An Algorithm for Maximizing Expected Log Investment Return

THOMAS M. COVER, FELLOW, IEEE

Abstract—Let the random (stock market) vector $X \geq 0$ be drawn according to a known distribution function $F(x)$, $x \in R^m$. A log-optimal portfolio b^* is any portfolio b achieving maximal expected log return $W^* = \sup_b E \ln b^t X$, where the supremum is over the simplex $b \geq 0$, $\sum_{i=1}^m b_i = 1$. An algorithm is presented for finding b^*. The algorithm consists of replacing the portfolio b by the expected portfolio b', $b_i' = E(b_i X_i / b^t X)$, corresponding to the expected proportion of holdings in each stock after one market period. The improvement in $W(b)$ after each iteration is lower-bounded by the Kullback–Leibler information number $D(b' \| b)$ between the current and updated portfolios. Thus the algorithm monotonically improves the return W. An upper bound on W^* is given in terms of the current portfolio and the gradient, and the convergence of the algorithm is established.

I. Introduction

Let X_i denote the random capital return from the investment of one unit in the ith stock, $i = 1, 2, \cdots, m$. For example, if stock i is bought for 20 and sold for 30, then $X_i = 1.5$. The stock vector X is a nonnegative vector-valued random variable drawn according to a known distribution function $F(x)$, $x \in R^m$. A portfolio

$$b = (b_1, b_2, \cdots, b_m)^t, \qquad b_i \geq 0, \quad \sum b_i = 1,$$

is an allocation of investment capital over the stocks $X = (X_1, X_2, \cdots, X_m)^t$. The expected log return $W(b)$ and the maximal expected log return W^* are given by

$$W(b) = E \ln b^t X = \int \ln b^t x \, dF(x),$$
$$W^* = \max_b W(b). \tag{1.1}$$

We wish to determine the portfolio b^* (unique if the support set of X is of full dimension) that maximizes the expected log return $W(b)$. A discussion of the naturalness of this objective can be found in the series of papers by Williams [1], Kelly [2], Latané [3], Breiman [4], Thorp [5], [6], [7], Samuelson [8], Hakansson [9], [10], Bell and Cover [11], [12], and Arrow [13]. Briefly, money compounds multiplicatively rather than additively, hence the naturalness of maximizing $E \ln b^t X$ instead of $E b^t X$. Also, under b^*, money grows exponentially to infinity at the highest possible rate and achieves distant goals in least time ([2], [4]). Finally, b^* is the heart of the game-theoretic solution of the two-person zero-sum game in which one player desires to outperform another in a single investment with payoff $E\varphi(b_1^t X / b_2^t X)$, where φ is any given nondecreasing function ([11], [12]). Thus b^* has both long-run and short-run optimality properties.

The problem of maximizing $E \ln b^t X$ can be viewed as one of maximizing a concave function over the simplex $B = \{b \in R^m\colon\ b \geq 0,\ \sum b_i = 1\}$. Thus a maximizing b^* exists. Optimization algorithms abound for problems of this kind. For example, the paper by Ziemba [15] applies the Frank–Wolfe algorithm to the portfolio selection problem; a succession of one-dimensional slices of the simplex B are searched for ϵ-optimal portfolios. Algorithms for special stock distributions are presented in Ziemba [16], where X is multivariate normal, and in Ziemba [17], where the X is discrete valued. See also Dexter, Yu, and Ziemba [17].

Special properties of the maximization suggest the use of an algorithm specific to the problem. In particular, because of the logarithmic objective function, an algorithm that takes multiplicative rather than additive steps seems natural.

The gradient of $W(b)$, which we denote by $\alpha(b)$, is given by

$$\alpha(b) = EX / b^t X = \nabla W(b). \tag{1.2}$$

The Algorithm: Generate a sequence of portfolio vectors $b^n \in B$, recursively according to

$$b_i^{n+1} = b_i^n \alpha_i(b^n), \qquad i = 1, 2, \cdots, m,$$
$$b^0 > 0. \tag{1.3}$$

The spirit of this algorithm is very close to that exhibited in the algorithms of Arimoto [19], Blahut [20], and Csiszár [21]. Their algorithms solve for channel capacity and the rate distortion function by multiplicatively updating the probability mass function in much the same manner as the portfolio vector is updated in (1.3). Also, Csiszár and Tusnády have investigated the convergence of the algorithm presented above and, in an as yet unpublished work [22], will present an alternate proof of its convergence. It should be noted that when we ran this algorithm on actual stock market data, we used a variety of *ad hoc* techniques to accelerate its convergence. Theorem 4 of Section V then became the primary tool for terminating the computation.

Manuscript received April 20, 1983; revised September 5, 1983. This work was partially supported by NSF Grant ECS78-23334 and JSEP Contract DAAG29-79-C-0047. This paper was presented in part at the Information Theory Symposium, Budapest, Hungary, August 1981.

The author is with the Departments of Electrical Engineering and Statistics, Stanford University, Stanford, CA 94305.

0018-9448/84/0300-0369$01.00

The algorithm multiplies the current portfolio vector $\boldsymbol{b}$ by the gradient, component by component. It has the following natural interpretation. Let $\boldsymbol{b}$ be the current allocation of resources across the stocks. The random vector $\boldsymbol{X}$ results in current holdings in the ith stock $b_i X_i$ and yields a total return $\boldsymbol{b}^t\boldsymbol{X}$. Thus the new proportion of capital in the ith stock is given by $b_i X_i/\boldsymbol{b}^t\boldsymbol{X}$, and the expected proportion in the ith stock is

$$b_i' = E(b_i X_i/\boldsymbol{b}_t \boldsymbol{X}) = b_i \alpha_i(\boldsymbol{b}).$$

This is the new portfolio induced by the algorithm. One replaces the portfolio $\boldsymbol{b}$ by the *expected* portfolio $\boldsymbol{b}'$ induced by one play of the market $\boldsymbol{X}$. Naturally one expects that the algorithm terminates at $\boldsymbol{b}$ such that $\boldsymbol{b}' = \boldsymbol{b}$. This is proved in Theorem 3.

Remark: The sequence $\{\boldsymbol{b}^{n+1}\}$ remains in the simplex, because

$$\begin{aligned}\sum b_i^{n+1} &= \sum b_i^n \alpha_i(\boldsymbol{b}^n) = \sum b_i^n E\left(X_i/\sum b_j^n X_j\right)\\ &= E\left(\left(\sum b_i^n X_i\right)/\sum b_j^n X_j\right) = E1 = 1. \quad (1.4)\end{aligned}$$

Example: Consider two stocks, X_1 and X_2. Let $X_1 = 1$ represent cash. Let X_2 take on the values 2 and 1/2 with equal probability. Set $\boldsymbol{X} = (X_1, X_2)$, and consider portfolios $\boldsymbol{b} = (b_1, b_2) \in B$. We calculate

$$W(b) = \frac{1}{2}\ln(b_1 + 2b_2) + \frac{1}{2}\ln\left(b_1 + \frac{1}{2}b_2\right);$$

$$\boldsymbol{b}' = \left(\frac{1}{2}, \frac{1}{2}\right)^t; \quad W^* = \frac{1}{2}\ln(9/8);$$

$$\alpha_2(\boldsymbol{b}) = (5 - b_2)/2(1 + b_2)(2 - b_2).$$

Inspection of W^* indicates that repeated independent investments in $\boldsymbol{X}$ will yield capital growing to infinity exponentially like $(9/8)^{n/2}$. This, despite the fact that either stock alone results in a median return stuck at 1 for any n. Writing the initial portfolio as

$$\boldsymbol{b} = \left(\frac{1}{2} + \epsilon, \frac{1}{2} - \epsilon\right)^t,$$

and the next iterate as

$$\boldsymbol{b}' = \left(\frac{1}{2} + \epsilon', \frac{1}{2} - \epsilon'\right)^t,$$

we derive

$$\epsilon' = 8\epsilon/(9 - 4\epsilon^2) = 8\epsilon/9 + O(\epsilon^3).$$

Thus, for this example, $\boldsymbol{b}^n \to \boldsymbol{b}^*$ exponentially fast.

The initial vector $\boldsymbol{b}^0$ is chosen to be positive in each component. Otherwise $\boldsymbol{b}^n$ would be confined to the boundary of the simplex. Let

$$W_n = W(\boldsymbol{b}^n) = E\ln \boldsymbol{b}^{nt}\boldsymbol{X}, \tag{1.5}$$

be the value (expected log return) of the nth iterate of the portfolio. Define

$$W^* = \sup_{\boldsymbol{b}} E\ln \boldsymbol{b}^t\boldsymbol{X}. \tag{1.6}$$

and let $\boldsymbol{b}^*$ be any portfolio achieving W^*.

In the course of this paper, we shall show that W_n is a monotonically nondecreasing sequence satisfying

$$W_{n+1} - W_n \geq \sum b_i^{n+1}\ln\frac{b_i^{n+1}}{b_i^n} \geqslant 0, \tag{1.7}$$

and

$$W_n \uparrow W^*. \tag{1.8}$$

II. Monotonicity

We wish to prove that each step of the algorithm yields an improvement in the expected log return W.

Denote the next portfolio iterate by $\boldsymbol{b}'$, where

$$b_i' = b_i E\left(X_i/\boldsymbol{b}^t\boldsymbol{X}\right), \tag{2.1}$$

and define the random variables

$$Y_i(\boldsymbol{b}) = X_i/\boldsymbol{b}^t\boldsymbol{X}, \qquad i = 1, 2, \cdots, m. \tag{2.2}$$

Let,

$$D(\boldsymbol{b}'\|\boldsymbol{b}) = \sum_{i=1}^m b_i'\ln\left(b_i'/b_i\right), \tag{2.3}$$

be the Kullback–Leibler information number (or divergence or relative entropy) between $\boldsymbol{b}'$ and $\boldsymbol{b}$. See Kullback [14] for extensive interpretations of this definition. We shall need the inequality

$$D(\boldsymbol{b}'\|\boldsymbol{b}) \geqslant 0, \text{ with equality if and only if } \boldsymbol{b}' = \boldsymbol{b}, \tag{2.4}$$

a consequence of the strict concavity of the logarithm.

Theorem 1 (Monotonicity):

$$W(\boldsymbol{b}') - W(\boldsymbol{b}) \geq D(\boldsymbol{b}'\|\boldsymbol{b}) \geq 0. \tag{2.5}$$

Proof: We have

$$\begin{aligned}W(\boldsymbol{b}') - W(\boldsymbol{b}) &= E\ln\left(\sum b_i' X_i/\sum b_i X_i\right)\\ &= E\ln\sum b_i' Y_i(\boldsymbol{b})\\ &= E\ln\sum b_i\left(EY_i(\boldsymbol{b})\right)Y_i(\boldsymbol{b}).\end{aligned}$$

Jensen's inequality on the expectation works in the wrong direction for our needs. On the other hand, we note that the random variables $b_i Y_i(\boldsymbol{b})$ satisfy

$$b_i Y_i(\boldsymbol{b}) \geq 0, \quad \text{almost everywhere},$$

$$\sum_{i=1}^m b_i Y_i(\boldsymbol{b}) = \sum b_i X_i/\sum b_j X_j = 1, \quad \text{almost everywhere}. \tag{2.6}$$

Thus applying Jensen's inequality for these mixing variables yields

$$\begin{aligned}W(\boldsymbol{b}') - W(\boldsymbol{b}) &\geq E\sum b_i Y_i(\boldsymbol{b})\ln EY_i(\boldsymbol{b})\\ &= \sum b_i EY_i(\boldsymbol{b})\ln\left(EY_i(\boldsymbol{b})\right)b_i/b_i\\ &= \sum b_i'\ln\left(b_i'/b_i\right)\\ &= D(\boldsymbol{b}'\|\boldsymbol{b}) \geq 0, \qquad \text{for all } \boldsymbol{b} \in B.\end{aligned} \tag{2.7}$$

This proves the theorem. □

Corollary: Under the algorithm $b_i' = b_i\alpha_i(\boldsymbol{b})$, we have $W(\boldsymbol{b}') = W(\boldsymbol{b})$ if and only if

$$b_i\alpha_i = b_i, \qquad \text{for all } i,$$

i.e., if and only if $\alpha_i = 1$, for i such that $b_i > 0$.

In addition to the desired inequality $E\ln(S_{n+1}/S_n) \geq 0$, where $S_n = \boldsymbol{b}^{n t}\boldsymbol{X}$, we also have the following theorem. This theorem will not be required for the proof of convergence.

Theorem 2: Monotonicity of ratio

$$E(S_{n+1}/S_n) \geq 1. \tag{2.8}$$

Proof: By (2.1), (2.6), and Jensen's inequality,

$$\begin{aligned} E\frac{S_{n+1}}{S_n} &= E\left(\sum b_i' X_i\right)/\left(\sum b_j X_j\right) \\ &= E\sum b_i(EY_i(\boldsymbol{b}))Y_i(\boldsymbol{b}) = \sum b_i(EY_i)^2 \\ &\geq \left(\sum b_i EY_i\right)^2 = 1. \end{aligned}$$

III. Preliminary Lemmas

Let B denote the simplex $\{\boldsymbol{b} \in \boldsymbol{R}^m: b_i \geq 0, \Sigma b_i = 1\}$. As before, let $\alpha(\boldsymbol{b}) = \nabla W(\boldsymbol{b})$, $W^* = \sup_{\boldsymbol{b}\in B} W(\boldsymbol{b})$.

Recall the Kuhn–Tucker conditions for this maximization. (See [11], [12].) A portfolio $\boldsymbol{b}$ achieves W^* if and only if

$$\alpha_i(\boldsymbol{b}) = 1, \qquad b_i > 0 \tag{3.1}$$

$$\alpha_i(\boldsymbol{b}) \leq 1, \qquad b_i = 0. \tag{3.2}$$

We designate these as the first and second parts of the Kuhn–Tucker conditions.

Definition: Let $\tilde{B}$ denote the set of accumulation points of $\{\boldsymbol{b}^n\}$.

We recall that $\boldsymbol{b}^0 > \boldsymbol{0}$ and that $b_i^{n+1} = \alpha_i(\boldsymbol{b}^n)b_i^n$. The following lemma is used prominently in the proof of $W_n \to W^*$.

Lemma 1: The set of accumulation points $\tilde{B}$ of $\{\boldsymbol{b}^n\}$ is nonempty, compact, and connected.

Proof: Since $\boldsymbol{b}^n \in B$, and B is compact, there must exist an accumulation point, by the Bolzano–Weierstrass theorem. Thus $\tilde{B}$ is nonempty. The set of accumulation points of a sequence is always closed. Since $\tilde{B}$ is also a bounded subset of $\boldsymbol{R}^n$, $\tilde{B}$ is compact.

If $\tilde{B}$ were not connected, the closedness of $\tilde{B}$ implies that there is an open cutset separating two components of $\tilde{B}$. This cutset contains a compact subset C with a nonempty interior. But the components of $\tilde{B}$ must be traversed infinitely often. Thus C must be crossed infinitely often. However,

$$D(\boldsymbol{b}^{n+1}\|\boldsymbol{b}^n) \leq W_{n+1} - W_n \to 0. \tag{3.3}$$

Thus we have convergence of the step sizes to zero:

$$\|\boldsymbol{b}^{n+1} - \boldsymbol{b}^n\| \to 0. \tag{3.4}$$

Hence C is entered infinitely often by elements of $\{\boldsymbol{b}^n\}$. But C is compact. Thus C contains an accumulation point of $\{\boldsymbol{b}^n\}$, violating the construction $C \cap \tilde{B} = \varphi$. Consequently $\tilde{B}$ is connected. □

Finally, we shall need continuity properties of the gradient $\alpha(\boldsymbol{b})$.

Note that $W^* > -\infty$ implies that $P(\boldsymbol{X} = 0) = 0$, and thus that $\boldsymbol{X}/\boldsymbol{b}^t\boldsymbol{X}$ is well defined with probability one. We shall consider the components $\alpha_i(\boldsymbol{b})$ as extended real valued functions of $\boldsymbol{b}$, possibly taking the value $+\infty$. We observe that if $W^* < \infty$, then $\alpha_i(\boldsymbol{b})$ is finite for all $\boldsymbol{b}$ in the interior of the simplex B. Finally, we note that $\alpha_i(\boldsymbol{b}) \geq 0$, for all $\boldsymbol{b} \in B$.

Lemma 2: Let $-\infty < W^* < \infty$. Then the components of the gradient vector $\alpha(\boldsymbol{b})$ are continuous extended real-valued functions of $\boldsymbol{b} \in B$.

Proof: Let $\boldsymbol{b}_0$ be any point in B and consider any sequence $\boldsymbol{b}_k \to \boldsymbol{b}_0$. Thus by continuity of $X_i/\boldsymbol{b}^t\boldsymbol{X}$,

$$\lim_{k\to\infty} \frac{X_i}{\boldsymbol{b}_k^t\boldsymbol{X}} = \frac{X_i}{\boldsymbol{b}_0^t\boldsymbol{X}}, \quad \text{almost surely.} \tag{3.5}$$

So by Fatou's lemma,

$$\alpha_i(\boldsymbol{b}_0) = E\frac{X_i}{\boldsymbol{b}_0^t\boldsymbol{X}} \leq \liminf_{k\to\infty} \alpha_i(\boldsymbol{b}_k). \tag{3.6}$$

Consequently, if $\alpha_i(\boldsymbol{b}_0) = \infty$, then $\boldsymbol{b}_0$ is a point of continuity of the extended real-valued function $\alpha_i(\boldsymbol{b})$.

We now show that $\boldsymbol{b}_0$ is also a point of continuity if $\alpha_i(\boldsymbol{b}_0) < \infty$. If $\boldsymbol{b}_k \in B$, $\boldsymbol{b}_k \to \boldsymbol{b}_0$, then there exists an integer k_0 such that $b_{ki} \geq (\frac{1}{2})b_{0i}$, for $k \geq k_0$, for $i = 1, 2, \cdots, m$. Thus for $k \geq k_0$,

$$\frac{X_i}{\boldsymbol{b}_k^t\boldsymbol{X}} \leq \frac{X_i}{(\frac{1}{2})\boldsymbol{b}_0^t\boldsymbol{X}}. \tag{3.7}$$

But $\alpha_i(\boldsymbol{b}_0) < \infty$, so by dominated convergence

$$\alpha_i(\boldsymbol{b}_k) \to \alpha_i(\boldsymbol{b}_0). \tag{3.8}$$

□

IV. Convergence of Algorithm

Since we have shown that $W(\boldsymbol{b}^n)$ is monotonically nondecreasing and thus has a limit, it remains to be shown that the limit is W^*.

Theorem 3 (Convergence): If $\boldsymbol{b}^0 > 0$, then

$$W(\boldsymbol{b}^n) \uparrow W^*.$$

Moreover, if $\boldsymbol{X}$ has full dimension, then $\boldsymbol{b}^n \to \boldsymbol{b}^*$.

Proof: We shall say $\boldsymbol{b}$ is stable if $\boldsymbol{b}' = \boldsymbol{b}$, i.e., if $\boldsymbol{b}$ satisfies the first of the Kuhn–Tucker conditions. The proof breaks into 2 parts:

1) showing that any accumulation point $\tilde{\boldsymbol{b}}$ of $\{\boldsymbol{b}^n\}$ is stable, and
2) showing that any limit point $\tilde{\boldsymbol{b}}$ satisfies the second Kuhn–Tucker condition.

Parts 1) and 2) then imply that $W(\tilde{b}) = W^*$, proving the theorem.

Part 1: Define $\Delta(b^n) = W(b^{n+1}) - W(b^n)$. Now $W(b^n)$ is monotonically nondecreasing and therefore has a limit W_0. So $\Delta(b^n) \to 0$. Let $\tilde{b}$ be any accumulation point of $\{b^n\}$, and let b^{n_k} be a subsequence converging to $\tilde{b}$. Existence of $\tilde{b}$ is guaranteed by the Bolzano–Weierstrass theorem.

Now $\Delta(b)$ is a continuous extended real valued function of b, by Lemma 2 and the continuity of $W(b)$. Thus $\Delta(b^{n_k}) \to \Delta(\tilde{b})$. But $\Delta(b^{n_k}) \to 0$. Thus $\Delta(\tilde{b}) = 0$. But by Theorem 1, $\Delta(\tilde{b}) \geq \Sigma \tilde{b}_i' \ln \tilde{b}_i'/\tilde{b}_i \geq 0$, with equality if and only if $\tilde{b}_i' = \tilde{b}_i$. Thus $\tilde{b}$ must be stable.

Part 2: We can easily show that the second Kuhn–Tucker condition is satisfied if in fact b^n has a limit $\tilde{b}$, because then, by continuity of $\alpha(b)$, we would have

$$\alpha_i(b^n) \to \alpha_i(\tilde{b}) = \tilde{\alpha}_i. \tag{4.1}$$

But

$$b_i^n = b_i^0 \prod_{i=1}^{n} \alpha_i(b^k), \tag{4.2}$$

which would diverge to ∞ if $\tilde{\alpha}_i > 1$. This contradiction to $b^n \in B$, leads us to conclude that $\tilde{\alpha}_i \leq 1$, for all i, thus establishing the second half of the Kuhn–Tucker conditions.

The remainder of the argument will go roughly as follows. We can argue that an accumulation point $\tilde{b}$ of $\{b^n\}$ must maximize $W(b)$ over the face of the simplex in which b lies. Now if X were of full dimension, then such a maximizing $\tilde{b}$ would be unique for each face. But this set of possible accumulation points, one per face of B, cannot be connected unless it consists of a single point. The argument above would then apply, finishing the proof. However, if X is not of full dimension, then it is no longer true that the maximum of $W(b)$ is uniquely attained. But, by projecting the portfolios b onto the linear subspace spanned by the support of X, uniqueness can be established for the projections of the maximizing portfolios for each face. This proves that $\tilde{B}$ projects into a single point, and again the argument above can be applied.

We proceed directly to the general case. For each $I \subseteq \{1, 2, \cdots, m\}$, let the face B_I be defined by

$$B_I = \left\{ b \in B : b_i > 0, i \in I, b_i = 0, i \in I^C \right\}. \tag{4.3}$$

We now show that all accumulation points of $\{b^n\}$ in face B_I have the same orthogonal projection onto the subspace spanned by the support set of X.

Let L be the subspace of R^m of least dimension satisfying $P(X \in L) = 1$. Let $\hat{b}$ denote the orthogonal projection of b onto L. Thus

$$(b - \hat{b})^t X = 0, \quad \text{almost surely.} \tag{4.4}$$

Suppose now that $b_1, b_2 \in \tilde{B} \cap B_I$. Thus, by the stability of accumulation points, we have

$$\alpha_i(b_1) = \alpha_i(b_2) = 1, \qquad \text{for } i \in I. \tag{4.5}$$

Hence, by (4.3), (4.5), and Jensen's inequality,

$$\begin{aligned} W(b_1) - W(b_2) &= E \ln \frac{b_1^t X}{b_2^t X} \\ &\leq \ln \sum_i b_{1i} \alpha_i(b_2) = \ln \sum_{i \in I} b_{1i} \alpha_i(b_2) \\ &= \ln 1 = 0, \end{aligned} \tag{4.6}$$

with equality if and only if

$$\frac{b_1^t X}{b_2^t X} = 1, \quad \text{almost surely.} \tag{4.7}$$

Reversing the roles of b_1 and b_2 then yields

$$W(b_2) - W(b_1) \leq 0, \tag{4.8}$$

which together with (4.6) yields

$$W(b_1) = W(b_2). \tag{4.9}$$

Closer inspection of the above argument reveals that b_1 and b_2 satisfy the Kuhn–Tucker conditions for face B_I and thus that W is indeed maximal for this face.

Since we have shown that equality holds in (4.6), it follows that (4.7) must hold. But from (4.4) and (4.7),

$$\frac{\hat{b}_1^t X}{\hat{b}_2^t X} = \frac{b_1^t X}{b_2^t X} = 1, \quad \text{almost surely.} \tag{4.10}$$

Thus

$$(\hat{b}_1 - \hat{b}_2)^t X = 0, \quad \text{almost surely.} \tag{4.11}$$

But $\hat{b}_1 - \hat{b}_2 \in L$, so, by the minimality of the dimension of L, we must have $\hat{b}_1 = \hat{b}_2$. This is the desired result. All accumulation points of $\{b^n\}$ that fall in face B_I have the same orthogonal projection onto L.

At this point, we realize that each face B_I of the simplex B generates at most one projected accumulation point in L. Moreover, $\cup B_I = B$, and there are precisely 2^m faces B_I partitioning B. Thus there are at most 2^m points in L corresponding to the projections of the accumulation points $\tilde{B}$ of $\{b^n\}$. By Lemma 1, $\tilde{B}$ is connected and hence its projection onto L must be connected. However, no finite nonempty set of points forms a connected set unless it consists of a single point. Thus $\tilde{B}$ projects onto a single point in L, which we designate by $\hat{b}_L$.

We now observe that, for all $b \in B$,

$$\alpha_i(b) = E \frac{X_i}{b^t X} = E \frac{X_i}{\hat{b}^t X} = \alpha_i(\hat{b}). \tag{4.12}$$

Thus $\alpha_i(b^k) = \alpha_i(\hat{b}^k) \to \alpha_i(\hat{b}_L)$, $i = 1, 2, \cdots, m$. Consequently,

$$\begin{aligned} b_i^n &= b_i^0 \prod_{i=1}^{n} \alpha_i(b^k) \\ &= b_i^0 \prod_{i=1}^{n} \alpha_i(\hat{b}^k), \end{aligned} \tag{4.13}$$

diverges to infinity unless

$$\alpha_i(\hat{\boldsymbol{b}}_L) \le 1, \qquad \text{for all } i. \tag{4.14}$$

This establishes the second half of the Kuhn–Tucker conditions for $\hat{\boldsymbol{b}}_L$ and thus for any accumulation point $\tilde{\boldsymbol{b}} \in \tilde{B}$.

We have now shown that any accumulation point $\tilde{\boldsymbol{b}}$ of $\{\boldsymbol{b}^n\}$ satisfies both halves of the Kuhn–Tucker conditions and is therefore optimal. Since $W(\boldsymbol{b}^n)$ is nondecreasing, we have $W(\boldsymbol{b}^n) \to W^*$, as desired. □

V. A Sequence of Upper Bounds on W^*

We now establish upper bounds on the error of approximation of $W(\boldsymbol{b}^n)$ to W^*. The bound is a function of the portfolio $\boldsymbol{b}$ and not of the algorithm used to guess $\boldsymbol{b}$.

Theorem 4 (Upper Bound): For any $\boldsymbol{b} \in B$,

$$W(\boldsymbol{b}) \le W^* \le W(\boldsymbol{b}) + \max_i \ln E\frac{X_i}{\boldsymbol{b}^t\boldsymbol{X}}. \tag{5.1}$$

Consequently, the algorithm (1.3) yields

$$W_n \le W^* \le W_n + \max_{i=1,2,\cdots,m} \{\ln \alpha_i^n\}. \tag{5.2}$$

Proof: The lower bound in (5.1) follows by the optimality of W^*. Let $S^* = \boldsymbol{b}^{*t}\boldsymbol{X}$ and $S(\boldsymbol{b}) = \boldsymbol{b}^t\boldsymbol{X}$. The upper bound follows from application of Jensen's inequality:

$$\begin{aligned} E\ln(S^*/S(\boldsymbol{b})) &\le \ln E(S^*/S(\boldsymbol{b})) \\ &= \ln \sum_i b_i^* E(X_i/S(\boldsymbol{b})) \\ &\le \ln \max_i E(X_i/S(\boldsymbol{b})). \end{aligned} \tag{5.3}$$

Thus $E\ln S^* \le E\ln S(\boldsymbol{b}) + \max_i \ln E(X_i/S(\boldsymbol{b}))$, as desired. □

Remark: Note that the Kuhn–Tucker conditions require that the term $\max \ln EX_i/S(\boldsymbol{b})$ be ≤ 0 for $W(\boldsymbol{b}) = W^*$. Thus the upper bound converges to W^* as $n \to \infty$.

Conclusion

We have shown that if we start the algorithm with $\boldsymbol{b}^0 > \boldsymbol{0}$, then $W_{n+1} \uparrow W^*$. Moreover, if there is a unique optimal portfolio $\boldsymbol{b}^*$, then $\boldsymbol{b}^n \to \boldsymbol{b}^*$. The effective computation of good portfolios is made feasible by Theorem 4, which enables one to stop the computation with an ϵ-optimal portfolio as soon as $\ln \alpha_i(\boldsymbol{b}^n) \le \epsilon$, $i = 1, 2, \cdots, m$.

Acknowledgment

The author would like to acknowledge the very helpful discussions with David Gluss, Max Costa, and David Larson concerning the proof of Theorem 3.

References

[1] J. Williams, "Speculation and the carryover," *Quart. J. Econom.*, vol. 50, pp. 436–455, 1936.

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

[3] H. Latané, "Criteria for choice among risk ventures," *J. Politic. Economy*, vol. 67, pp. 144–155, 1959.

[4] L. Breiman, "Optimal gambling systems for favorable games," in *Proc. Fourth Berkeley Symp.*, 1961, pp. 65–78.

[5] E. Thorp, "Optimal gambling systems for favorable games," *Rev. Internat. Statist.*, vol. 37, pp. 273–293, 1969.

[6] ——, "Portfolio choice and the Kelly criterion," *Business Econom. Statist. Proc. Amer. Stat. Assoc.*, pp. 215–224, 1971.

[7] J. Bicksler and E. Thorp, "The capital growth model: An empirical investigation," *J. Financial and Quantitative Anal.*, vol. VIII, pp. 273–287, 1973.

[8] P. Samuelson, "General proof that diversification pays," *J. Financial and Quantitative Anal.*, vol. II, pp. 1–13, 1967.

[9] N. Hakansson and T. Liu, "Optimal growth portfolios when yields are serially correlated," *Rev. Econom. Statist.*, pp. 385–394, 1970.

[10] N. Hakansson, "Capital growth and the mean-variance approach to portfolio selection," *J. Financial and Quantitative Anal.*, vol. VI, pp. 517–557, 1971.

[11] R. Bell and T. Cover, "Competitive optimality of logarithmic investment," *Math. of O.R.*, vol. 5, no. 2, pp. 161–166, May 1980.

[12] ——, "Competitive portfolio selection," in preparation.

[13] K. Arrow, *Essays in Theory of Risk Bearing.* Amsterdam: North Holland, 1974.

[14] S. Kullback, *Information Theory and Statistics.* New York: Dover, 1959.

[15] W. Ziemba, "An expected Frank–Wolfe algorithm with application to portfolio selection problems," *Recent Results in Stochastic Programming*, Kall and Prekopa, Eds. New York: Springer-Verlag, 1980.

[16] ——, "On the robustness of the Arrow–Pratt risk aversion measure," *Econom. Lett.*, vol. 2, pp. 21–26, 1979.

[17] ——, "Note on optimal growth portfolios when yields are serially correlated," *J. Financial and Quantitative Anal.*, vol. VII, pp. 1995–2000, Sept. 1972.

[18] A. S. Dexter, J. N. W. Yu, and W. Ziemba, "Portfolio selection in a lognormal market when the investor has a power utility function: Computational results," in *Stochastic Programming*, M.A.H. Dempster, Ed. New York: Academic, 1980.

[19] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, Jan. 1972.

[20] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 4, July 1972.

[21] I. Csiszár, "On the computation of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 1, pp. 122–125, Jan. 1974.

[22] I. Csiszár and G. Tusnady, "Information geometry and alternating minimization procedures," preprint, Math. Inst. Hungarian Academy of Sciences, 1982.