each unit capital invested in the horse. We denote by  $S(x^n||y^n)$  the gambler's wealth after  $n$  races where the race outcomes were  $x^n$  and the side information that was causally available was  $y^n$ . The *growth*, denoted by  $W(X^n||Y^n)$ , is defined as the expected logarithm (base 2) of the gambler's wealth, i.e.,

$$W(X^n||Y^n) \triangleq \mathbb{E}[\log S(X^n||Y^n)]. \quad (2)$$

Finally, the *growth rate*  $\frac{1}{n}W(X^n||Y^n)$  is defined as the normalized growth.

Here is a summary of the notation:

- $X_i$  is the outcome of the horse race at time  $i$ .
- $Y_i$  is the the side information at time  $i$ .
- $o(X_i|X^{i-1})$  is the payoffs at time  $i$  for horse  $X_i$  given that in the previous race the horses  $X^{i-1}$  won.
- $b(X_i|Y^i, X^{i-1})$  the fractions of the gambler's wealth invested in horse  $X_i$  at time  $i$  given that the outcome of the previous races are  $X^{i-1}$  and the side information available at time  $i$  is  $Y^i$ .
- $S(X^n||Y^n)$  the gambler's wealth after  $n$  races when the outcomes of the races are  $X^n$  and the side information  $Y^n$  is causally available.
- $\frac{1}{n}W(X^n||Y^n)$  is the growth rate.

Without loss of generality, we assume that the gambler's capital is 1 initially; therefore  $S_0 = 1$ .

### IV. MAIN RESULTS

In Subsection IV-A, we assume that the gambler invests all his money in the horse race, while in Subsection IV-B, we allow the gambler to invest only part of the money. Using Kelly's result, it is shown in Subsection IV-B that if the odds are fair with respect to some distribution, then the gambler should invest all his money in the race.

#### A. Investing all the money in the horse race

We assume that at any time  $n$  the gambler invests all his capital and, therefore,

$$S(X^n||Y^n) = b(X_n|X^{n-1}, Y^n)o(X_n|X^{n-1})S(X^{n-1}||Y^{n-1}).$$

This also implies that

$$S(X^n||Y^n) = \prod_{i=1}^n b(X_i|X^{i-1}, Y^i)o(X_i|X^{i-1}).$$

The following proposition characterizes the optimal betting strategy and the corresponding growth of wealth.

*Theorem 1:* For any finite horizon  $n$ , the maximum growth rate is achieved when the gambler invests the money proportional to the causal conditioning distribution, i.e.,

$$b^*(x_i|x^{i-1}, y^i) = p(x_i|x^{i-1}, y^i), \quad \forall x^i, y^i, i \leq n, \quad (3)$$

and the growth is

$$W^*(X^n||Y^n) = \mathbb{E}[\log o(X^n)] - H(X^n||Y^n).$$

Note that the sequence  $\{p(x_i|x^{i-1}, y^i)\}_{i=1}^n$  uniquely determines  $p(x^n||y^n)$ . Moreover, for all pairs  $(x^n, y^n)$  such that  $p(x^n||y^n) > 0$ , the sequence  $\{p(x_i|x^{i-1}, y^i)\}_{i=1}^n$  is determined uniquely by  $p(x^n||y^n)$  by the identity

$$p(x_i|x^{i-1}, y^i) = \frac{p(x^i||y^i)}{p(x^{i-1}||y^{i-1})}.$$

A similar argument applies for  $\{b^*(x_i|x^{i-1}, y^i)\}_{i=1}^n$  and  $b^*(x^n||y^n)$ , and therefore (3) is equivalent to

$$b^*(x^n||y^n) = p(x^n||y^n), \quad \forall x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n.$$

*Proof of Theorem 1:* We have

$$\begin{aligned} W^*(X^n|Y^n) &= \max_{b(x^n|y^n)} \mathbb{E}[\log b(X^n|Y^n)o(X^n)] \\ &= \max_{b(x^n|y^n)} \mathbb{E}[\log b(X^n|Y^n)] + \mathbb{E}[\log o(X^n)] \\ &= -H(X^n|Y^n) + \mathbb{E}[\log o(X^n)], \end{aligned}$$

where the last equality is achieved by choosing  $b(x^n|y^n) = p(x^n|y^n)$ , and it is justified by the following upper bound

$$\begin{aligned} &\mathbb{E}[\log b(X^n|Y^n)] \\ &= \sum_{x^n, y^n} p(x^n, y^n) \left[ \log p(x^n|y^n) + \log \frac{b(x^n|y^n)}{p(x^n|y^n)} \right] \\ &= -H(X^n|Y^n) + \sum_{x^n, y^n} p(x^n, y^n) \log \frac{b(x^n|y^n)}{p(x^n|y^n)} \\ &\stackrel{(a)}{\leq} -H(X^n|Y^n) + \log \sum_{x^n, y^n} p(x^n, y^n) \frac{b(x^n|y^n)}{p(x^n|y^n)} \\ &\stackrel{(b)}{\leq} -H(X^n|Y^n) + \log \sum_{x^n, y^n} p(y^n|x^{n-1})b(x^n|y^n) \\ &= -H(X^n|Y^n), \end{aligned} \quad (4)$$

where (a) follows from Jensen's inequality and (b) from the fact that  $\sum_{x^n, y^n} p(y^n|x^{n-1})b(x^n|y^n) = 1$ . All summations in (4) are over the arguments  $(x^n, y^n)$ , for which  $p(x^n, y^n) > 0$ . This ensures that  $p(x^n|y^n) > 0$ , and therefore, we can multiply and divide by  $p(x^n|y^n)$  in the first step of (4). ■

In a case in which the odds are fair and uniform, i.e.,  $o(X_i|X^{i-1}) = \frac{1}{|\mathcal{X}|}$ , then

$$\frac{1}{n}W^*(X^n|Y^n) = \log |\mathcal{X}| - \frac{1}{n}H(X^n|Y^n).$$

Thus, the sum of the growth rate  $\frac{1}{n}W(X^n|Y^n)$  and the entropy rate  $\frac{1}{n}H(X^n|Y^n)$  of the horse race process conditioned causally on the side information is constant, and one can see a duality between  $H(X^n|Y^n)$  and  $W^*(X^n|Y^n)$ ; cf. [15, th. 6.1.3]

Let us denote by  $\Delta W$  the increase in the growth rate due to causal side information, i.e.,

$$\Delta W = \frac{1}{n}W^*(X^n|Y^n) - \frac{1}{n}W^*(X^n). \quad (5)$$

Thus  $\Delta W$  characterizes the value of side information  $Y^n$ . Theorem 1 leads to the following proposition, which gives a new operational meaning of Massey's directed information.

*Corollary 1:* The increase in growth rate due to causal side information  $Y^n$  for horse races  $X^n$  is

$$\Delta W = \frac{1}{n}I(Y^n \rightarrow X^n). \quad (6)$$

*Proof:* From Theorem 1, we have

$$\begin{aligned} W^*(X^n|Y^n) - W^*(X^n) &= -H(X^n|Y^n) + H(X^n) \\ &= I(Y^n \rightarrow X^n). \quad \blacksquare \end{aligned}$$

## B. Investing only part of the money

In this subsection, we consider the case wherein the gambler does not necessarily invest all his money in the gambling. Let  $b_0(y^i, x^{i-1})$  be the portion of money that the gambler does not invest in gambling at time  $i$  given that the previous race results were  $x^{i-1}$  and the side information is  $y^i$ . In this setting, the wealth is given by

$$\begin{aligned} S(X^n|Y^n) &= \prod_{i=1}^n (b_0(X^{i-1}, Y^i) + (b(X_i|X^{i-1}, Y^i)o(X_i|X^{i-1}))), \end{aligned}$$

and the growth  $W(X^n|Y^n)$  is defined as before in (2).

The term  $W(X^n|Y^n)$  obeys a chain rule similar to the causal conditioning entropy definition  $H(X^n|Y^n)$ , i.e.,

$$W(X^n|Y^n) = \sum_{i=1}^n W(X_i|X^{i-1}, Y^i),$$

where

$$\begin{aligned} W(X_i|X^{i-1}, Y^i) &\triangleq \mathbb{E}[\log(b_0(X^{i-1}, Y^i) + b(X_i|X^{i-1}, Y^i)o(X_i|X^{i-1}))]. \end{aligned}$$

Note that for any given history  $(x^{i-1}, y^i) \in \mathcal{X}^{i-1} \times \mathcal{Y}^i$ , the betting scheme  $\{b_0(x^{i-1}, y^i), b(x_i|x^{i-1}, y^i)\}$  influences only  $W(X_i|X^{i-1}, Y^i)$ , so that we have

$$\begin{aligned} &\max_{\{b_0(x^{i-1}, y^i), b(x_i|x^{i-1}, y^i)\}_{i=1}^n} W(X^n|Y^n) \\ &= \sum_{i=1}^n \max_{b_0(x^{i-1}, y^i), b(x_i|x^{i-1}, y^i)} W(X_i|X^{i-1}, Y^i) \\ &= \sum_{i=1}^n \sum_{x^{i-1}, y^i} p(x^{i-1}, y^i) \max_{b_0(x^{i-1}, y^i), b(x_i|x^{i-1}, y^i)} W(X_i|x^{i-1}, y^i). \end{aligned}$$

The optimization problem in the last equation is equivalent to the problem of finding the optimal betting strategy in the memoryless case where the winning horse distribution  $p(x)$  is  $p(x) = \Pr(X_i = x|x^{i-1}, y^i)$ , the odds  $o(x)$  are  $o(x) = o(X_i = x|x^{i-1})$ , and the betting strategy  $(b_0, b(x))$  is respectively  $(b_0(x^i, y^{i-1}), b(X_i = x|x^{i-1}, y^i))$ . Hence, the optimization,  $\max W(X_i|x^{i-1}, y^i)$ , is equivalent to the following convex problem:

$$\begin{aligned} &\text{maximize} && \sum_x p(x) \log(b_0 + b(x)o(x)) \\ &\text{subject to} && b_0 + \sum_x b(x) = 1, \\ &&& b_0 \geq 0, b(x) \geq 0, \quad \forall x \in \mathcal{X}. \end{aligned}$$

The solution to this optimization problem was given by Kelly [4]. If the odds are *super-fair*, namely,  $\sum_x \frac{1}{o(x)} \leq 1$ , then the gambler will invest all his wealth in the race rather than leave some as cash, since by betting  $b(x) = \frac{c}{o(x)}$ , where  $c = 1/\sum_x \frac{1}{o(x)}$ , the gambler's money will be multiplied by  $c \geq 1$ , regardless of the race outcome. Therefore, for this case, the solution is given by Theorem 1, where the gambler invests proportional to the causal conditioning distribution  $p(x^n|y^n)$ .

If the odds are sub-fair, i.e.,  $\sum_x \frac{1}{o(x)} > 1$ , then it is optimal to bet only some of the money, namely  $b_0 > 0$ . The solution to this problem is given in terms of an algorithm in [4, p. 925].

## V. AN EXAMPLE

Here we consider betting on a horse race, where the winning horse can be represented as a Markov process, and causal side information is available.

*Example 1:* Consider the horse race process depicted in Figure 1, where two horses are racing and the winning horse  $X_i$  behaves as a Markov process. A horse that won will win again with probability  $1 - p$  and lose with probability  $p$ . At time zero, we assume that both horses have probability  $\frac{1}{2}$  of winning. The side information  $Y_i$  at time  $i$  is a noisy observation of the horse race outcome  $X_i$ . It has probability  $1 - q$  of being equal to  $X_i$ , and probability  $q$  of being different from  $X_i$ .

For this example, the increase in growth rate due to side information as  $n$  goes to infinity is

$$\Delta W = h(p * q) - h(q),$$

where the function  $h(\cdot)$  denotes the binary entropy, i.e.,  $h(x) = -x \log x - (1 - x) \log(1 - x)$ , and  $p * q$  denotes the parameter of a Bernoulli distribution that results from convolving two Bernoulli distributions with parameters  $p$  and  $q$ , i.e.,  $p * q = (1 - p)q + (1 - q)p$ .

The increase in the growth rate  $\Delta W$  for this example can be obtained using first principles as follows:

$$\begin{aligned} \Delta W &= \lim_{n \rightarrow \infty} \frac{1}{n} I(Y^n \rightarrow X^n) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(Y^i | X^{i-1}) - H(Y^i | X^i) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [H(Y^i | X_{i-1}) - H(Y_2^i | X_2^i) - H(Y_1 | X_1)] \\ &\stackrel{(a)}{=} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [H(Y^i | X_{i-1}) - H(Y^{i-1} | X^{i-1}) - H(Y_1 | X_1)] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [H(Y_i | Y^{i-1}, X_{i-1}) - H(Y_1 | X_1)] \\ &\stackrel{(b)}{=} H(Y_1 | X_0) - H(Y_1 | X_1) = h(p * q) - h(q), \end{aligned} \quad (7)$$

where steps (a) and (b) are due to the stationarity of the process  $(X_i, Y_i)$ . Alternatively, the sequence of equalities up to step (b) in (7) can be derived directly using

$$\begin{aligned} \frac{1}{n} I(Y^n \rightarrow X^n) &\stackrel{(a)}{=} \frac{1}{n} \sum_{i=1}^n I(Y_i; X_i^n | X^{i-1}, Y^{i-1}) \\ &\stackrel{(b)}{=} H(Y_1 | X_0) - H(Y_1 | X_1), \end{aligned} \quad (8)$$

where (a) is the identity given in [11, eq. (9)] and (b) is due to the stationarity of the process.

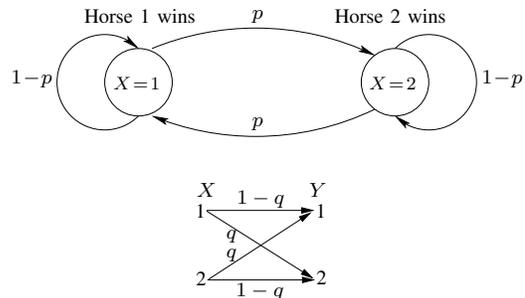


Fig. 1. The setting of Example 1. The winning horse  $X_i$  is represented as a Markov process with two states. In state 1, horse number 1 wins, and in state 2, horse number 2 wins. The side information,  $Y_i$ , is a noisy observation of the winning horse,  $X_i$ .

If the side information is known with some lookahead  $k \in \{0, 1, \dots\}$ , that is, if the gambler knows  $Y^{i+k}$  at time  $i$ , then the increase in growth rate is given by

$$\begin{aligned} \Delta W &= \lim_{n \rightarrow \infty} \frac{1}{n} I(Y^{n+k} \rightarrow X^n) \\ &= H(Y_{k+1} | Y^k, X_0) - H(Y_1 | X_1), \end{aligned} \quad (9)$$

where the last equality is due to the same arguments as in (8).

Figure 2 shows the increase in growth rate  $\Delta W$  due to side information as a function of the side information parameters  $(q, k)$ . The left plot shows  $\Delta W$  as a function of  $q$ , where  $p = 0.2$  and no lookahead,  $k = 0$ . The right plot shows  $\Delta W$  as a function of  $k$ , where  $p = 0.2$  and  $q = 0.25$ . If the entire side information sequence  $Y_1, Y_2, \dots$  is known to the gambler ahead of time, then we should have mutual information rather than directed information, i.e.,

$$\begin{aligned} \Delta W &= \lim_{n \rightarrow \infty} \frac{1}{n} I(Y^n; X^n) \\ &= \lim_{n \rightarrow \infty} \frac{H(Y^n)}{n} - H(Y_1 | X_1), \end{aligned} \quad (10)$$

and this coincides with the fact that for a stationary hidden Markov process  $\{Y_1, Y_2, \dots\}$  the sequence  $H(Y_{k+1} | Y^{k-1}, X_0)$  converges to the entropy rate of the process.

## VI. CONCLUSION AND FURTHER EXTENSIONS

We have shown that directed information arises naturally in gambling as the gain in the maximum achievable capital growth due to the availability of causal side information. We now outline two extensions: stock market portfolio strategies and data compression in the presence of causal side information. Details are given in [16].

### A. Stock market

Using a notation similar to that in [15, ch. 16], a stock market at time  $i$  is represented as a vector of stocks  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{im})$ , where  $m$  is the number of stocks, and the price relative  $X_{ik}$  is the ratio of the price of stock- $k$  at the end of day  $i$  to the price of stock- $k$  at the beginning of day  $i$ . We assume that at time  $i$  there is side information  $Y^i$  that is known to the investor. A *portfolio* is an allocation of wealth across the stocks. A nonparticipating or causal portfolio strategy with

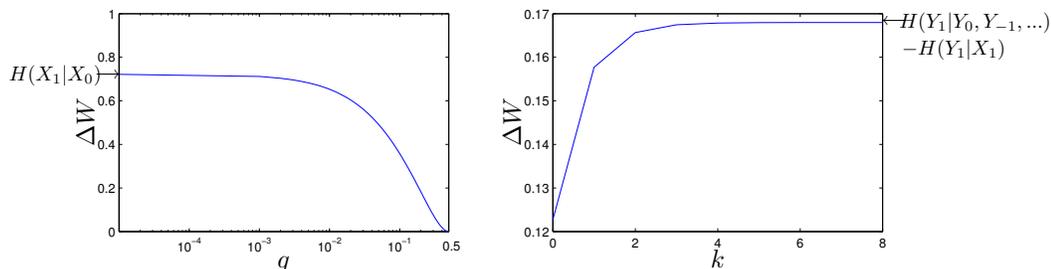


Fig. 2. Increase in the growth rate, in Example 1, as a function of the side information parameters  $(q, k)$ . The left plot of the figure shows the increase of the growth rate  $\Delta W$  as a function of  $q = \Pr(X_i \neq Y_i)$  and there is no lookahead. The right plot shows the increase of the growth rate as function of lookahead  $k$ , where  $q = 0.25$ . The horse race outcome is assumed to be a first-order binary symmetric Markov process with parameter  $p = 0.2$ .

causal side information at time  $i$  is denoted as  $\mathbf{b}(\mathbf{x}^{i-1}, y^i)$ , and it satisfies  $\sum_{k=1}^m b_k(\mathbf{x}^{i-1}, y^i) = 1$ , and  $b_k(\mathbf{X}^{i-1}, Y^i) \geq 0$  for all possible  $\mathbf{x}^{i-1}, y^i$ . We define  $S(\mathbf{x}^n || y^n)$  as the wealth at the end of day  $n$  for a stock sequence  $\mathbf{x}^n$  and causal side information  $y^n$ . We can write

$$S(\mathbf{x}^n || y^n) = (\mathbf{b}^t(\mathbf{x}^{n-1}, y^n) \mathbf{x}_n) S(\mathbf{x}^{n-1} || y^{n-1})$$

where  $(\cdot)^t$  denotes the transpose of a vector. The goal is to maximize the growth  $W(\mathbf{X}^n || Y^n) = \mathbb{E}[\log S(\mathbf{X}^n || Y^n)]$ . We also define  $W(\mathbf{X}_n | \mathbf{X}^{n-1}, Y^n) = \mathbb{E}[\log(\mathbf{b}^t(\mathbf{X}^{n-1}, Y^n) \mathbf{X}_n)]$ . From this definition, we can write the chain rule

$$W(\mathbf{X}^n || Y^n) = \sum_{i=1}^n W(\mathbf{X}_i | \mathbf{X}^{i-1}, Y^i).$$

The gambling in horse races with  $m$  horses studied in the previous section is a special case of investing the stock market with  $m + 1$  stocks. The first  $m$  stocks correspond to the  $m$  horses and at the end of the day one of the stocks, say  $k \in \{1, \dots, m\}$ , gets the value  $o(k)$  with probability  $p(k)$  and all other stocks become zero. The  $m + 1$ -st stock is always one, and it allows the gambler to invest only part of the wealth in the horse race.

The developments in the previous section can be expanded to characterize the increase in growth rate due to side information, where again directed information emerges as the key quantity, upper-bounding the value of causal side information; cf. [17]. Details will be given in [16].

### B. Instantaneous compression with causal side information

Let  $X_1, X_2, \dots$  be a source and  $Y_1, Y_2, \dots$  its side information sequence. The source is to be losslessly encoded instantaneously, with causal available side information. More precisely, an instantaneous lossless source encoder with causal side information consists of a sequence of mappings  $\{M_i\}_{i \geq 1}$  such that each  $M_i : \mathcal{X}^i \times \mathcal{Y}^i \rightarrow \{0, 1\}^*$  has the property that for every  $x^{i-1}$  and  $y^i$ ,  $M_i(x^{i-1}, y^i)$  is an instantaneous (prefix) code for  $X_i$ .

An instantaneous lossless source encoder with causal side information operates sequentially, emitting the concatenated bit stream  $M_1(X_1, Y_1)M_2(X_2, Y_2) \dots$ . The defining property that  $M_i(x^{i-1}, y^i)$  is an instantaneous code for every  $x^{i-1}$  and  $y^i$  is a necessary and sufficient condition for the existence of

a decoder that can losslessly recover  $x^i$  based on  $y^i$  and the bit stream  $M_1(x_1, y_1)M_2(x_2, y_2) \dots$  just as soon as it sees  $M_1(x_1, y_1)M_2(x_2, y_2) \dots M_i(x^i, y^i)$  for all sequence pairs  $(x_1, y_1), (x_2, y_2) \dots$  and all  $i \geq 1$ . Using natural extensions of standard arguments, we show in [16] that  $I(Y^n \rightarrow X^n)$  is essentially (up to terms that are sublinear in  $n$ ) the rate savings in optimal sequential lossless compression of  $X^n$  due to the causal availability of the side information.

### REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
- [2] —, "Coding theorems for a discrete source with fidelity criterion," in *Information and Decision Processes*, R. E. Machol, Ed. McGraw-Hill, 1960, pp. 93–126.
- [3] R. G. Gallager, "Source coding with side information and universal coding," Sept. 1976, unpublished manuscript.
- [4] J. L. Kelly, "A new interpretation of information rate," *Bell System Technical Journal*, vol. 35, pp. 917–926, 1956.
- [5] J. Massey, "Causality, feedback and directed information," *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, pp. 303–305, 1990.
- [6] G. Kramer, "Directed information for channels with feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH) Zurich, 1998.
- [7] S. Tatikonda, "Control under communication constraints," *Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA*, 2000.
- [8] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 4–21, 2003.
- [9] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," Sept. 2006, submitted to *IEEE Trans. Inf. Theory*. Available at [arxiv.org/pdf/cs.IT/0608070](http://arxiv.org/pdf/cs.IT/0608070).
- [10] S. C. Tatikonda and S. Mitter, "The capacity of channels with feedback," September 2006, submitted to *IEEE Trans. Inf. Theory*. Available at [arxiv.org/cs.IT/0609139](http://arxiv.org/cs.IT/0609139).
- [11] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," Jan 2007, submitted to *IEEE Trans. Inf. Theory*. Available at [arxiv.org/cs.IT/0701041](http://arxiv.org/cs.IT/0701041).
- [12] H. H. Permuter and T. Weissman, "Capacity region of the finite-state multiple access channel with and without feedback," August 2007, submitted to *IEEE Trans. Inf. Theory*. Available at [arxiv.org/pdf/cs.IT/0608070](http://arxiv.org/pdf/cs.IT/0608070).
- [13] B. Shradar and H. H. Permuter, "On the compound finite state channel with feedback," in *Proc. Internat. Symp. Inf. Theory*, Nice, France, 2007.
- [14] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inf. Theory*, vol. IT-53, pp. 2154–2179, 2007.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York: Wiley, 2006.
- [16] Y.-H. Kim, H. H. Permuter, and T. Weissman, "An interpretation of directed information in gambling, portfolio theory and data compression," Jan. 2007, in preparation.
- [17] A. R. Barron and T. M. Cover, "A bound on the financial value of information," *IEEE Trans. Inf. Theory*, vol. IT-34, pp. 1097–1100, 1988.