The Efficiency of Investment Information

Elza Erkip, *Member, IEEE*, and Thomas M. Cover, *Fellow, IEEE*

***Abstract*— We investigate how the description of a correlated information V improves the investment in the stock market $\boldsymbol{X}$. The objective is to maximize the growth rate of wealth in repeated investments. We find a single-letter characterization of the incremental growth rate $\Delta(R)$, the maximum increase in growth rate when V is described to the investor at rate R. The incremental growth rate specialized to the horse race market is related to source coding with side information of Wyner and Ahlswede–Körner. We provide two horse race examples: jointly binary and jointly Gaussian. The initial efficiency $\Delta'(0)$ is the maximum possible increase in the growth rate per bit of description. We show that the initial efficiency is related to the dependency between V and the market. In particular, for the horse race market, the initial efficiency is the square of the Hirschfeld–Gebelein–Rényi maximal correlation between V and $\boldsymbol{X}$. This provides a connection with the hypercontraction of the Markov operator of Ahlswede and Gács. For the general market the initial efficiency is 1 when the side information V is equal to the stock market outcome $\boldsymbol{X}$.**

***Index Terms*— Investment, maximal correlation, portfolio, source coding with side information.**

I. Introduction

SUPPOSE an investor wishes to invest in the stock market. The market consists of m stocks. We denote these m stocks by the vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_m)$, where $X_j \geq 0$ is the price relative for the stock j on that day. We assume $\boldsymbol{X}$ is random with a known underlying distribution $F(\boldsymbol{x})$. The investor divides his money among these m stocks according to the portfolio $\boldsymbol{b} = (b_1, b_2, \cdots, b_m)$, where $b_j \geq 0$

$$\sum_{j=1}^{m} b_j = 1.$$

The wealth of the investor after one day is given by $S = \boldsymbol{b}^t \boldsymbol{X}$.

Now suppose the investor invests in the stock market for n consecutive days. We assume the investor uses the same betting strategy $\boldsymbol{b}$ each day. Then the wealth at the end of day n is given by

$$S_n = \prod_{i=1}^{n} \boldsymbol{b}^t \boldsymbol{X}_i.$$

Manuscript received September 10, 1996; revised December 1, 1997. This work was supported in part by NSF under Grant NCR-9205663, by JSEP under Contract DAAH04-94-G-0058, and by ARPA under Contract J-FBI-94-218-2. The material in this paper was presented in part at the IEEE Information Theory Workshop, Rydzyna, Poland, 1995, and at the IEEE International Symposium on Information Theory, Whistler, BC, Canada, 1995.

E. Erkip is with the Department of Electrical and Computer Engineering, MS-366, Rice University, Houston, TX 77005-1892 USA.

T. M. Cover is with the Departments of Statistics and Electrical Engineering, Stanford University, Stanford, CA 94305-4055 USA.

Publisher Item Identifier S 0018-9448(98)02355-4.

If $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots$ are independent and identically distributed according to F, then

$$\frac{1}{n} \log S_n \rightarrow W(\boldsymbol{b}, F) \quad \text{a.e.}$$

where

$$W(\boldsymbol{b}, F) = \int \log \boldsymbol{b}^t \boldsymbol{x} \, dF(\boldsymbol{x})$$

is the *growth rate* of wealth. Unless otherwise stated, we will use logarithm to the base 2. Thus (see [12])

$$S_n = 2^{nW(\boldsymbol{b}, F) + o(\sqrt{n})}.$$

The optimum growth rate $W^\star(F)$ is the maximum of $W(\boldsymbol{b}, F)$ over all choices of portfolio $\boldsymbol{b}$. The $\log$ optimum portfolio that achieves this maximum is denoted by $\boldsymbol{b}^\star$. Therefore,

$$W^\star(F) = \max_{\boldsymbol{b}} W(\boldsymbol{b}, F) = W(\boldsymbol{b}^\star F),$$

and

$$\boldsymbol{b}^\star = \arg\max_{\boldsymbol{b}} W(\boldsymbol{b}, F). \tag{1}$$

Now suppose that the investor has some side information V about the stock market. For example, this may be some insider information about some of the companies whose stocks are being traded in the market. Then the investor can improve his growth rate due to the presence of this side information.

We assume the sequence (X_i, V_i), $i = 1, \cdots, n$, is independent and identically distributed with a given distribution $F(\boldsymbol{x}, v)$. Then the increase in growth rate due to the side information is given by

$$\begin{aligned} \Delta W &= E_V W^\star(F(\cdot \mid V)) - W^\star(F(\cdot)) \\ &= \max_{\boldsymbol{b}(\cdot)} E \log \frac{\boldsymbol{b}^t(V)\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} \end{aligned}$$

where the maximum is over all portfolio choices $\boldsymbol{b} : \mathcal{V} \rightarrow \mathcal{B}$. Here $\mathcal{V}$ denotes the alphabet for V and $\mathcal{B}$ is the $(m-1)$-dimensional probability simplex

$$\mathcal{B} = \left\{ \boldsymbol{b} \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^{m} b_i = 1 \right\}.$$

The idea of maximizing growth rate of wealth in the horse race market was introduced by Kelly [16]. In the general market, no closed form for the optimum portfolio $\boldsymbol{b}^\star$ or the optimum growth rate $W^\star$ exists. Barron and Cover [6] showed that, in the presence of side information V, the increase in growth rate ΔW is bounded above by $I(X; V)$ where the growth rate and mutual information are both defined with logarithms of the same base.

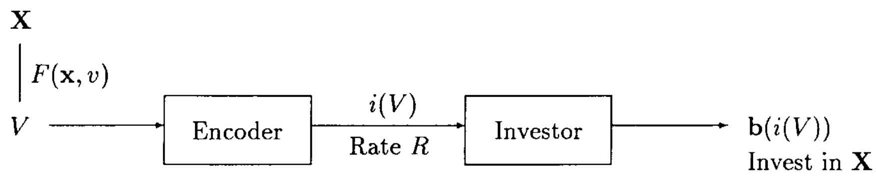

Fig. 1. Descriptions and investment.

In this work, we consider a slightly different situation. Now the investor cannot observe the side information V directly, but can access a rate R description of it. We would like to know what R bits should be given to the investor and what this information is worth in terms of increasing the growth rate. What is the essential part of V, such that its knowledge will be sufficient for investment in $\boldsymbol{X}$, and by how much can this help? We do not wish to reconstruct V from the encoded data; we only use the rate R description to provide a large increase in growth rate for investment in $\boldsymbol{X}$. We wish to find the incremental growth rate $\Delta(R)$, the maximum increase in the growth rate for rate R descriptions.

We will see that this problem can be formulated as a rate distortion problem, and a single-letter characterization of $\Delta(R)$ can be found as given in Theorem 1. Our main result is that if there is a rate constraint R in describing side information V, the maximum increase in the growth rate is given by

$$\Delta(R) = \max_{F(\boldsymbol{b}|v):\ I(\boldsymbol{b};V)\le R,\ \boldsymbol{b}\to V\to \boldsymbol{X}} E\log\frac{\boldsymbol{b}^t\boldsymbol{X}}{\boldsymbol{b}^{*t}\boldsymbol{X}}.$$

We then specialize the general $\Delta(R)$ to the horse race market. We observe that finding the incremental growth rate for the horse race market can be reduced to source coding with side information of Wyner [22] and Ahlswede–Körner [5]. We solve for the incremental growth rate for jointly Gaussian and jointly binary horse race markets. The jointly Gaussian source coding with side information has not been previously treated, since noiseless source coding is unmotivated for continuous random variables.

After observing that the incremental growth rate is a concave and nondecreasing function of R, we turn our attention to the derivative of $\Delta(R)$ at the origin. We call $\Delta'(0)$ the *initial efficiency*. Initial efficiency shows how the first few bits about V improve the growth rate in the investment in the stock market $\boldsymbol{X}$.

In the horse race market, we show that the initial efficiency is the square of the Hirschfeld–Gebelein–Rényi maximal correlation between V and X. Maximal correlation, introduced by Hirschfeld [15] for discrete random variables and by Gebelein [14] for absolutely continuous distributions, is a generalized version of the correlation coefficient between two random variables. Rényi [17], [18] showed that it is a natural measure of dependence, investigated conditions under which maximal correlation is attained and provided an alternate characterization. Maximal correlation has appeared numerous times before in the information theory literature. Of these, one paper to note is by Witsenhausen [20], where a meaningful definition for common information between two random variables involves the maximal correlation. Another work of considerable interest is by Ahlswede and Gács [2] on the hypercontraction of the Markov operator. The relationship of their results to the initial efficiency in the horse race market is presented in Section IV-E. We show in Section V that for the general market the initial efficiency is 1 when the side information V is equal to the stock vector $\boldsymbol{X}$.

In the next section we formulate the problem and establish the incremental growth rate for the general market. Section III is devoted to a simple example where the incremental growth rate and the initial efficiency are calculated. In Section IV we define the horse race market, specialize the incremental growth rate to the horse race market, and solve two fundamental horse race examples: jointly binary and jointly Gaussian. We next find that the initial efficiency in the horse race market is the square of the Hirschfeld–Gebelein–Rényi maximal correlation. Section V is about the initial efficiency for the general market when $V = \boldsymbol{X}$.

II. Formulation of the Problem

Consider a stock market consisting of m investment opportunities. We call these investment opportunities *stocks* or *assets*. Let X_j denote the ratio of closing price to the opening price of stock j on a given day. Here $X_j \ge 0$ is the price relative for stock j. We denote the stock market by the vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_m)$. We assume that $\boldsymbol{X}$ has a known distribution. The investor distributes his money among the m stocks according to the portfolio

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_m) \qquad b_j \ge 0 \qquad \sum_{j=1}^{m} b_j = 1$$

where b_j denotes the proportion of money invested in stock i.

Suppose that some other person, whom we call the encoder, has access to a piece of relevant information about the stock market. We denote this side information by V. The investor cannot observe V directly, but both the investor and the encoder know the joint distribution $F(\boldsymbol{x}, v)$ of $\boldsymbol{X}$ and V. The task of the encoder is to provide the investor with an R-bit description of V, which in turn is used by the investor to improve the growth rate. Fig. 1 illustrates this setup.

We assume that the investment is done for n consecutive days and that $(\boldsymbol{X}_i, V_i)$ are i.i.d. according to $F(\boldsymbol{x}, v)$. The encoder observes V^n and chooses an index from the set $\{1, 2, \cdots, 2^{nR}\}$ to describe V^n. The investor uses this compressed data to devise a sequence of portfolios $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ to invest in the stock market for n consecutive days. We would like to know how much this description improves the growth rate of the investor and what the best such description should be. The improvement is with respect to the optimal growth rate for distribution $F(\boldsymbol{x})$.

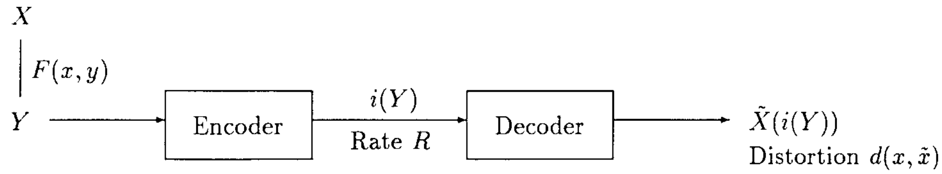

Fig. 2. Rate distortion with remote source.

Definition: A $(2^{nR}, n)$ code $\mathcal{C}$ consists of an encoding function $i_n : \mathcal{V}^n \rightarrow \{1, 2, \cdots, 2^{nR}\}$ and an investment strategy $\boldsymbol{b}^n : \{1, 2, \cdots, 2^{nR}\} \rightarrow \mathcal{B}^n$ where $\mathcal{B}$ denotes the $(m-1)$-dimensional probability simplex. The *rate* of the code is R. The *increase in growth rate* Δ_n associated with the code $\mathcal{C}$ is given by

$$\Delta_n = \frac{1}{n} E \log \frac{\prod_{j=1}^{n} \boldsymbol{b}_j^t(i_n(V^n))\boldsymbol{X}_j}{\prod_{j=1}^{n} \boldsymbol{b}^{\star t}\boldsymbol{X}_j}.$$

The expectation is taken over the joint distribution of V^n and $\boldsymbol{X}^n$, and $\boldsymbol{b}^\star$ is the log optimum portfolio with respect to the marginal distribution $F(\boldsymbol{x})$ of $\boldsymbol{X}$ as in (1).

Definition: A description-rate growth-rate pair (R, Δ) is said to be *achievable* if there exists a sequence of $(2^{nR}, n)$ codes $(i_n(\cdot), \boldsymbol{b}^n(\cdot))$ for which

$$\lim_{n\rightarrow\infty} \frac{1}{n} E \log \frac{\prod_{j=1}^{n} \boldsymbol{b}_j^t(i_n(V^n))\boldsymbol{X}_j}{\prod_{j=1}^{n} \boldsymbol{b}^{\star t}\boldsymbol{X}_j} \geq \Delta.$$

Definition: The *description-growth rate region* is the closure of the set of achievable description-rate growth-rate pairs (R, Δ).

Definition: The *incremental growth rate* $\Delta(R)$ is the supremum of all Δ such that (R, Δ) is in the description-growth rate region for a given R.

Note that $\Delta(R)$ can be thought of as the maximum increase in growth rate for transmission rate R. The next theorem is a characterization of the incremental growth rate for the stock market.

Theorem 1: For investment in stock market $\boldsymbol{X}$ with side information V described at rate R, the incremental growth rate is given by

$$\Delta(R) = \max E \log \frac{\boldsymbol{b}^t\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}$$

where the maximum is over all conditional distributions $F(\boldsymbol{b} \mid v)$ satisfying the rate constraint $I(\boldsymbol{b}; V) \leq R$, and $\boldsymbol{b} \rightarrow V \rightarrow \boldsymbol{X}$ forms a Markov chain.

In order to see the intuition behind Theorem 1, we consider a specific scheme for achieving $\Delta(R)$. We fix a distribution $F(\boldsymbol{b} \mid v)$ satisfying the Markov relationship and the rate constraint in Theorem 1. We denote the marginal distribution of $\boldsymbol{b}$ by $F(\boldsymbol{b})$. A random codebook consisting of 2^{nR} sequences $\boldsymbol{b}^n$ drawn i.i.d. $\sim \prod_{i=1}^n F(\boldsymbol{b}_i)$ is generated. We denote these sequences by $\boldsymbol{b}^n(1), \cdots, \boldsymbol{b}^n(2^{nR})$. The encoder observes the side information V^n and finds an index w such that $\boldsymbol{b}^n(w)$ is jointly typical with V^n. Since $R \geq I(\boldsymbol{b}; V)$, the existence of such a w is guaranteed with high probability. The encoder sends the index w to the investor and the investor uses the portfolio choice $\boldsymbol{b}^n(w) = (\boldsymbol{b}_1(w), \cdots, \boldsymbol{b}_n(w))$. The Markov relationship $\boldsymbol{b} \rightarrow V \rightarrow \boldsymbol{X}$ ensures that $\boldsymbol{X}^n$ and $\boldsymbol{b}^n(w)$ are jointly typical. Thus it appears that $(\boldsymbol{X}, \boldsymbol{b}) \sim F(\boldsymbol{x}, \boldsymbol{b})$. Therefore, an increase in growth rate of

$$\Delta = E \log \frac{\boldsymbol{b}^t\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}$$

can be achieved. Maximizing over the choice of the random portfolio $\boldsymbol{b}$, we observe that $\Delta(R)$ in Theorem 1 is achievable.

We will prove Theorem 1 by showing that this problem can be viewed as a rate distortion problem with a remote source. We first define the distortion function $d(\boldsymbol{b}, \boldsymbol{X})$.

Definition: The *distortion function* $d(\boldsymbol{b}, \boldsymbol{X})$ between the portfolio $\boldsymbol{b} \in \mathcal{B}$ and the stock vector $\boldsymbol{X}$ given reference portfolio $\boldsymbol{b}^\star$ is defined by

$$d(\boldsymbol{b}, \boldsymbol{X}) = -\log \frac{\boldsymbol{b}^t\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}.$$

Definition: The *distortion between sequences* $\boldsymbol{b}^n$ and $\boldsymbol{X}^n$ is defined by

$$d(\boldsymbol{b}^n, \boldsymbol{X}^n) = \frac{1}{n} \sum_{j=1}^{n} d(\boldsymbol{b}_j, \boldsymbol{X}_j).$$

Thus

$$d(\boldsymbol{b}^n, \boldsymbol{X}^n) = \frac{1}{n} \log \left(\prod_{j=1}^{n} \boldsymbol{b}_j^t\boldsymbol{X}_j \Big/ \prod_{j=1}^{n} \boldsymbol{b}^{\star t}\boldsymbol{X}_j \right)$$

is the difference in empirical growth rate between portfolios $\boldsymbol{b}^n$ and $\boldsymbol{b}^{\star n}$.

Rate distortion with a remote source, otherwise known as the indirect rate distortion problem, was investigated by Berger [8, pp. 78–81]. The model used is shown in Fig. 2. The aim is to reconstruct the source X from the encoded data with minimal distortion. The distortion function $d(x, \tilde{x})$ is defined between the source X and the reconstruction $\tilde{X}$. However, unlike the standard rate distortion problem, the encoder cannot observe the source X, but can only access Y, a noisy version of X. The joint distribution $F(x, y)$ of X and Y is known both at the encoder and the decoder. The decoder outputs the reconstruction $\tilde{X}$ based on the encoded data.

For a conditional probability distribution $F(\tilde{x} \mid y)$ we define the expected distortion $\bar{d}(F)$ for the indirect rate distortion problem as

$$\bar{d}(F) = \int d(x,\tilde{x})\,dF(x)\,dF(y \mid x)\,dF(\tilde{x} \mid y) = Ed(X,\tilde{X})$$

where the expectation is taken over the joint distribution of $(\tilde{X}, Y, X)$ with the assumption that $\tilde{X}$ and X are conditionally independent given Y.

Berger showed that the distortion rate function $D(R)$ can be written as

$$D(R) = \min_{F(\tilde{x}|y):\ I(Y;\tilde{X})\le R} \bar{d}(F) = \min_{F(\tilde{x}|y):\ I(Y;\tilde{X})\le R,\ \tilde{X}\to Y\to X} Ed(X,\tilde{X}). \qquad (2)$$

We now prove the theorem.

Proof of Theorem 1: For any code $(i_n(\cdot), \boldsymbol{b}^n(\cdot))$ the increase in growth rate Δ_n associated with the code $\mathcal{C}$ is

$$\begin{aligned}\Delta_n &= \frac{1}{n} E \log \prod_{j=1}^{n} \frac{\boldsymbol{b}_j^t(i_n(V^n))\boldsymbol{X}_j}{\boldsymbol{b}^{\star t}\boldsymbol{X}_j} \\ &= \frac{1}{n}\sum_{j=1}^{n} E\log \frac{\boldsymbol{b}_j^t(i_n(V^n))\boldsymbol{X}_j}{\boldsymbol{b}^{\star t}\boldsymbol{X}_j} \\ &= -\frac{1}{n}\sum_{j=1}^{n} Ed(\boldsymbol{b}_j(i_n(V^n)), \boldsymbol{X}_j) \\ &= -Ed(\boldsymbol{b}^n(i_n(V^n)), \boldsymbol{X}^n).\end{aligned}$$

Therefore, a description-rate growth-rate pair (R,Δ) is achievable if there exists a sequence of $(2^{nR}, n)$ codes $(i_n(\cdot), \boldsymbol{b}^n(\cdot))$ for which

$$\lim_{n\to\infty} Ed(\boldsymbol{b}^n(i_n(V^n)), \boldsymbol{X}^n) \le -\Delta.$$

Also, $-\Delta(R)$ is the infimum of all $-\Delta$ (hence $\Delta(R)$ is the supremum of all Δ) such that (R,Δ) is in the closure of the set of all achievable pairs.

Note that the encoder observes the side information V, and the distortion $d(\boldsymbol{b}, \boldsymbol{X})$ is a function of the investor's portfolio choice $\boldsymbol{b}$ and the stock outcome $\boldsymbol{X}$. Hence this is an indirect rate distortion problem. The equivalence can also be observed by comparing Fig. 1 and Fig. 2.

Note that

$$d(\boldsymbol{b}, \boldsymbol{X}) = \log \frac{\boldsymbol{b}^{\star t}\boldsymbol{X}}{\boldsymbol{b}^t\boldsymbol{X}} \ge \log \min_i b_i^\star.$$

If $\min_i b_i^\star > 0$, then $d(\boldsymbol{b}, \boldsymbol{X})$ is bounded below and we can apply existing theory. Otherwise, let

$$\hat{\boldsymbol{b}} = \left(\frac{1}{m}, \cdots, \frac{1}{m}\right)$$

denote the uniform portfolio. We can write

$$\begin{aligned} d(\boldsymbol{b}, \boldsymbol{X}) &= \log \frac{\boldsymbol{b}^{\star t}\boldsymbol{X}}{\boldsymbol{b}^t\boldsymbol{X}} \frac{\hat{\boldsymbol{b}}^t\boldsymbol{X}}{\hat{\boldsymbol{b}}^t\boldsymbol{X}} \\ &= \log \frac{\boldsymbol{b}^{\star t}\boldsymbol{X}}{\hat{\boldsymbol{b}}^t\boldsymbol{X}} + \log \frac{\hat{\boldsymbol{b}}^t\boldsymbol{X}}{\boldsymbol{b}^t\boldsymbol{X}}.\end{aligned}$$

The first term does not depend on the portfolio choice $\boldsymbol{b}$, so it can be ignored in the calculations. The second term is bounded below by $-\log m$. Therefore, we can use the distortion measure $\log(\hat{\boldsymbol{b}}^t\boldsymbol{X}/\boldsymbol{b}^t\boldsymbol{X})$, and incorporate the first term later. Either way will give the same result.

Using Berger's result (2) we can write the incremental growth rate as

$$\begin{aligned}\Delta(R) &= -\min_{F(\boldsymbol{b}|v):\ I(\boldsymbol{b};V)\le R,\ \boldsymbol{b}\to V\to \boldsymbol{X}} Ed(\boldsymbol{b}, \boldsymbol{X}) \\ &= \max_{F(\boldsymbol{b}|v):\ I(\boldsymbol{b};V)\le R,\ \boldsymbol{b}\to V\to \boldsymbol{X}} E\log \frac{\boldsymbol{b}^t\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}.\end{aligned}$$

This proves Theorem 1. □

Generally, in rate distortion theory the aim of the decoder is to reconstruct the source so as not to exceed a certain distortion level. The distortion function shows the distortion between the source outcome and the reproduction. Although the investment problem is mathematically equivalent to a rate distortion problem, the interpretation and motivation are slightly different. The investor is not trying to reconstruct the source. Rather, he takes an action by devising an investment strategy which will be effective on the particular outcome. The distortion is a function of the outcome and the action taken. It represents the gain of using an investment strategy $\boldsymbol{b}$ over that of using $\boldsymbol{b}^\star$ on the stock vector $\boldsymbol{X}$.

An equivalent characterization of Theorem 1 is

$$\Delta(R) = \max_{F(\tilde{v}|v):\ I(\tilde{V};V)\le R,\ \tilde{V}\to V\to \boldsymbol{X}} E\log \frac{\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}. \qquad (3)$$

Here the randomness of $\boldsymbol{b}$ in Theorem 1 is captured by the random variable $\tilde{V}$, and the portfolio $\boldsymbol{b}^\star$ depends on the conditional distribution $F(\tilde{v} \mid v)$ chosen for $\tilde{V}$. For a given distribution of $\tilde{V}$, $\boldsymbol{b}^\star(\tilde{v})$ denotes the optimum portfolio when the stocks are distributed according to the conditional distribution $F(\boldsymbol{x} \mid \tilde{v})$ of $\boldsymbol{X}$ given $\tilde{V} = \tilde{v}$.

We now show that $\Delta(R)$ is bounded above by R.

Theorem 2:

$$\Delta(R) \le R.$$

Proof: From Barron and Cover [6], we know that side information $\tilde{V}$ increases the investor's growth rate by at most

$$\Delta W = E\log \frac{\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} \le I(\tilde{V};\boldsymbol{X}).$$

Also, for $\tilde{V} \to V \to \boldsymbol{X}$

$$I(\tilde{V};\boldsymbol{X}) \le I(\tilde{V};V)$$

by the data processing inequality [12, p. 32].

Combining this with (3) we have

$$\begin{aligned}\Delta(R) &\le \max_{F(\tilde{v}|v):\ I(\tilde{V};V)\le R,\ \tilde{V}\to V\to \boldsymbol{X}} I(\tilde{V};\boldsymbol{X}) \\ &\le \max_{F(\tilde{v}|v):\ I(\tilde{V};V)\le R,\ \tilde{V}\to V\to \boldsymbol{X}} I(\tilde{V};V) \\ &\le R.\end{aligned}$$ □

Using rate distortion theory, we argue that the incremental growth rate $\Delta(R)$ is a concave, nondecreasing, and differentiable function of R. Thus $\Delta'(R)$, the derivative of $\Delta(R)$, decreases as the rate of description increases. Every additional bit of description causes less increase in the growth rate. We are interested in $\Delta'(0)$, the largest increase in growth rate per bit of description. This motivates our next definition.

Definition: The *initial efficiency* $\Delta'(0)$ is the derivative of $\Delta(R)$ at the origin.

The initial efficiency is a measure of the effectiveness of the description of V in improving the growth rate. Theorem 2 implies that $\Delta'(0) \le 1$ for any market. We expect the initial efficiency to be related to the dependency between V and $\boldsymbol{X}$. In Section IV we analyze the horse race market and observe that the initial efficiency for the horse race market is the square of the Hirschfeld–Gebelein–Rényi maximal correlation between V and $\boldsymbol{X}$. We prove in Section V that the initial efficiency is 1 when the encoder can observe the stock vector outcomes (i.e., $V = \boldsymbol{X}$).

In the next section we calculate $\Delta(R)$ and $\Delta'(0)$ for a particular example.

III. An Example

This section is devoted to the calculation of the incremental growth rate $\Delta(R)$ and the initial efficiency $\Delta'(0)$ for a specific example. Recall that $\Delta(R)$ is the maximum increase in growth rate of investment in the stock market $\boldsymbol{X}$ when the correlated information V is described at rate R. The incremental growth rate $\Delta(R)$ is given by

$$\Delta(R) = \max_{F(\tilde{v}|v):\ I(\tilde{V};V)\le R,\ \tilde{V}\to V\to X} E \log \frac{\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}$$

as in (3). The initial efficiency $\Delta'(0)$ is the derivative of $\Delta(R)$ at the origin.

In this example we let $V = \boldsymbol{X}$. Therefore, the stock vector is described to the investor at rate R. The "side information" V is the market itself and the only difficulty is the rate constraint R. What should be said about $\boldsymbol{X}$?

Suppose the stock market consists of only two stocks. The first one is cash; it is always equal to 1. The second stock (the "hot" stock) either goes up by a factor of 2 (with probability α) or goes down by a factor of 2 (with probability $1-\alpha$). That is,

$$\boldsymbol{X} = \begin{cases} (1,2), & \text{with probability } \alpha \\ (1,\frac{1}{2}), & \text{with probability } 1-\alpha. \end{cases}$$

We assume $1/3 < \alpha < 2/3$.

The investor distributes his money between cash and the hot stock according to the portfolio $\boldsymbol{b}$, where

$$\boldsymbol{b} = (b, 1-b), \qquad 0 \le b \le 1.$$

Then the growth rate $W(\boldsymbol{b}, p(\boldsymbol{x}))$ is given by

$$\begin{aligned} W(\boldsymbol{b}, p(\boldsymbol{x})) &= W(b,\alpha) = E\log S \\ &= \alpha \log(b + 2(1-b)) \\ &\quad + (1-\alpha)\log\left(b + \frac{1}{2}(1-b)\right). \end{aligned}$$

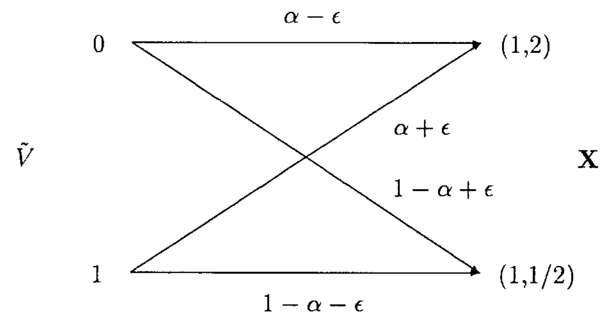

Fig. 3. $\tilde{V}$ and $\boldsymbol{X}$ for the hot-stock example.

For $1/3 < \alpha < 2/3$, the optimum portfolio is

$$b^\star = 2 - 3\alpha \triangleq b^\star_\alpha$$

and the optimum growth rate is given by

$$\begin{aligned} W^\star(p(\boldsymbol{x})) = W^\star(\alpha) &= W\left(b^\star_\alpha, \alpha\right) \\ &= \log\frac{3}{2} + \alpha - H(\alpha) \end{aligned}$$

where $H(\alpha) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ is the binary entropy function.

In order to calculate $\Delta(R)$, consider $\tilde{V} \sim \mathrm{Bern}(1/2)$ where $\tilde{V}$ and $\boldsymbol{X}$ are related through a binary channel as shown in Fig. 3. We will choose ϵ such that

$$0 \le \epsilon < \min\left(\alpha - \frac{1}{3}, \frac{2}{3} - \alpha\right). \tag{4}$$

Then

$$R = I(\tilde{V};V) = I(\tilde{V};\boldsymbol{X}) = H(\alpha) - \frac{1}{2}H(\alpha-\epsilon) - \frac{1}{2}H(\alpha+\epsilon)$$

and the increase in the growth rate is given by

$$\begin{aligned} E\log\frac{\boldsymbol{b}^t(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} &= E_{\tilde{V}} W^\star(p(\boldsymbol{x} \mid \tilde{V})) - W^\star(p(\boldsymbol{x})) \\ &= \frac{1}{2}\left(\log\frac{3}{2} + \alpha - \epsilon - H(\alpha-\epsilon)\right) \\ &\quad + \frac{1}{2}\left(\log\frac{3}{2} + \alpha + \epsilon - H(\alpha+\epsilon)\right) \\ &\quad - \left(\log\frac{3}{2} + \alpha - H(\alpha)\right) \\ &= H(\alpha) - \frac{1}{2}H(\alpha-\epsilon) - \frac{1}{2}H(\alpha+\epsilon) \\ &= I(\tilde{V};\boldsymbol{X}). \end{aligned}$$

Therefore, we have $\Delta = R$ for this choice of $p(\tilde{v} \mid v)$. Note that $\Delta = R$ as long as ϵ satisfies (4). By Theorem 2, $\Delta(R) \le R$, so we conclude

$$\Delta(R) = R, \qquad \text{for small } R$$

and

$$\Delta'(0) = 1.$$

In this simple "hot stock" example, we have seen that when $V = \boldsymbol{X}$, the initial efficiency is 1. Also for small rates the upper bound $\Delta(R) = R$ is achieved. We will see in Section V that the initial efficiency $\Delta'(0) = 1$ for all stock markets with side information $V = \boldsymbol{X}$.

IV. Horse Race Market

Theorem 1 provides an expression for the incremental growth rate in the general market. In this section, we specialize this result to an important market, the *horse race market*. We will see that the incremental growth rate can now be expressed in terms of mutual information. We solve some examples in the horse race market and calculate the initial efficiency. We will show that the initial efficiency for the horse race market is the square of the maximal correlation between V and the market $\boldsymbol{X}$.

A. Introduction

Consider a special kind of market in which only one of the stocks can be positive at a given time. If the price relative for some stock, say stock i, is nonzero, then all the other price relatives are equal to 0. This is equivalent to saying that the probability distribution on the stock vector $\boldsymbol{X}$ is

$$\boldsymbol{X} = (0, \cdots, 0, o_i, 0, \cdots, 0) = o_i \boldsymbol{e}_i, \quad \text{with probability } p_i \tag{5}$$

where $\sum_{i=1}^m p_i = 1$ and $\boldsymbol{e}_i$ is the ith standard basis vector for $\mathbb{R}^m$.

We call this market the *horse race market* because it is equivalent to gambling in a horse race. Note that no cash reserve is allowed. All available wealth must be invested in the horses, none can be set aside. The cash-aside option is no loss of generality in the growth rate of wealth if the odds are fair ($\sum_{i=1}^m 1/o_i = 1$) or superfair ($\sum_{i=1}^m 1/o_i < 1$).

Another characterization of the horse race market is through a discrete random variable X that takes on m values $\{1, \cdots, m\}$. The random variable X denotes the winning horse in the race. Horse i wins with probability p_i, i.e., $P(X = i) = p_i$. The investor distributes his money among the m horses according to the portfolio $\boldsymbol{b} = (b_1, \cdots, b_m)$, where $\sum_{i=1}^m b_i = 1$.

The investment pays off o_i odds for one. If $X = i$, then the investor receives o_i dollars for each dollar invested in i, and he loses all the money invested in $j \neq i$. We will use the notation b_i and $b(i)$ (similarly, o_i and $o(i)$) interchangeably.

The wealth of the gambler is multiplied by the factor

$$S(X) = b(X)o(X).$$

It is easy to see the equivalence of this scalar notation with the vector notation in (5). We will use both the vector notation $\boldsymbol{X}$ and the scalar notation X for the horse race market.

We can obtain a closed-form solution for the optimum portfolio $\boldsymbol{b}^\star$ in the horse race market. We first calculate the growth rate $W(b, p)$.

$$\begin{aligned} W(b,p) &= \int \log \boldsymbol{b}^t \boldsymbol{X} \, dF(\boldsymbol{X}) \\ &= \sum_{i=1}^m p_i \log b_i o_i \\ &= \sum_{i=1}^m p_i \log \frac{b_i}{p_i} + \sum_{i=1}^m p_i \log p_i + \sum_{i=1}^m p_i \log o_i \\ &= -D(p \,\|\, b) - H(p) + \sum_{i=1}^m p_i \log o_i. \end{aligned}$$

Note that $D(p \,\|\, b)$ is minimized (and is equal to 0) when $\boldsymbol{b} = \boldsymbol{p}$. Therefore, the optimum portfolio is

$$b_i^\star = p_i, \qquad \text{for } i = 1, \cdots, m.$$

The optimum strategy for the gambler is to bet in proportion to the underlying probability distribution. This was established by Kelly [16]. Proportional gambling is also known as *Kelly gambling*. The optimum growth rate $W^\star$ is then given by

$$W^\star(X) = \sum_{i=1}^m p_i \log o_i - H(X). \tag{6}$$

Now suppose the gambler has side information Y about the outcome of the race. We assume that X and Y have a known joint distribution. Then the optimum portfolio is given by

$$b_{i|y}^\star = p_{i|y} = \Pr\{X = i \mid Y = y\}$$

which results in the optimum growth rate

$$W^\star(X \mid Y) = \sum_{i=1}^m p_i \log o_i - H(X \mid Y). \tag{7}$$

From (6) and (7), the increase in growth rate due to the presence of side information is given by

$$\Delta = W^\star(X \mid Y) - W^\star(X) = I(X;Y). \tag{8}$$

Thus in the horse race market the increase in growth rate with side information Y is the mutual information between X and Y.

In the next subsection, we find the incremental growth rate $\Delta(R)$ for the horse race market.

B. Incremental Growth Rate in Horse Race

The following theorem specializes the general expression for $\Delta(R)$ to the horse race market.

Theorem 3: If side information V is available in the horse race market X, the incremental growth rate is given by

$$\Delta(R) = \max I(\tilde{V}; X),$$

where $\tilde{V} \to V \to X$ forms a Markov chain and the maximum is over all conditional distributions $F(\tilde{v} \mid v)$ satisfying the rate constraint $I(\tilde{V}; V) \leq R$

Proof: From (3), we must maximize

$$E \log(\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X} / \boldsymbol{b}^{\star t}\boldsymbol{X})$$

over all $\tilde{V}$ satisfying the Markov condition and the rate constraint. Let $F(\tilde{v})$ be the marginal distribution of $\tilde{V}$ and $F(\tilde{v} \mid i)$ be the conditional distribution conditioned on $X = i$. Similarly, let $p(i) = \Pr(X = i)$ be the marginal and $p(i \mid \tilde{v})$ be the conditional distribution for X. We recall that $\boldsymbol{b}^\star(\tilde{v})$ is the optimum portfolio when X is distributed according to $p(i \mid \tilde{v})$. Then

$$\begin{aligned} E \log \frac{\boldsymbol{b}^t(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} &= \int_{\tilde{v}} \sum_{i=1}^m p_i \log \frac{b_i^\star(\tilde{v}) o_i}{p_i o_i} \, dF(\tilde{v} \mid i) \\ &= \int_{\tilde{v}} \sum_{i=1}^m p_i \log \frac{b_i^\star(\tilde{v})}{p_i} \, dF(\tilde{v} \mid i) \end{aligned}$$

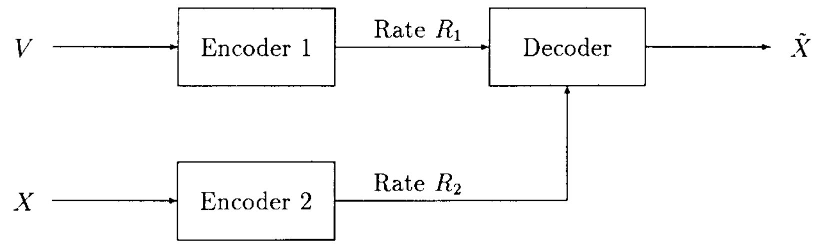

Fig. 4. Source coding with side information.

$$= H(X) + \int_{\tilde{v}} \sum_{i=1}^{m} p(i \mid \tilde{v}) \log\left(\frac{b_i^\star(\tilde{v})}{p(i \mid \tilde{v})} p(i \mid \tilde{v})\right) dF(\tilde{v})$$
$$= H(X) - H(X \mid \tilde{V}) - \int_{\tilde{v}} D\big(p(i \mid \tilde{v}) \,\|\, b_i^\star(\tilde{v})\big)\, dF(\tilde{v})$$
$$= I(\tilde{V}; X) - \int_{\tilde{v}} D\big(p(i \mid \tilde{v}) \,\|\, b_i^\star(\tilde{v})\big)\, dF(\tilde{v}).$$

The above expression is maximized when the last term is zero, that is, when $b_i^\star(\tilde{v}) = p(i \mid \tilde{v})$. Therefore,

$$\max_{F(\tilde{v}|v)} E \log \frac{\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} = \max_{F(\tilde{v}|v)} I(\tilde{V}; X)$$

where $\tilde{V}$ is chosen to satisfy the Markov relationship and the rate constraint. □

The problem of finding the incremental growth rate in the horse race market can be reduced to that of source coding with side information. Source coding with side information for discrete random variables was independently investigated by Wyner [22] and by Ahlswede and Körner [5]. The block diagram for source coding with side information is illustrated in Fig. 4. Suppose (V_i, X_i), $i = 1, \cdots, n$, are independent, identically distributed copies of the pair (V, X). The first encoder observes V^n and encodes it using R_1 bits per symbol. The second encoder observes X^n and uses R_2 bits per symbol to compress it. The decoder receives these two descriptions and forms $\tilde{X}^n$, the reproduction of X^n with small probability of error. The rate region is given by

$$\mathcal{R} = \{(R_1, R_2) : I(\tilde{V}; V) \le R_1, H(X \mid \tilde{V}) \le R_2 \text{ for some } \tilde{V} \to V \to X\}.$$

Let $C(R)$ be the minimum rate at which the second encoder must transmit when the first encoder uses R bits per symbol. Then

$$C(R) = \min_{(R, R_2) \in \mathcal{R}} R_2 = \min_{F(\tilde{v}|v):\ I(\tilde{V};V) \le R,\ \tilde{V} \to V \to X} H(X \mid \tilde{V}).$$

The quantity $C(R)$ represents the minimum descriptive complexity of X in the presence of a rate R description of V. Comparing the functional forms of $C(R)$ and $\Delta(R)$, we can write

$$\Delta(R) = H(X) - C(R).$$

Note that $H(X) - C(R)$ is the maximum decrease in the descriptive complexity of X when V is described at rate R.

The reasoning behind this relationship is as follows. The amount of decrease in the descriptive complexity of X^n when V^n is encoded by encoder $i_n(\cdot)$ is

$$H(X^n) - H(X^n \mid i_n(V^n)) = I(X^n; i_n(V^n)).$$

By (8), this is exactly the increase in the growth rate when the compressed version $i_n(V^n)$ of V^n is available at the investor. Therefore, the maximum decrease in the descriptive complexity is exactly equal to the maximum increase in the growth rate.

The incremental growth rate in the horse race market is also closely related to the entropy characterization problem of Csiszár and Körner [13, pp. 303–358]. The form appearing in Theorem 3 was used by Ahlswede, Gács, and Körner to prove strong converses of some multi-user coding theorems [3]. This form also appears in hypothesis testing with communication constraints of Ahlswede and Csiszár [1]. When one tests the hypothesis of a given bivariate distribution $p(x, v)$ against the alternative of independence under a communication constraint on V, for any fixed type I error, the type II error decreases exponentially with the sample size n. Ahlswede and Csiszár have defined $\theta(R)$ as the largest exponent of decay when V is described at rate R. The functional form of the incremental growth rate $\Delta(R)$ is exactly equal to this best exponent $\theta(R)$.

Witsenhausen and Wyner [21] investigated the extremization appearing in Theorem 3 closely. In particular, for V discrete they showed that to compute the required maximization, $\tilde{V}$ needs to take at most $|\mathcal{V}| + 1$ values, where $|\mathcal{V}|$ is the alphabet size for V. This leads us to the following corollary.

Corollary 1: For the horse race market with discrete side information V described at rate R, the incremental growth rate is given by

$$\Delta(R) = \max_{p(\tilde{v}|v):\ |\tilde{\mathcal{V}}| \le |\mathcal{V}|+1,\ I(\tilde{V};V) \le R,\ \tilde{V} \to V \to X} I(\tilde{V}; X).$$

Next we find $\Delta(R)$ for $V = X$.

Corollary 2: For the horse race market when $V = X$

$$\Delta(R) = R, \qquad \text{for } R \le H(X)$$

and $\Delta'(0) = 1$.

Proof: Taking $V = X$ in Theorem 3 gives $R = I(\tilde{V}; X)$ and $\Delta = I(\tilde{V}; X)$. Hence

$$\Delta(R) = R$$

and

$$\Delta'(0) = 1.$$

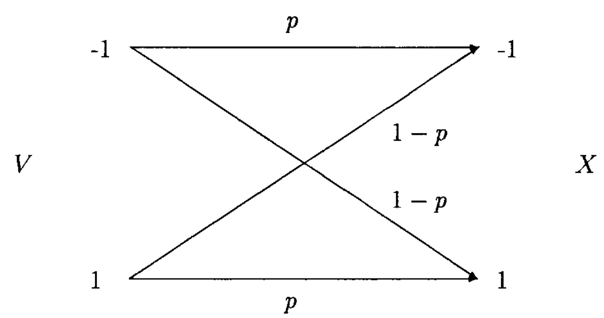

Fig. 5. V and X binary.

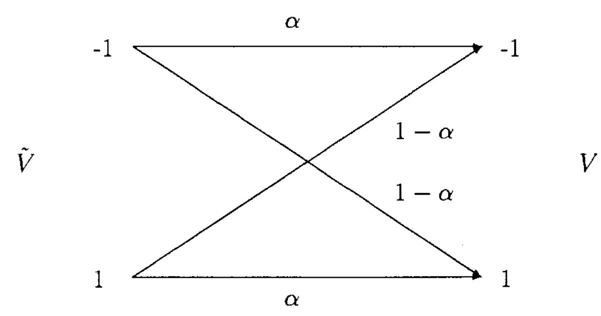

Fig. 6. $\tilde{V}$ and V binary.

For $R \geq H(X)$, the encoder describes X perfectly and no further improvement can be made. □

Note that when $V = X$, any $\tilde{V}$ used in he descriptions is optimal in the sense that $\Delta = R = I(\tilde{V}; X)$. Hence any bit of description of X improves the growth rate by exactly one bit. Every description is maximally efficient.

We now investigate the incremental growth rate for some horse race examples.

C. Horse Race Market: V and X Binary

In this first example for the horse race market, V and X are binary random variables with a symmetric joint distribution. The theorem below establishes $\Delta(R)$.

Theorem 4: Suppose $V \sim \text{Bern}(1/2)$ and is related to X through a binary symmetric channel (BSC) with crossover probability p as in Fig. 5. Let

$$H(p) = -p \log p - (1-p) \log(1-p)$$

be the binary entropy function and $\alpha * p = \alpha(1-p) + (1-\alpha)p$ denote the parameter of a BSC that is the cascade of two BSC's with crossover probabilities α and p.

Then

$$(R, \Delta(R)) = (1 - H(\alpha), 1 - H(\alpha * p))$$

where $0 \leq \alpha \leq 1$.

The form appearing in the above theorem was established by Witsenhausen and Wyner [21] using convexity arguments. Their results also lead to the calculation of $\Delta(R)$ when V and X are related through a general binary-input binary-output channel, a binary erasure channel, or a Hamming channel with V uniform. Similar convexity arguments for all binary-input binary-output channels also appear in Witsenhausen [19] and Ahlswede–Körner [4]. Chayat and Shamai [11] have extended the convexity to the class of binary-input discrete- or continuous-output memoryless symmetric channels.

In order to keep the example self-contained, we provide a proof of Theorem 4 using the following lemma of Wyner and Ziv [23].

Lemma 1 (Corollary to "Mrs. Gerber's Lemma"): Suppose V and X are two binary random variables connected through a BSC(p) as in Fig. 5. Then

$$H(V \mid \tilde{V}) \geq a \Rightarrow H(X \mid \tilde{V}) \geq H(p * H^{-1}(a))$$

for any $p(\tilde{v} \mid v)$ satisfying $\tilde{V} \rightarrow V \rightarrow X$.

Proof of Theorem 4: The proof involves choosing an auxiliary random variable $\tilde{V}$ such that $\tilde{V} \rightarrow V \rightarrow X$, $I(\tilde{V}; V) \leq R$, and the maximum in Theorem 3 is achieved. For any $\tilde{V}$ satisfying $I(\tilde{V}; V) \leq R$, we have $H(V \mid \tilde{V}) \geq 1 - R$. By Lemma 1, this implies $H(X \mid \tilde{V}) \geq H(p * H^{-1}(1-R))$. Therefore,

$$I(\tilde{V}; X) \leq 1 - H(p * H^{-1}(1-R)).$$

In order to achieve this upper bound we consider $\tilde{V}$ that is related to V through another binary symmetric channel of crossover probability α as in Fig. 6. Here α is chosen to satisfy the rate constraint.

Then

$$\begin{aligned} I(\tilde{V}; V) &= H(V) - H(V \mid \tilde{V}) \\ &= 1 - H(\alpha) \\ &= R \end{aligned} \tag{9}$$

and

$$\begin{aligned} I(\tilde{V}; X) &= H(X) - H(X \mid \tilde{V}) \\ &= 1 - H(p * \alpha) \\ &= 1 - H(p * H^{-1}(1-R)). \end{aligned} \tag{10}$$

Thus the above choice of $\tilde{V}$ is optimal, and using (9) and (10) we can characterize the $(R, \Delta(R))$ pairs as

$$(R, \Delta(R)) = (1 - H(\alpha), 1 - H(p * \alpha)).$$ □

A typical $\Delta(R)$ curve can be found in Fig. 7. Next we calculate the initial efficiency.

Theorem 5: When X and V are binary as in Theorem 4, the initial efficiency for the horse race is

$$\Delta'(0) = (1 - 2p)^2.$$

Proof: We use $\Delta(R)$ as in Theorem 4 to calculate the initial efficiency

$$\begin{aligned} \Delta'(R) &= \frac{d\Delta(R)}{dR} \\ &= \frac{d\Delta/d\alpha}{dR/d\alpha} \\ &= \frac{\log\left(\frac{\alpha * p}{1 - \alpha * p}\right)(1-2p)}{\log\left(\frac{\alpha}{1-\alpha}\right)}. \end{aligned}$$

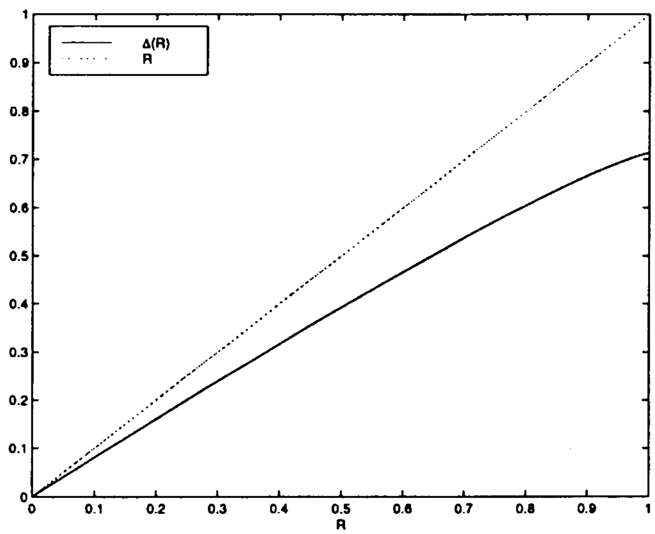

Fig. 7. $\Delta(R)$ for binary horse race.

To evaluate $\Delta'(R)$ at the origin, we let $\alpha \rightarrow 1/2$ $(R = 1 - h(\alpha) \rightarrow 0)$ and use L'Hôpital's rule

$$\Delta'(0) = \lim_{\alpha \rightarrow 1/2} \frac{\log\left(\frac{\alpha * p}{1-\alpha * p}\right)(1-2p)}{\log\left(\frac{\alpha}{1-\alpha}\right)} = (1-2p)^2. \quad \square$$

We note that $(1-2p)$ is the correlation between V and X. Therefore, the initial efficiency is the square of the correlation, that is, $\Delta'(0) = \rho^2$.

D. Horse Race: X and V Gaussian

In this subsection we calculate $\Delta(R)$ when X and V are jointly Gaussian. The notion of a horse race market and the result of Theorem 3 can be generalized to continuous X, in which case the portfolio $b(x)$ becomes a probability density function.

Theorem 6: Suppose X and V are jointly Gaussian with correlation ρ. Then

$$\Delta(R) = \frac{1}{2} \log \frac{1}{1 - \rho^2(1 - 2^{-2R})}.$$

In the proof we will use the following result by Bergmans [9], which is the conditional version of the entropy power inequality.

Lemma 2 (Conditional Entropy Power Inequality): Suppose V and X are related through an additive white Gaussian noise (AWGN) channel as in Fig. 8. Also suppose $\tilde{V} \rightarrow V \rightarrow X$. Then

$$h(V \mid \tilde{V}) \geq a \Rightarrow h(X \mid \tilde{V}) \geq g\left(\sigma_Z^2 + g^{-1}(a)\right)$$

where $g(x) = \frac{1}{2} \log 2\pi e x$ is the entropy function for a Gaussian random variable with variance x.

The lemma follows from the entropy power inequality and from the convexity of $g(\sigma_Z^2 + g^{-1}(a))$ as a function of a.

Next we prove the theorem.

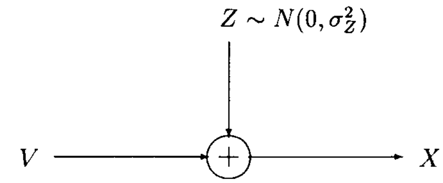

Fig. 8. Additive white Gaussian noise channel.

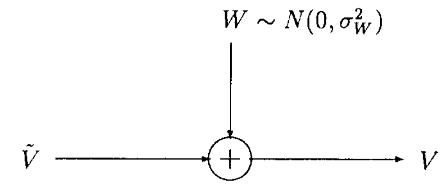

Fig. 9. $\tilde{V}$ and V Gaussian.

Proof of Theorem 6: Without loss of generality we assume that V and X are related through an additive white Gaussian noise channel as in Fig. 8 and $V \sim N(0,1)$. That is, $X = V + Z$, V and Z are independent, and $Z \sim N(0, \sigma_Z^2)$ with

$$\sigma_Z^2 = \frac{1-\rho^2}{\rho^2}.$$

We now use we use Lemma 2 to find an upper bound for $\Delta(R)$. For $\tilde{V}$ satisfying the rate constraint $I(\tilde{V}; V) \leq R$ of Theorem 3

$$h(V \mid \tilde{V}) \geq \frac{1}{2} \log(2\pi e 2^{-2R}).$$

Then by Bergmans' lemma

$$\begin{aligned} I(\tilde{V}; X) &= h(X) - h(X \mid \tilde{V}) \\ &\leq \frac{1}{2} \log 2\pi e\left(1 + \sigma_Z^2\right) \\ &\quad - \frac{1}{2} \log 2\pi e\left(\sigma_Z^2 + g^{-1}\left(\frac{1}{2} \log 2\pi e 2^{-2R}\right)\right) \\ &= \frac{1}{2} \log \frac{1+\sigma_Z^2}{\sigma_Z^2 + 2^{-2R}} \\ &= \frac{1}{2} \log \frac{1}{1-\rho^2(1-2^{-2R})}. \qquad (11) \end{aligned}$$

In order to achieve this upper bound, a natural choice for $\tilde{V}$ is to relate it to V through another AWGN channel with the noise variance chosen to satisfy the rate constraint (see Fig. 9). That is, $V = \tilde{V} + W$, where $\tilde{V}$ and W are independent and $W \sim N(0, \sigma_W^2)$. We choose σ_W so that $I(\tilde{V}; V) = R$. Hence $\sigma_W^2 = 2^{-2R}$. Then

$$\begin{aligned} I(\tilde{V}; X) &= h(X) - h(X \mid \tilde{V}) \\ &= h(X) - h(W + Z) \\ &= \frac{1}{2} \log \frac{1+\sigma_Z^2}{\sigma_Z^2 + \sigma_W^2} \\ &= \frac{1}{2} \log \frac{1}{1-\rho^2(1-2^{-2R})}. \end{aligned}$$

Therefore, the upper bound in (11) is achieved and

$$\Delta(R) = \frac{1}{2} \log \frac{1}{1-\rho^2(1-2^{-2R})}. \quad \square$$

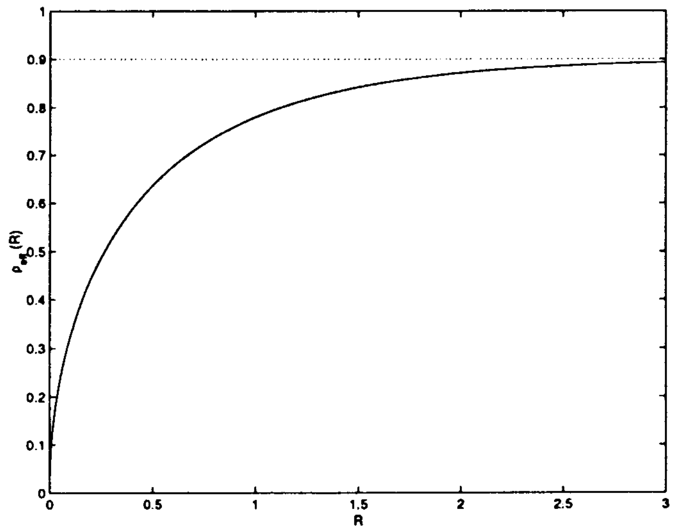

Fig. 10. Effective correlation, $\rho = 0.9$, $\rho_{\text{eff}} = \rho\sqrt{1-2^{-2R}}$.

The expression for $\Delta(R)$ suggests the definition of an effective correlation.

Definition: We define the *effective correlation* as

$$\rho_{\text{eff}} \triangleq \rho\sqrt{1-2^{-2R}}.$$

In fact, ρ_{eff} is the effective correlation between V and X when V is described at rate R. Recall that in the proof of Theorem 6, $\tilde{V}$ represents the information sent by the encoder. Let $\rho_{\tilde{V},X}$ be the correlation between this optimal description and X. Then

$$\begin{aligned}\rho_{\tilde{V},X} &= \frac{E(\tilde{V}X)}{\sqrt{E(\tilde{V}^2)E(X^2)}}\\ &= \frac{1-\sigma_W^2}{\sqrt{\left(1-\sigma_W^2\right)\left(1+\sigma_Z^2\right)}}\\ &= \rho\sqrt{1-2^{-2R}}\\ &= \rho_{\text{eff}}.\end{aligned}$$

From Fig. 10, we can observe that $\rho_{\text{eff}} \to \rho$ as $R \to \infty$. Fig. 11 shows the $\Delta(R)$ curve for $\rho = 0.9$.

Next we calculate the initial efficiency.

Theorem 7: For V and X jointly Gaussian with correlation ρ

$$\Delta'(0) = \rho^2.$$

Proof: Follows from taking the derivative of $\Delta(R)$ in Theorem 6. □

As in the binary horse race, the initial efficiency is the square of the correlation ρ between V and X. The next subsection explores the general relationship between initial efficiency and the correlation for the horse race market.

E. Initial Efficiency for the Horse Race Market

In Theorem 3, we showed that the incremental growth rate for the horse race market is given by

$$\Delta(R) = \max_{p(\tilde{v}|v):\ I(\tilde{V};V)=R,\ \tilde{V}\to V\to X} I(\tilde{V};X).$$

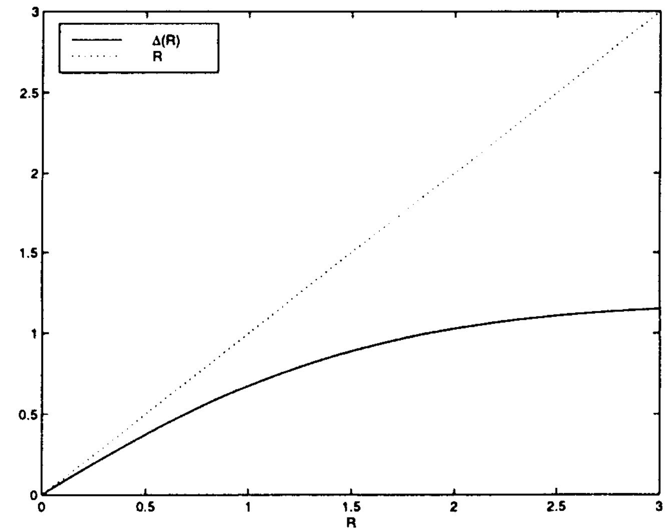

Fig. 11. $\Delta(R)$ for Gaussian horse race $\Delta(R) = \frac{1}{2}\log(1/(1-\rho^2(1-2^{-2R})))$.

We showed that $\Delta(R)$ is a concave, nondecreasing function of R. Hence $\Delta'(R)$ is largest at the origin. We called $\Delta'(0)$ the *initial efficiency*. It is a measure of how efficiently one can describe V for the investor who gambles on X.

When we calculated the initial efficiency for the binary and Gaussian examples, we observed that it is equal to the square of the correlation between V and X. In this subsection, we show that in the general horse race market the initial efficiency is related to the Hirschfeld–Gebelein–Rényi maximal correlation between V and X. Maximal correlation is a measure of dependence that is stronger and more general than the correlation coefficient.

Definition: The *Hirschfeld–Gebelein–Rényi maximal correlation* $\rho_m(V,X)$ between two random variables V and X is defined by

$$\rho_m(V,X) = \sup Eg(V)h(X)$$

where the supremum is over all functions f and g such that

$$Eg(V) = Eh(X) = 0 \qquad Eg^2(V) = Eh^2(X) = 1.$$

Note that ρ_m is obtained by taking the supremum of the correlation coefficient of $g(V)$ and $h(X)$ over functions g, h. Therefore, it is easy to see that

$$0 \le \rho_m \le 1.$$

Note, $\rho \le \rho_m$, where ρ is the correlation between V and X.

The maximal correlation was first introduced by Hirschfeld [15] for discrete random variables and by Gebelein [14] for absolutely continuous random variables. Rényi [17], [18] compared maximal correlation to other measures of dependence and provided a set of sufficient conditions for which the supremum in the definition of the maximal correlation is achieved. However, these conditions are not necessary, as shown by a simple example. For discrete-valued random variables the sufficient conditions are met and the supremum in the maximal correlation is in fact a maximum.

The maximal correlation is more general than the regular correlation in that it allows arbitrary zero-mean, variance-one functions of V and X. The actual labeling has no effect on the maximal correlation, only the joint distribution of V and X is important.

First we will observe that maximal correlation provides the unifying answer to the initial efficiency in the binary and Gaussian examples. Then we will establish the relationship between the initial efficiency and the maximal correlation for the general horse race market.

Jointly Binary: Suppose V and X are binary and distributed as in Theorem 4. Then the correlation ρ between V and X is

$$\rho(V, X) = (1 - 2p).$$

In order to calculate the maximal correlation ρ_m, we must consider all functions $g(V)$ and $h(X)$ which have mean zero and variance one. Since V and X take on values 1 and -1 both with probability $1/2$, they have mean zero and variance one. In fact, $g(V) = V$ and $h(X) = X$ are the the only functions of V and X that satisfy the mean and variance constraints. Therefore,

$$\rho_m(V, X) = \rho(V, X).$$

Hence by Theorem 5,

$$\Delta'(0) = (1 - 2p)^2 = \rho^2(V, X) = \rho_m^2(V, X).$$

Thus the initial efficiency is the square of the maximal correlation between V and X.

Jointly Gaussian: For V and X jointly Gaussian with correlation ρ, we know from [10] that

$$\rho_m = \rho.$$

Combining with Theorem 7, we have

$$\Delta'(0) = \rho^2(V, X) = \rho_m^2(V, X).$$

Hence, once again, the initial efficiency is the square of the maximal correlation.

Below we state the main theorem of this subsection.

Theorem 8: The initial efficiency in describing a random variable V for the investor who gambles on X is given by the square of the Hirschfeld–Gebelein–Rényi maximal correlation between V and X, i.e.,

$$\Delta'(0) = \rho_m^2(V, X).$$

Combining Theorem 8 with the concavity of $\Delta(R)$ and the characterization of the incremental growth rate in Theorem 3, we can write

$$\sup_{F(\tilde{v}|v):\ \tilde{V} \to V \to X} \frac{I(\tilde{V}; X)}{I(\tilde{V}; V)} = \rho_m^2(V, X).$$

This provides a connection with the hypercontraction of the Markov operator of Ahlswede and Gács [2].

The following characterization of maximal correlation due to Rényi [18] will be useful in the proof of the theorem:

$$\rho_m^2(V, X) = \sup_{Eg(V)=0,\ Eg^2(V)=1} E(E^2(g(V) \mid X)). \quad (12)$$

Proof: We prove the theorem for finite-valued side information V. Quantization allows us to convert continuous random variables to discrete ones, and by taking the supremum over quantizations we can show that the theorem also holds for continuous V.

We let $|\mathcal{V}| = k - 1$. Then by Corollary 1 we have $|\tilde{\mathcal{V}}| = k$. In order to find $\Delta'(0)$, we evaluate the ratio $\Delta(R)/R$ for small R and then take the limit as $R \to 0$. From Theorem 3, we know that R is related to $I(\tilde{V}; V)$ and Δ is related to $I(\tilde{V}; X)$. We use natural logarithm instead of logarithm to the base 2 in the calculations. Since we are interested in the ratio $I(\tilde{V}; V)/I(\tilde{V}; X)$, this will not affect our answer.

We first examine small $I(\tilde{V}; V)$ and show that if $I(\tilde{V}; V) \le \epsilon$, then $p(\tilde{v} \mid v)$ and $p(\tilde{v})$ are close in variational distance. We know from [13, p. 58] that

$$D(p \,\|\, q) \ge \frac{1}{2} \|p - q\|^2$$

where

$$D(p \,\|\, q) = \sum_x p(x) \ln(p(x)/q(x))$$

is the relative entropy distance, and

$$\|p - q\| = \sum_{x \in \mathcal{X}} |p(x) - q(x)|$$

is the variational distance between two distributions p and q.

Since

$$I(\tilde{V}; V) = D(p(\tilde{v}, v) \,\|\, p(\tilde{v}), p(v))$$

$I(\tilde{V}; V) \le \epsilon$ implies that

$$\frac{1}{2} \|p(\tilde{v}, v) - p(\tilde{v})p(v)\|^2 \le \epsilon.$$

Hence,

$$\|p(\tilde{v}, v) - p(\tilde{v})p(v)\| \le \sqrt{2\epsilon}. \quad (13)$$

But

$$\begin{aligned} \|p(\tilde{v}, v) - p(\tilde{v})p(v)\| &= \sum_{\tilde{v}, v} |p(\tilde{v}, v) - p(\tilde{v})p(v)| \\ &= \sum_{\tilde{v}, v} |p(v)p(\tilde{v} \mid v) - p(\tilde{v})p(v)| \\ &= \sum_v p(v) \sum_{\tilde{v}} |p(\tilde{v} \mid v) - p(\tilde{v})|. \quad (14) \end{aligned}$$

Combining (13) and (14), we have

$$|p(\tilde{v} \mid v) - p(\tilde{v})| \le \lambda \quad (15)$$

for all $\tilde{v}, v$ such that $p(v) \ne 0$ and

$$\lambda = \sqrt{2\epsilon} / \min\{p(v) \colon p(v) \ne 0\}.$$

Note that $\lambda \to 0$ as $\epsilon \to 0$.

Another way to express (15) is

$$p(\tilde{v} \mid v) - p(\tilde{v}) = \lambda u(\tilde{v}, v) \quad (16)$$

where $\max_{v,\tilde{v}} |u(\tilde{v},v)| = 1$ and

$$\sum_{\tilde{v}} u(\tilde{v},v) = \frac{1}{\lambda}\sum_{\tilde{v}}(p(\tilde{v}\mid v)-p(\tilde{v})) = 0 \tag{17}$$

$$\sum_{v} p(v)u(\tilde{v},v) = \frac{1}{\lambda}\sum_{v} p(v)(p(\tilde{v}\mid v)-p(\tilde{v})) = p(\tilde{v})-p(\tilde{v}) = 0. \tag{18}$$

We have seen that when $I(\tilde{V};V)$ is small, (16) can be used to express $p(\tilde{v}|v)$. We will observe that the reverse implication is also true; all such distributions $p(\tilde{v}|v)$ lead to small $I(\tilde{V};V)$. Hence, we can use (16), consider all $p(\tilde{v})$, $u(\tilde{v},v)$ satisfying the constraints and let $\lambda \to 0$ to calculate $\Delta'(0)$.

Using (16) for $p(\tilde{v}|v)$, we expand $H(\tilde{V}\mid V=v)$ in a Taylor series around $p(\tilde{v})$

$$\begin{aligned} H(\tilde{V}\mid V=v) &= H(p(\tilde{v}\mid v)) \\ &= H(p(\tilde{v})+\lambda u(\tilde{v},v)) \\ &= H(p(\tilde{v}))+\nabla H(p(\tilde{v}))\lambda \boldsymbol{u}(\tilde{v},v) \\ &\quad + \frac{1}{2}\lambda \boldsymbol{u}^t(\tilde{v},v)M_H(p(\tilde{v}))\lambda \boldsymbol{u}(\tilde{v},v)+o(\lambda^2) \end{aligned}$$

where $\boldsymbol{u}(\tilde{v},v) = (u(\tilde{v}_1,v),\cdots,u(\tilde{v}_k,v))^t$ and M_H is the Hessian of H. We can calculate $\nabla H(p(\tilde{v}))$ and $M_H(p(\tilde{v}))$ as

$$\begin{aligned} \nabla H(p(\tilde{v})) &= (-\ln p(\tilde{v}_1)-1,\cdots,-\ln p(\tilde{v}_k)-1) \\ M_H(p(\tilde{v})) &= -\mathrm{diag}(1/p(\tilde{v}_1),\cdots,1/p(\tilde{v}_k)). \end{aligned}$$

Therefore,

$$H(\tilde{V}\mid V=v) = H(p(\tilde{v}))-\sum_{\tilde{v}}\lambda(\ln p(\tilde{v})+1)u(\tilde{v},v) - \frac{1}{2}\lambda^2\sum_{\tilde{v}}\frac{u^2(\tilde{v},v)}{p(\tilde{v})}+o(\lambda^2)$$

and

$$\begin{aligned} H(\tilde{V}\mid V) &= \sum_v p(v)H(\tilde{V}\mid V=v) \\ &= H(p(\tilde{v}))-\lambda\sum_{\tilde{v}}(\ln p(\tilde{v})+1)\sum_v p(v)u(\tilde{v},v) \\ &\quad -\frac{1}{2}\lambda^2\sum_v p(v)\sum_{\tilde{v}}\frac{u^2(\tilde{v},v)}{p(\tilde{v})}+o(\lambda^2). \end{aligned}$$

By (18), the second term in the summation above is zero. Hence we can write $I(\tilde{V};V)$ as

$$\begin{aligned} I(\tilde{V};V) &= H(\tilde{V})-H(\tilde{V}\mid V) \\ &= \frac{1}{2}\lambda^2\sum_v p(v)\sum_{\tilde{v}}\frac{u^2(\tilde{v},v)}{p(\tilde{v})}+o(\lambda^2). \end{aligned}$$

Note that $I(\tilde{V};V)\to 0$ as $\lambda\to 0$.

Similarly, we can expand $H(\tilde{V}\mid X)$ in a Taylor series. Since $\tilde{V}\to V\to X$, we have

$$\begin{aligned} p(\tilde{v}\mid x) &= \sum_v p(\tilde{v}\mid v)p(v\mid x) \\ &= p(\tilde{v})+\lambda\sum_v u(\tilde{v},v)p(v\mid x). \end{aligned}$$

Thus

$$\begin{aligned} &H(\tilde{V}\mid X=x) \\ &\quad = H(p(\tilde{v}))-\lambda\sum_{\tilde{v}}(\ln p(\tilde{v})+1)\sum_v u(\tilde{v},v)p(v\mid x) \\ &\qquad -\frac{1}{2}\lambda^2\sum_{\tilde{v}}\frac{(\sum_v u(\tilde{v},v)p(v\mid x))^2}{p(\tilde{v})}+o(\lambda^2). \end{aligned}$$

We note that

$$\lambda\sum_{v,x}p(x)p(v\mid x)u(\tilde{v},v) = \lambda\sum_v p(v)u(\tilde{v},v) = 0.$$

Hence

$$\begin{aligned} I(\tilde{V};X) &= H(\tilde{V})-H(\tilde{V}\mid X) \\ &= \frac{1}{2}\lambda^2\sum_{\tilde{v}}\frac{1}{p(\tilde{v})}\sum_x p(x)\left(\sum_v u(\tilde{v},v)p(v\mid x)\right)^2 \\ &\quad +o(\lambda^2). \end{aligned}$$

Thus for $I(\tilde{V};V)$ small we can write the ratio of $I(\tilde{V};X)$ to $I(\tilde{V};V)$ as

$$\frac{I(\tilde{V};X)}{I(\tilde{V};V)} = \frac{\sum_{\tilde{v}}\frac{1}{p(\tilde{v})}\sum_x p(x)\big(\sum_v u(\tilde{v},v)p(v\mid x)\big)^2+o(1)}{\sum_{\tilde{v}}\frac{1}{p(\tilde{v})}\sum_v p(v)u^2(\tilde{v},v)+o(1)}. \tag{19}$$

In order to relate this ratio to maximal correlation, we observe that for fixed $\tilde{v}$, $u(\tilde{v},v)$ is a function of v satisfying

$$E(u(\tilde{v},V)) = 0.$$

We have used the fact that $u(\tilde{v},v)$ satisfies (18). Thus using Rényi's characterization of the maximal correlation (12), we can write

$$\frac{E(E^2(u(\tilde{v},V)\mid X))}{E(u^2(\tilde{v},V))}\le \rho_m^2$$

for all $\tilde{v}$ and $u(\tilde{v},v)$ as in (18).

Note that

$$E(u^2(\tilde{v},V)) = \sum_v p(v)u^2(\tilde{v},v)$$

and

$$E(E^2(u(\tilde{v},V)\mid X)) = \sum_x p(x)\left(\sum_v u(\tilde{v},v)p(v\mid x)\right)^2.$$

Therefore,

$$\sum_x p(x)\left(\sum_v u(\tilde{v},v)p(v\mid x)\right)^2 \le \rho_m^2\sum_v p(v)u^2(\tilde{v},v)$$

and for any probability distribution $p(\tilde{v})$

$$\sum_{\tilde{v}}\frac{1}{p(\tilde{v})}\sum_x p(x)\left(\sum_v u(\tilde{v},v)p(v\mid x)\right)^2 \le \rho_m^2\sum_{\tilde{v}}\frac{1}{p(\tilde{v})}\sum_v p(v)u^2(\tilde{v},v). \tag{20}$$

To calculate $\Delta'(0)$, we must take the supremum of the ratio $I(\tilde{V};X)/I(\tilde{V};V)$ in (19) over all $p(\tilde{v})$ and $u(\tilde{v},v)$ satisfying (17) and (18), and let $\lambda \to 0$. Using (20), we conclude that

$$\Delta'(0) \le \rho_m^2.$$

In order to complete the proof, we must show that this upper bound can actually be achieved. Suppose $g_m(v)$ is the function for which the supremum in (12) is achieved. The existence of such a g_m is guaranteed for discrete random variables [18]. Hence,

$$\rho_m^2 = E(E^2(g_m(V) \mid X)).$$

Let $\max_v |g_m(v)| = S$, $\tilde{V} \sim \text{Bern}(1/2)$ and

$$u_m(\tilde{v},v) = \begin{cases} g_m(v)/S, & \text{if } \tilde{v}=0 \\ -g_m(v)/S, & \text{if } \tilde{v}=1. \end{cases}$$

Since

$$\sum_{\tilde{v}} u_m(\tilde{v},v) = 0$$

and

$$\sum_v p(v) u_m(\tilde{v},v) = 0$$

(17) and (18) are satisfied. Also,

$$\max_{v,\tilde{v}} |u_m(\tilde{v},v)| = 1.$$

Then for $\tilde{v} = 0, 1$

$$\begin{aligned} \sum_v p(v) u_m^2(\tilde{v},v) &= \frac{1}{S^2} \sum_v p(v) g_m^2(v) \\ &= \frac{1}{S^2} E\big(g_m^2(V)\big) \\ &= \frac{1}{S^2} \end{aligned}$$

and

$$\begin{aligned} \sum_x \left(\sum_v u_m(\tilde{v},v) p(v\,|\,x) \right)^2 &= \frac{1}{S^2} \sum_x \left(\sum_v g_m(v) p(v\,|\,x) \right)^2 \\ &= \frac{1}{S^2} E(E^2(g_m(V) \mid X)) \\ &= \frac{\rho_m^2}{S^2}. \end{aligned}$$

Hence, for this $u_m(\tilde{v},v)$ and $p(\tilde{v})$

$$\begin{aligned} \frac{I(\tilde{V};X)}{I(\tilde{V};V)} &= \frac{\rho_m^2/S^2 + o(1)}{1/S^2 + o(1)} \\ &\to \rho_m^2 \quad \text{as } \lambda \to 0. \end{aligned}$$

Thus the upper bound is achieved and

$$\Delta'(0) = \rho_m^2.$$

This completes the proof. □

Note that the proof suggests a way to describe V efficiently at small rates. To form the descriptions, we should use a binary $\tilde{V}$. The conditional distribution $p(\tilde{v} \mid v)$ of $\tilde{V}$ is obtained from the marginal $p(\tilde{v})$ by moving in the direction of $g_m(v)$, the maximizing function in the maximal correlation.

Theorem 8 shows that the maximally efficient description of V pays off as the square of the maximal correlation in improving the growth rate of the investor. Initial efficiency, which shows how the first few bits help the investor, is in fact related to the dependency between V and X.

Recall that the calculation of the incremental growth rate in the horse race market can be reduced to source coding with side information and

$$C(R) = H(X) - \Delta(R)$$

where $C(R)$ is the minimum descriptive complexity of X when V is described at rate R. Hence the initial efficiency is the maximum decrease in the descriptive complexity of X per bit of description of V. This maximum decrease is the square of the Hirschfeld–Gebelein–Rényi maximal correlation.

V. General Initial Efficiency

This section explores the initial efficiency in the general market when $V = \boldsymbol{X}$. For the hot-stock example of Section III and the horse race market, we have observed that the initial efficiency is 1 if $V = \boldsymbol{X}$. Initially, the payoff is exactly 1 bit of increase in the growth rate per bit of description rate. The next theorem generalizes this result to the stock market.

Theorem 9: When $V = \boldsymbol{X}$

$$\Delta'(0) = 1$$

for the general market.

Thus when the encoder can observe the stock vector outcomes, the initial efficiency achieves its largest possible value. In fact, the proof of Theorem 9 shows that

$$\Delta(R) = R, \qquad \text{for small } R.$$

Hence, it is possible to attain the upper bound on $\Delta(R)$ for small rates.

Proof: We prove the theorem when $\boldsymbol{X}$ has a density $f(\boldsymbol{x})$. A similar proof works for general $F(\boldsymbol{x})$.

We call the stocks for which $b_i^\star > 0$ the *active* stocks. Let $A = \{i : b_i^\star > 0\}$ be the set of active stocks. Without loss of generality, we assume that $A = \{1, \cdots, k\}$. We also assume that $k \ge 2$, that is, at least two stocks are active.

Let $V = \boldsymbol{X}$. We will use the equivalent characterization of Theorem 1 given in (3)

$$\Delta(R) = \max_{F(\tilde{v}|x):\, I(\tilde{V};\boldsymbol{X}) \le R} E \log \frac{\boldsymbol{b}^{\star t}(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}.$$

Here $\boldsymbol{b}^\star$ depends on the distribution $F(\tilde{v}|x)$ and for fixed $F(\tilde{v}|x)$, $\boldsymbol{b}^\star(\tilde{v})$ is the log optimum portfolio when the distribution on the stocks is given by $F(\boldsymbol{x}|\tilde{v})$.

To find the initial efficiency, we will use a specific density $f(\tilde{v}|\boldsymbol{x})$ and a possibly suboptimal portfolio assignment $\boldsymbol{b}(\tilde{v})$. However, we will see that this achieves $\Delta'(0) = 1$ and therefore is optimal. Our choice for $f(\tilde{v}|\boldsymbol{x})$ is motivated by the proof of initial efficiency for the horse race market.

We know from the Kuhn–Tucker conditions (see [7]) that for $i \in A$

$$E\frac{X_i}{\boldsymbol{b}^{\star t}\boldsymbol{X}} = 1.$$

We choose a vector $\boldsymbol{c} \neq \boldsymbol{0}$ such that $c_i = 0$ for $i \notin A$ and

$$\sum_{i=1}^{k} c_i = 0.$$

The assumption $k \geq 2$ ensures that such a $\boldsymbol{c}$ exists. We let

$$g(\boldsymbol{X}) = \frac{\boldsymbol{c}^t\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}}.$$

Note that $Eg(\boldsymbol{X}) = 0$. Also

$$|g(\boldsymbol{x})| \leq \frac{\max_{i\in A}|c_i|}{\min_{i\in A} b_i^\star}.$$

We choose $\tilde{V} \sim \mathrm{Bern}\,(1/2)$ and

$$f(\boldsymbol{x} \mid \tilde{v}) = \begin{cases} f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x}), & \text{for } \tilde{v} = 0 \\ f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x}), & \text{for } \tilde{v} = 1 \end{cases} \tag{21}$$

where $\lambda > 0$ and small. Hence $f(\boldsymbol{x}|\tilde{v})$ is obtained from $f(\boldsymbol{x})$ by perturbing it in the direction of $f(\boldsymbol{x})g(\boldsymbol{x})$.

We now argue that $f(\boldsymbol{x}|\tilde{v})$ is a density and that the marginal density of $\boldsymbol{X}$ is $f(\boldsymbol{x})$. Since $|g(\boldsymbol{x})|$ is bounded, for small enough λ

$$\begin{aligned} f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x}) &= f(\boldsymbol{x})(1 + \lambda g(\boldsymbol{x})) \\ &\geq 0 \end{aligned}$$

and, similarly, $f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x}) \geq 0$.

Also

$$\begin{aligned} \int_{\boldsymbol{x}} f(\boldsymbol{x} \mid \tilde{v} = 0)\, d\boldsymbol{x} &= \int_{\boldsymbol{x}} f(\boldsymbol{x})\, d\boldsymbol{x} + \lambda \int_{\boldsymbol{x}} f(\boldsymbol{x})g(\boldsymbol{x})\, d\boldsymbol{x} \\ &= 1. \end{aligned}$$

and

$$\int_{\boldsymbol{x}} f(\boldsymbol{x} \mid \tilde{v} = 1)\, d\boldsymbol{x} = 1.$$

Therefore, $f(\boldsymbol{x} \mid \tilde{v})$ represents a density. Note that

$$\sum_{\tilde{v}} p(\tilde{v}) f(\boldsymbol{x} \mid \tilde{v}) = f(\boldsymbol{x})$$

so the marginal density of $\boldsymbol{X}$ is $f(\boldsymbol{x})$ as required.

Instead of using the optimum portfolio $\boldsymbol{b}^\star(\tilde{v})$ for $f(\boldsymbol{x} \mid \tilde{v})$, we will use a possibly suboptimal portfolio $\boldsymbol{b}(\tilde{v})$. Our choice of $\boldsymbol{b}(\tilde{v})$ is concentrated on the active stocks $\boldsymbol{X}_A$ and is given by

$$\boldsymbol{b}(\tilde{v}) = \begin{cases} \boldsymbol{b}^\star + \lambda \boldsymbol{c}, & \text{if } \tilde{v} = 0 \\ \boldsymbol{b}^\star - \lambda \boldsymbol{c}, & \text{if } \tilde{v} = 1. \end{cases} \tag{22}$$

Hence $\boldsymbol{b}(\tilde{v})$ is obtained from the log optimum portfolio $\boldsymbol{b}^\star$ by perturbation in the direction of $\boldsymbol{c}$. Note that $0 \leq b_i(\tilde{v}) \leq 1$ for small enough λ. Also $\sum_{i=1}^{k} b_i(\tilde{v}) = 1$; thus $\boldsymbol{b}(\tilde{v})$ is a portfolio.

The increase in growth rate Δ for this choice of $f(\tilde{v}, \boldsymbol{x})$ and $\boldsymbol{b}(\tilde{v})$ is then given by

$$\begin{aligned} \Delta &= E \log \frac{\boldsymbol{b}^t(\tilde{V})\boldsymbol{X}}{\boldsymbol{b}^{\star t}\boldsymbol{X}} \\ &= \frac{1}{2}\int_{\boldsymbol{x}} (f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log \frac{(\boldsymbol{b}^{\star t}\boldsymbol{x} + \lambda \boldsymbol{c}^t\boldsymbol{x})}{\boldsymbol{b}^{\star t}\boldsymbol{x}}\, d\boldsymbol{x} \\ &\quad + \frac{1}{2}\int_{\boldsymbol{x}} (f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log \frac{(\boldsymbol{b}^{\star t}\boldsymbol{x} - \lambda \boldsymbol{c}^t\boldsymbol{x})}{\boldsymbol{b}^{\star t}\boldsymbol{x}}\, d\boldsymbol{x} \\ &= \frac{1}{2}\int_{\boldsymbol{x}} (f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log(1 + \lambda g(\boldsymbol{x}))\, d\boldsymbol{x} \\ &\quad + \frac{1}{2}\int_{\boldsymbol{x}} (f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log(1 - \lambda g(\boldsymbol{x}))\, d\boldsymbol{x}. \end{aligned}$$

Similarly, the rate R for this $f(\tilde{v}, \boldsymbol{x})$ can be written as

$$\begin{aligned} R &= I(\tilde{V}; \boldsymbol{X}) \\ &= \frac{1}{2}\int (f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log \frac{f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x})}{f(\boldsymbol{x})}\, d\boldsymbol{x} \\ &\quad + \frac{1}{2}\int (f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log \frac{f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x})}{f(\boldsymbol{x})}\, d\boldsymbol{x} \\ &= \frac{1}{2}\int (f(\boldsymbol{x}) + \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log(1 + \lambda g(\boldsymbol{x}))\, d\boldsymbol{x} \\ &\quad + \frac{1}{2}\int (f(\boldsymbol{x}) - \lambda f(\boldsymbol{x})g(\boldsymbol{x})) \log(1 - \lambda g(\boldsymbol{x}))\, d\boldsymbol{x}. \end{aligned}$$

Hence $\Delta = R$ for small enough λ. Note that we need λ to be small so that $f(\boldsymbol{x} \mid \tilde{v})$ in (21) is a density and $\boldsymbol{b}(\tilde{v})$ in (22) is a portfolio. Also note R is small for small λ, with $R = 0$ at $\lambda = 0$. Therefore,

$$\Delta = R, \qquad \text{for small } R.$$

From Theorem 2 we know that $\Delta(R) \leq R$, which implies

$$\Delta(R) = R, \qquad \text{for small } R.$$

Hence the portfolio in (22) is in fact the optimum portfolio for the distribution in (21). Therefore,

$$\Delta'(0) = 1.$$

This proves the theorem. □

VI. Conclusion

We attempt to answer the question of what we should say about V when an investor wants to invest in the correlated stock market $\boldsymbol{X}$, and what this information is worth. How does one invest when there are communication constraints on the side information? In particular, how much do the first few bits about V help?

We provide a single-letter characterization for the incremental growth rate $\Delta(R)$, the maximum increase in growth rate of investment in the market $\boldsymbol{X}$ when correlated side information V is described at rate R. The incremental growth rate is a concave and nondecreasing function of the description rate, and the initial efficiency $\Delta'(0)$ shows the maximum increase in growth rate per bit of description of V.

When we specialize incremental growth rate to the horse race market, where only one of the stocks can be positive at a given time, we observe the equivalence to the source coding with side information of Wyner and Ahlswede–Körner. The investigation of jointly binary and jointly Gaussian horse race market shows that the initial efficiency in both examples is the square of the correlation between the side information V and the market X. The unifying answer for the initial efficiency is given by Hirschfeld–Gebelein–Rényi maximal correlation. The initial efficiency for the horse race market is the square of the maximal correlation between V and X. Hence even the most efficient description of V pays off as the square of the maximal correlation. The counterpart in source coding is that the maximum decrease in the descriptive complexity of X per bit of description of V is the square of the maximal correlation.

For the general market, the initial efficiency is 1 when side information $V = \boldsymbol{X}$. When the encoder can observe the stock market outcomes, initially every bit of description adds one bit to the growth rate, thereby doubling the wealth. A simple characterization of the initial efficiency for the general market for arbitrary side information V is not yet known.

References

[1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 533–542, 1986.

[2] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probab.*, vol. 4, pp. 925–939, 1976.

[3] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. für Wahrscheinlichkeitsth. verwandte Gebiete*, vol. 34, pp. 157–177, 1976.

[4] R. Ahlswede and J. Körner, "On the connection between the entropies of input and output distributions of discrete memoryless channels," in *Proc. 5th Conf. Probability Theory* (Brasov, Rumania, 1974). Bucharest, Rumania: Editura Acad. Rep. Soc. Romania, 1977, pp. 13–23.

[5] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, 1975.

[6] A. R. Barron and T. M. Cover, "A bound on the financial value of information," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1097–1100, 1988.

[7] R. Bell and T. M. Cover, "Game-theoretic optimal portfolios," *Manag. Sci.*, vol. 34, no. 6, pp. 724–733, June 1988.

[8] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[9] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 279–280, 1974.

[10] A. Buja, "Remarks on functional canonical variates, alternating least squares method and ACE," *Ann. Statist.*, vol. 18, pp. 1032–1069, 1990.

[11] N. Chayat and S. Shamai (Shitz), "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1077–1079, 1989.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[14] H. Gebelein, "Das statistische problem der Korrelation als variations—und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung," *Z. für angewandte Math. und Mech.*, vol. 21, pp. 364–379, 1941.

[15] H. O. Hirschfeld, "A connection between correlation and contingency," *Proc. Cambridge Philosophical Soc.*, vol. 31, pp. 520–524, 1935.

[16] J. Kelly, "A new interpretation of information rate," *Bell Syst. Tech. J.*, vol. 35, pp. 917–926, 1956.

[17] A. Rényi, "New version of the probabilistic generalization of the large sieve," *Acta Math. Hung.*, vol. 10, pp. 217–226, 1959.

[18] ——, "On measures of dependence," *Acta Math. Hung.*, vol. 10, pp. 441–451, 1959.

[19] H. S. Witsenhausen, "Entropy inequalities for discrete channels," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 610–616, 1974.

[20] ——, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, 1975.

[21] H. S. Witsenhausen and A. D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 493–501, 1975.

[22] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, 1975.

[23] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and its applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 769–772, 1973.