An Adaptive Algorithm for log-optimal Portfolio and its Theoretical Analysis*

HUANG JIANGGUO YE ZHONGXING

(Department of Applied Mathematics, Shanghai Jiao Tong University, Shanghai, 200240)

Abstract

In this paper, an effectively adaptive algorithm for solving log-optimal portfolio problem is proposed. It is a variant type of stochastic approximation algorithm. Since the problem considered here is a constrained optimization problem, the gradient ascent direction used conventionally is replaced by the steepest ascent tangent vector on the corresponding constraint manifold. Under some reasonable conditions, the convergence property of this algorithm is also demonstrated. Finally, this algorithm is applied to search optimal portfolio with real data of the Exchange Institute of Shanghai Security, the obtained numerical results are satisfactory.

Keywords: log-optimal portfolio, stochastic approximation, adaptive algorithm.

AMS Subject Classification: 91B28, 62M99.

§1. Introduction

The purpose of this paper is to provide an effectively adaptive numerical algorithm for solving the log-optimal portfolio problem arising from the field of financial investment and then to consider the theoretical analysis of this algorithm.

Suppose an investor wishes to invest in the stock market which consists of m stocks. We denote the market by vector $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_m)^T$, $\boldsymbol{X}_i \geq 0, i = 1, 2, \cdots, m$, where $\boldsymbol{X}_i$ is the price relative of stock i which represents the ratio of the price at the end of the day to the price at the beginning of the day. We denote by $F(\boldsymbol{x})$ the joint distribution of the vector of price relatives $\boldsymbol{X}$.

The investor distributes his money among these m stocks according to the portfolio $\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^T$, where $b_i \geq 0$, $\sum\limits_{i=1}^{m} b_i = 1$. Here, b_i denotes the fraction of one's wealth invested in stock i. Thus, the relative wealth of the investor after one day is given by $S = \boldsymbol{b}^T\boldsymbol{X}$.

Furthermore, assume that this investor is interested in the long-term investment on these stocks. Without loss of generality, suppose that he has initial wealth of investment 1 and chooses the portfolio $\boldsymbol{b}(t)$ as the investment strategy at the tth day. Thus, after n consecutive days, his wealth becomes $S_n = \prod\limits_{t=1}^{n} \boldsymbol{b}^T(t)\boldsymbol{X}(t)$, where $\boldsymbol{X}(t)$ stands for the stock gain vector at the tth day. The aim of the optimal investment is then to find the optimal portfolio sequence $\{\boldsymbol{b}(t)\}$ in order to maximize the wealth gain S_n.

Suppose that the sequence $\{\boldsymbol{X}(t)\}$ of stock gain vectors are statistically independent and identically distributed (as usual, called i.i.d. for simplicity in what follows), according to the distribution $F(\boldsymbol{x})$. It was shown in [6] that, in the sense of asymptotic optimality, the $\{\boldsymbol{b}(t)\}$ can be chosen a constant vector $\boldsymbol{b}^*$ independent of time t, which maximizes the doubling rate,

$$W(\boldsymbol{b}, F) = \mathsf{E}[\log(\boldsymbol{b}^t\boldsymbol{X})] = \int \log(\boldsymbol{b}^t\boldsymbol{x})dF(\boldsymbol{x}), \tag{1}$$

*This work was supported by the National Natural Science Foundation of China under the grant No. 79790130 and grant No. 19901008.

Received 1999.11.10.

i.e.,

$$b^* = \arg\max_{b\in B} W(b, F), \tag{2}$$

where $B \equiv \left\{b : \sum_{i=1}^{m} b_i = 1,\ b_i \ge 0\right\}$.

In other words, let $X(1), X(2), \cdots, X(n)$ be a sequence of i.i.d. stock gain vectors drawn according to $F(x)$, and $S_n^* = \prod_{t=1}^{n} b^{*T} X(t)$, where b^* is the log-optimal portfolio in the sense of (2). Let $S_n = \prod_{t=1}^{n} b^T(t) X(t)$ be the wealth resulting from any other sequence of causal portfolio $\{b(t)\}$. Then it was shown in [6] that

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \le 0 \quad \text{with probability 1.}$$

Furthermore, for a practical investment problem, given the corresponding sequence $\{x(t)\}$ of stock gain vectors (sample vectors), we are interested in the dervation of the log-optimal portfolio b^*. In general, no closed form for b^* exists, thus we have to obtain it numerically. One way consisting in first approximating the doubling rate (1) with the sample average, i.e., $\frac{1}{n} \sum_{t=1}^{n} \log(b^T x(t))$, and then solving this problem by conventional or genetic type optimization algorithms, people are referred to [14], [15] for more detail.

This is a batch way but not adaptive (on line) method. That means, if we are known more data $\{x(t)\}_{t=n+1}^{n_1}$, we have to solve the problem again in the same way and can not borrow the known result concerning the data $\{x(t)\}_{t=1}^{n}$ sufficiently.

In this paper, we first propose an effectively adaptive algorithm for solving the problem (1) and (2). This algorithm is a stochastic approximation algorithm [11] - a stochastic gradient ascent method. Since the problem considered here is a constrained optimization problem, the gradient ascent direction used conventionally is replaced by the steepest ascent tangent vector on the corresponding constraint manifold. Moreover, in order to obtain the corresponding tangent vector more easily, we suggest a quadratic parameter transformation, to change the original constraint manifold, a $(m-1)$-dimensional simplex, $B = \left\{b : \sum_{i=1}^{m} b_i = 1,\ b_i \ge 0\right\}$, into a $(m-1)$-dimensional smooth unit surface, $S^{m-1} = \left\{w = (w_1, w_2, \cdots, w_m)^T : \sum_{i=1}^{m} w_i^2 = 1\right\}$. The derivation of this algorithm is inspired heavily by the related ideas in [3], [4]. Secondly, under some reasonable conditions, we demonstrate the convergence of this algorithm. The crucial step is to show that the corresponding stochastic sequence is upper bounded and has a positive lower bound. Finally, we apply this algorithm to solve a practical stock investment problem. We choose 20 stocks operated in the Exchange Institute of Shanghai Security, with their whole stock gain vectors in 1997 given. We obtain the related log-optimal portfolio by this algorithm. Those numerical results are satisfactory.

The rest of this paper is organized as follows. We describe the algorithm in section 2 and give the proof of convergence for this algorithm in section 3. Finally we provide the numerical simulation results in section 4.

§2. Algorithm descriptions

Since the objective function of problem (1) and (2) is described in regression form, it is natural to construct the stochastic approximation algorithm, an iterative method of stochastic gradient type, to solve it [11]. On the other hand, it is a constrained optimization problem, we should choose the desired ascent direction which is not only the fastestbut also in the related tangent space of the constraint manifold so as to keep the constrained condition invariant in infinitesimal displacement sense. This idea was used by Brockett ([3], [4]) to deal with least square matching problems and dynamical systems that sort lists, diagonalize matrices and solve linear programming problems, though the corresponding objective functions considered there are deterministic, not in regression form.

In fact, this direction is just the intrinsic gradient associated with the related constraint manifold. Thus, to obtain the desired adaptive algorithm for solving (1) and (2), we should derive the intrinsic gradient concerning this problem. For this purpose, we first give some definitions and a useful result which provides us a standard way to compute the gradient corresponding to a manifold M from the bases of M's ambient manifold N (i.e., M is the submanifold of N). We will use the standard conventions concerning Riemannian geometry in what follows, people are referred to [13] for more complete detail.

Let (M, g) denote a n-dimensional Riemannian manifold with the metric $g = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j$, where $\{dx^i\}$ denote the natural bases of the 1-form (cotangent space) T^*M. That means that if we denote by $\{\partial_i\}$ the related natural bases of the tangent space TM, which are dual with the bases of $\{dx^i\}$, then for any two tangent vector fields $X = \sum_{i=1}^{n} a_i \partial_i$, $Y = \sum_{i=1}^{n} b_i \partial_i$,

$$g(X, Y) = \sum_{i,j=1}^{n} a_i b_j g_{ij}.$$

Moreover, as usual [13], for a function $f \in C^1(M)$, the related (intrinsic) gradient corresponding to manifold M is defined by

$$\text{grad}_M f = \sum_{i,j=1}^{n} (g^{ij} \partial_j f) \partial_i, \tag{3}$$

where g^{ij} is the (i,j)-entry of the matrix $[g^{rs}]_{1 \le r,s \le n}$ which is the inverse matrix of the metric matrix $[g_{rs}]_{1 \le r,s \le n}$. Or alternatively,

$$\begin{aligned} &\text{grad}_M f \in TM, \\ &X(f) = g(\text{grad}_M f, X), \qquad \forall\, X \in TM. \end{aligned} \tag{4}$$

Lemma 1 Let (N, g) denote a n-dimensional Riemannian manifold with the metric g, M is a m-dimensional induced Riemannian submanifold of N. Moreover, Let $\{\tau_i\}_{i=1}^{n}$ be the moving frame of N, i.e., the local orthonormal vector fields, on N. Suppose that $\{\tau_i\}_{i=1}^{m}$ also forms the moving frame on M, i.e., $\{\tau_i\}_{i=1}^{m}$ are the orthonormal vector fields on M. Then we have

$$\text{grad}_M f = \text{grad}_N f - \sum_{i=m+1}^{n} \tau_i(f) \tau_i. \tag{5}$$

Proof From the definition (4) we have

$$g(\text{grad}_M f, X) = g(\text{grad}_N f, X), \qquad \forall\, X \in TM.$$

This means that there exist some constants $\alpha_{m+1}, \alpha_{m+1}, \cdots, \alpha_n$, such that,

$$\text{grad}_N f - \text{grad}_M f = \sum_{i=m+1}^{n} \alpha_i \tau_i.$$

Hence, from the orthonormal property of $\{\tau_i\}_{i=1}^{n}$, we know

$$\alpha_i = g(\text{grad}_N f - \text{grad}_M f, \tau_i) = \tau_i(f) - g(\text{grad}_M f, \tau_i) = \tau_i(f).$$

The desired result then follows. □

Suppose that we want to solve the following constrained optimization problem

$$w^* = \arg\max_{w \in M} \mathsf{E}[f(w, X)], \tag{6}$$

where $f(\boldsymbol{w}, \boldsymbol{x})$ is a given scalar-valued function, $\mathsf{E}[\cdot]$ denotes the mathematical expectation associated with some unknown distribution density function $p(\boldsymbol{x})$, $M \subset R^n$ is a submanifold of the Euclidean space R^n. Then the variant type of stochastic approximation method for solving (6) can be written as

$$\Delta \boldsymbol{w}^{(t)} = \eta(t)\mathrm{grad}_M f(\boldsymbol{w}^{(t)}, \boldsymbol{x}^{(t)}), \qquad t = 0, 1, \cdots, \tag{7}$$

where grad_M denotes the intrinsic gradient related to the manifold M, $\eta(t)$ is the step size (also called learning rate in general) at the tth iteration step, $\boldsymbol{x}^{(t)}$ is the tth sample drawn from the distribution $p(\boldsymbol{x})$.

Compared to the ordinary stochastic approximation method used in unconstrained optimization problems [11], the unique variant is that here, the chosen ascent direction is changed from ordinary gradient direction corresponding to R^n into the intrinsic gradient direction corresponding to the constraint manifold M.

Now we proceed to derive the gradient corresponding to the problem (1) and (2). At this time, the related constraint manifold $\boldsymbol{B} = \left\{\boldsymbol{b} : \sum\limits_{i=1}^{m} b_i = 1,\ b_i \geq 0\right\}$ is a irregular manifold with boundary. In order to avoid unnecessary complexity, we first make a quadratic transformation $b_i = w_i^2$, $i = 1, 2, \cdots, m$, and then the constraint manifold becomes

$$V = S^{m-1} = \left\{\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^T : \sum_{i=1}^{m} w_i^2 = 1\right\}, \tag{8}$$

which is a smooth compact manifold without boundary. And the problem (1) and (2) read as follows.

$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in V} \mathsf{E}[l(\boldsymbol{w}, \boldsymbol{X})], \tag{9}$$

where $l(\boldsymbol{w}, \boldsymbol{x}) = \log\left(\sum\limits_{i=1}^{m} w_i^2 x_i\right)$.

We can view the manifold V as an induced submanifold of the Euclidean space R^m, a flat manifold. And the unique normal vector field for V in R^m is just $r(\boldsymbol{w}) \equiv \boldsymbol{w}$, $\boldsymbol{w} \in V$. (Here and hereafter, we only list the corresponding coordinate vector to denote a vector field)

Furthermore, it is easy to know

$$\mathrm{grad}_{R^m} l(\boldsymbol{w}, \boldsymbol{x}) = \left(\frac{2x_1 w_1}{\sum\limits_{i=1}^{m} w_i^2 x_i}, \frac{2x_2 w_2}{\sum\limits_{i=1}^{m} w_i^2 x_i}, \cdots, \frac{2x_m w_m}{\sum\limits_{i=1}^{m} w_i^2 x_i}\right)^T,$$

$$r(\boldsymbol{w})(l(\boldsymbol{w}, \boldsymbol{x})) = (r(\boldsymbol{w}), \mathrm{grad}_{R^m} l(\boldsymbol{w}, \boldsymbol{x})) = 2\boldsymbol{w}, \qquad \boldsymbol{w} \in V,$$

where $r(\boldsymbol{w})(l(\boldsymbol{w}, \boldsymbol{x}))$ denotes the direction derivative of $l(\boldsymbol{w}, \boldsymbol{x})$ according to the direction $r(\boldsymbol{w})$.

Then, from Lemma 1 we have

$$\mathrm{grad}_V l(\boldsymbol{w}, \boldsymbol{x}) = \mathrm{grad}_{R^m} l(\boldsymbol{w}, \boldsymbol{x}) - 2\boldsymbol{w}.$$

Now, according to the general algorithm (7), we finally obtain an adaptive algorithm for searching $\boldsymbol{w}^*$.

Algorithm P:

Let $\{\boldsymbol{x}(t)\}$ denote the i.i.d. sequence of stock gain vectors. Then the optimal solution $\boldsymbol{w}^*$ related to (9) is obtained iteratively in the following way.

Step 0 Give any initial guess $\boldsymbol{w}^{(0)} \in V$, $t = 0$;

Step 1 compute the modified results $\boldsymbol{w}^{(t+1)} \equiv (w_1^{(t+1)}, w_2^{(t+1)}, \cdots, w_m^{(t+1)})^T$,

$$w_i^{(t+1)} = w_i^{(t)} + \eta(t)\left[\frac{x_i(t) w_i^{(t)}}{\sum\limits_{j=1}^{m} (w_j^{(t)})^2 x_j(t)} - w_i^{(t)}\right], \qquad i = 1, 2, \cdots, m, \tag{10}$$

where $\eta(t)$ is some chosen positive step size, called learning rate in general, it should satisfies some conditions given below;

Step 2 halt if the iteration number of times t reaches some given positive integer or if the norm of the difference $\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}$ is small than some given control precision; otherwise, $t \leftarrow (t+1)$, go to Step 1.

Similar to the general stochastic approximation algorithms, in order to make the algorithm convergent, we assume the learning rate sequence $\{\eta(t)\}$ satisfies that

$$\eta(t) > 0, \quad \sum \eta(t) = +\infty, \qquad \eta(t) \to 0 \text{ as } t \to +\infty. \tag{11}$$

In practice, if we only know the stock gain vectors, say $\boldsymbol{x}(1), \boldsymbol{x}(2), \cdots, \boldsymbol{x}(n)$, in some time period, we can choose the $\boldsymbol{x}(t)$ used in (10) from $\boldsymbol{x}(1), \boldsymbol{x}(2), \cdots, \boldsymbol{x}(n)$ with equal probability.

§3. Convergence analysis of Algorithm P

In order to analyze the convergence property of the **Algorithm P**, we first give a reasonable assumption concerning the sequence $\{\boldsymbol{x}(t)\}$ of stock gain vectors. That is,

A1: The sequence $\{\boldsymbol{x}(t)\}$ of stock gain vectors used in **Algorithm P** have an upper bound B and have a positive lower bound A; in other words,

$$0 < A \le x_i(t) \le B < +\infty, \qquad i = 1, 2, \cdots, m, \quad t = 1, 2, \cdots, \text{ a.s..}$$

For their probability distribution $F(x)$, this means

$$\int_{[A,B]^m} dF(x) = 1.$$

Under this assumption, we then have an important result concerning the iteration sequence $\{\boldsymbol{w}^{(t)}\}$ given in (10).

Lemma 2 For any $x \in \left[0, \frac{1}{2}\right]$,

$$1 - x \ge e^{-3x}.$$

Proof As a matter of fact, let $f(x) = (1-x)e^{3x}$. Then $f'(x) = e^{3x}(2-3x) > 0$, and the minimal value is attained at $x = 0$ with value 1. The desired inequality then follows easily. □

Theorem 1 For the **Algorithm P**, assume that the learning rate sequence $\{\eta(t)\}$ take values in $\left(0, \frac{1}{2}\right)$ and satisfy the condition (11), and the sequence $\{\boldsymbol{x}(t)\}$ of stock gain vectors satisfies assumption **A1**. Then, for any given nonzero initial vector $\boldsymbol{w}^{(0)}$, there exist two positive constants A_1 and B_1, such that, for the iteration sequence $\{\boldsymbol{w}^{(t)}\}$ given in (10), we have

$$A_1 \le \|\boldsymbol{w}^{(t)}\| \le B_1, \qquad t = 1, 2, \cdots, \text{ a.s.,} \tag{12}$$

where $\|\boldsymbol{w}^{(t)}\|$ denotes the standard Euclidean norm of the vector $\boldsymbol{w}^{(t)}$.

Proof We first verify the left hand side inequality of (12). For each i, we multiply both sides of (10) by $w_i^{(t)}$, and take the summation over i to obtain

$$\sum_{i=1}^{m} w_i^{(t)} w_i^{(t+1)} = (1 - \eta(t)) \sum_{i=1}^{m} (w_i^{(t)})^2 + \eta(t).$$

Therefore, by the Cauchy inequality

$$w_i^{(t)} w_i^{(t+1)} \le \frac{1}{2}\left[(1 - \eta(t))(w_i^{(t)})^2 + \frac{1}{1 - \eta(t)}(w_i^{(t+1)})^2\right],$$

we then have

$$(1 - \eta(t)) \sum_{i=1}^{m} (w_i^{(t)})^2 + \eta(t) \le \frac{1}{2}(1 - \eta(t)) \sum_{i=1}^{m} (w_i^{(t)})^2 + \frac{1}{2(1 - \eta(t))} \sum_{i=1}^{m} (w_i^{(t+1)})^2,$$

i.e.,

$$(1-\eta(t))^2\delta^{(t)} + 2\eta(t)(1-\eta(t)) \le \delta^{(t+1)}, \tag{13}$$

where, $\delta^{(t)} \equiv \sum_{i=1}^{m}(w_i^{(t)})^2$, for simplicity of presentation.

Thus, by induction, from the recursive inequality (13), we also have

$$\begin{aligned}\delta^{(t+1)} &\ge \prod_{i=0}^{t}(1-\eta(t))^2\delta^{(0)} + \eta(t-1) + \sum_{i=1}^{t}\eta(i-1)\prod_{j=i}^{t}(1-\eta(j))^2 \\ &\ge \sum_{i=1}^{t}\eta(i-1)\prod_{j=i}^{t}(1-\eta(j))^2 \equiv I(t).\end{aligned} \tag{14}$$

With the help of Lemma 2, from (14) we know that for $t \ge 2$,

$$I(t) \ge \sum_{i=1}^{t}\eta(i-1)e^{-6\sum_{j=i}^{t}\eta(j)} \ge e^{-6\sum_{j=0}^{t}\eta(j)}\sum_{i=2}^{t}\eta(i-1)e^{6\sum_{j=0}^{i-1}\eta(j)}. \tag{15}$$

On the other hand,

$$\eta(i-1)e^{6\sum_{j=0}^{i-1}\eta(j)} \ge \int_{\sum_{j=0}^{i-2}\eta(j)}^{\sum_{j=0}^{i-1}\eta(j)} e^{6x}dx.$$

Hence,

$$\sum_{i=2}^{t}\eta(i-1)e^{6\sum_{j=0}^{i-1}\eta(j)} \ge \int_{\eta(0)}^{\sum_{j=0}^{t-1}\eta(j)} e^{6x}dx = \frac{1}{6}(e^{6\sum_{j=0}^{t-1}\eta(j)} - e^{6\eta(0)}),$$

which together with (15) yields

$$I(t) \ge \frac{1}{6}[e^{6(\eta(0)-\eta(t))} - e^{-6\sum_{j=1}^{t}\eta(j)}].$$

It follows from the condition (11) that

$$\lim_{t\to+\infty}(e^{6(\eta(0)-\eta(t))} - e^{-6\sum_{j=1}^{t}\eta(j)}) = e^{6\eta(0)}.$$

Consequently, there exists some positive integer $t_0 > 2$ such that

$$I(t) \ge \frac{1}{12}e^{6\eta(0)} > 0 \qquad \forall\, t \ge t_0. \tag{16}$$

Combining (14) and (16) we have that for any nonnegative integer t,

$$\delta^{(t+1)} \ge \min\left\{\eta(0), \eta(1), \cdots, \eta(t_0), \frac{1}{12}e^{6\eta(0)}\right\} > 0.$$

The left hand side of the inequality (12) is then proved.

Now, let us proceed to demonstrate the right hand side inequality of (12). For each i, we take the squares at both sides of (10), and sum them together, we have

$$\sum_{i=1}^{m}(w_i^{(t+1)})^2 = (1-\eta(t))^2\sum_{i=1}^{m}(w_i^{(t)})^2 + 2\eta(t)(1-\eta(t)) + \eta(t)^2\frac{\sum_{i=1}^{m}(x_i(t))^2(w_i^{(t)})^2}{\left(\sum_{i=1}^{m}x_i(t)(w_i^{(t)})^2\right)^2}.$$

Then, from assumption **A1** and the left hand side of the inequality (12) just proved, we know

$$\begin{aligned}\delta^{(t+1)} &\le (1-\eta(t))^2\delta^{(t)} + 2\eta(t)(1-\eta(t)) + \frac{B}{AA_1}\eta(t)^2 \\ &\le (1-\eta(t))^2\delta^{(t)} + \left(2+\frac{B}{AA_1}\right)\eta(t),. \qquad \text{a.s.,}\end{aligned}$$

which leads to

$$\delta^{(t+1)} \leq \prod_{i=0}^{t}(1-\eta(i))^2\delta^{(0)} + \Big(2+\frac{B}{AA_1}\Big)\eta(t-1) + \sum_{i=1}^{t}\Big(2+\frac{B}{AA_1}\Big)\eta(i-1)\prod_{j=i}^{t}(1-\eta(j))^2 \qquad \text{a.s..} \tag{17}$$

It follows from the fact $\lim_{t\to+\infty}\prod_{i=0}^{t}(1-\eta(t))^2 = 0$ that there exists some positive constant C_1 such that

$$\prod_{i=0}^{t}(1-\eta(i))^2\delta^{(0)} \leq C_1. \tag{18}$$

In addition, defining $\overline{I}(t) \equiv \sum_{i=1}^{t}\eta(i-1)\prod_{j=i}^{t}(1-\eta(j))^2$, with the help of the conventional inequality $1-x \leq e^{(-x)}$ for $x \in (0,1)$, we see immediately that, for $t \geq 1$,

$$\overline{I}(t) \leq \sum_{i=1}^{t}\eta(i-1)e^{-2\sum_{j=i}^{t}\eta(j)}. \tag{19}$$

To forward our estimation, we now introduce an index sequence $\{n_k\}_{k=0}^{+\infty}$ defined by

$$\begin{cases} n_0 = \min\Big\{n : \sum_{i=0}^{n}\eta(i) \geq 1\Big\}, \\ n_{k+1} = \min\Big\{n : n > n_k,\ \sum_{n_k+1}^{n}\eta(i) \geq 1\Big\}. \end{cases} \tag{20}$$

We also denote by $m(t)$ the minimal positive intege1 such that $n_{m(t)} \geq t$. Thus, it follow from (19) and the definitions of $\{n_k\}_{k=0}^{+\infty}$ that

$$\begin{aligned}
\overline{I}(t) &\leq \sum_{i=1}^{m(t)}\eta(i-1)e^{-2\sum_{j=i}^{t}\eta(j)} \\
&= \sum_{i=1}^{n_0}\eta(i-1)e^{-2\sum_{j=i}^{t}\eta(j)} + \sum_{k=0}^{m(t)}\sum_{i=n_k+1}^{n_{k+1}}\eta(i-1)e^{-2\sum_{j=i}^{t}\eta(j)} \\
&\leq \sum_{i=1}^{n_0}\eta(i-1)e^{-2\sum_{j=n_0+1}^{n_{m(k)}-1}\eta(j)} + \sum_{k=0}^{m(t)}\sum_{i=n_k+1}^{n_{k+1}}\eta(i-1)e^{-2\sum_{j=n_{k+1}+1}^{n_{m(t)}-1}\eta(j)} \\
&\leq e^{-2m(t)}\sum_{i=0}^{n_0-1}\eta(i) + \sum_{k=0}^{m(t)}e^{-2(m(t)-k-2)}\sum_{i=n_k+1}^{n_{k+1}}\eta(i-1) \\
&\leq e^{-2m(t)} + \frac{3}{2}\sum_{k=0}^{m(t)}e^{-2(m(t)-k-2)} \\
&\leq e^{-2m(t)} + \frac{3}{2}e^{-2(m(t)-2)}\Big[e^{2m(t)} + \int_0^{m(t)}e^{2x}dx\Big].
\end{aligned}$$

Thus, there exists some positive constant C_2 such that

$$II(t) \leq C_2, \qquad t = 0,1,2,\cdots.$$

Now, substituting this and (18) into (17), we have

$$\delta^{(t+1)} \leq C_1 + \Big(2+\frac{B}{AA_1}\Big)(1-C_2) \qquad t = 0,1,\cdots, \ \text{a.s..}$$

We then obtain the desired result. □

Next we consider the property of the objective function $W(\boldsymbol{b},F)$ given in (1) (which is denoted by $W(\boldsymbol{b})$ in the following for simplicity) and the solution set of problem (1) and (2). It was shown in [5] that $W(\boldsymbol{b})$ is concave on $\boldsymbol{B}$, and the set of log-optimal portfolios forms a convex set. Under some reasonable assumptions, we will claim some further results in the succedent steps. We first give an assumption.

A2: There are not two different portfolios $\boldsymbol{b}$ and $\overline{\boldsymbol{b}}$ such that their gains $\boldsymbol{b}^T\boldsymbol{X}$ and $\overline{\boldsymbol{b}}^T\boldsymbol{X}$ are identical almost surely, where $\boldsymbol{X}$ is the stochastic vector obeying the probability distribution $F(x)$, the same as the i.i.d. sequence $\{\boldsymbol{x}(t)\}$ of stock gain vectors.

Lemma 3 Under the assumptions **A1** and **A2**, the objective function $W(\boldsymbol{b})$ is strictly concave on the simplex $\boldsymbol{B}$, and so the log-optimal portfolio of problem (1) and (2) is unique.

Proof In suffices to verify the strict concavity of $W(\boldsymbol{b})$ on $\boldsymbol{B}$. For any two vectors $\boldsymbol{b}, \overline{\boldsymbol{b}} \in \boldsymbol{B}$, $\boldsymbol{b} \neq \overline{\boldsymbol{b}}$ and $\lambda \in (0,1)$, due to the strict concavity of $\log(x)$, we have

$$\log((1-\lambda)\boldsymbol{b}^T\boldsymbol{x} + \lambda\overline{\boldsymbol{b}}^T\boldsymbol{x}) \geq (1-\lambda)\log(\boldsymbol{b}^T\boldsymbol{x}) + \lambda\log(\overline{\boldsymbol{b}}^T\boldsymbol{x})$$

and with the strict inequality if $\boldsymbol{b}^T\boldsymbol{x} \neq \overline{\boldsymbol{b}}^T\boldsymbol{x}$.

Set $K = \{\boldsymbol{x} : \boldsymbol{b}^T\boldsymbol{x} = \overline{\boldsymbol{b}}^T\boldsymbol{x}\}$. Then from assumption **A2** we know

$$\int_K dF(\boldsymbol{x}) = \delta < 1.$$

Thus, there exists an open set H, $K \subset H$, such that

$$\int_H dF(\boldsymbol{x}) = \overline{\delta} < 1.$$

Therefore, on the compact set $[A,B]^m \setminus H$,

$$\log((1-\lambda)\boldsymbol{b}^T\boldsymbol{x} + \lambda\overline{\boldsymbol{b}}^T\boldsymbol{x}) - (1-\lambda)\log(\boldsymbol{b}^T\boldsymbol{x}) - \lambda\log(\overline{\boldsymbol{b}}^T\boldsymbol{x}) > 0.$$

Moreover, from the property of continuous function on compact set, we further know, there exists some positive constant ϵ, such that, $\forall\, \boldsymbol{x} \in [A,B]^m \setminus H$,

$$\log((1-\lambda)\boldsymbol{b}^T\boldsymbol{x} + \lambda\overline{\boldsymbol{b}}^T\boldsymbol{x}) - (1-\lambda)\log(\boldsymbol{b}^T\boldsymbol{x}) - \lambda\log(\overline{\boldsymbol{b}}^T\boldsymbol{x}) \geq \epsilon > 0.$$

Hence, from assumption **A1** and above analysis,

$$\begin{aligned}
W((1-\lambda)\boldsymbol{b} + \lambda\overline{\boldsymbol{b}}) = &\int_{[A,B]^m} \log((1-\lambda)\boldsymbol{b}^T\boldsymbol{x} + \lambda\overline{\boldsymbol{b}}^T\boldsymbol{x})dF(\boldsymbol{x}) \\
&\geq (1-\lambda)\int_H \log(\boldsymbol{b}^T\boldsymbol{x})dF(\boldsymbol{x}) + \lambda\int_H \log(\overline{\boldsymbol{b}}^T\boldsymbol{x})dF(\boldsymbol{x}) \\
&+ \int_{[A,B]^m\setminus H} [(1-\lambda)\log(\boldsymbol{b}^T\boldsymbol{x}) + \lambda\log(\overline{\boldsymbol{b}}^T\boldsymbol{x}) + \epsilon]dF(\boldsymbol{x}) \\
&\geq (1-\lambda)W(\boldsymbol{b}) + \lambda W(\overline{\boldsymbol{b}}) + \epsilon(1-\overline{\delta}) \\
&> (1-\lambda)W(\boldsymbol{b}) + \lambda W(\overline{\boldsymbol{b}}).
\end{aligned}$$

The desired result then follows. □

Now we can obtain our main result of this paper.

Theorem 2 For the **Algorithm P**, if the learning rate sequence $\{\eta(t)\}$ are chosen to satisfy (11) and, there exists some $q \geq 2$ such that

$$\sum_{t=0}^{+\infty} \eta^{1+\frac{q}{2}}(t) < +\infty; \tag{21}$$

moreover, the sequence $\{\boldsymbol{x}(t)\}$ of stock gain vectors satisfy the assumptions **A1** and **A2**. Then for any initial guess vector $\boldsymbol{w}^{(0)} \setminus \{0\}$, the iteration sequence $\{\boldsymbol{w}^{(t)}\}$ is convergent a.s..

Proof In order to apply the techniques given in [1], [2], [11], we introduce the following strict Lyapunov function

$$V(\boldsymbol{w}) = \sum_{i=1}^{m} w_i^2 - \mathsf{E}\Big[\log\Big(\sum_{i=1}^{m} w_i^2 X_i\Big)\Big] \tag{22}$$

with the related equilibria set

$$J = \Big\{ w \neq 0 : \mathsf{E}\Big[\frac{X_i w_i}{\sum\limits_{j=1}^{m} X_j w_j^2}\Big] = w_i, \quad i = 1, 2, \cdots, m \Big\}. \tag{23}$$

The induced gradient flow system is

$$\frac{dw_i(t)}{dt} = \mathsf{E}\Big[\frac{X_i w_i(t)}{\sum\limits_{j=1}^{m} X_j w_j^2(t)}\Big] - w_i(t), \qquad i = 1, 2, \cdots, m. \tag{24}$$

We first show that the nonlinear system (23) has only finite number of solutions. Obviously, it suffices to prove that for the system (23), there are only finite number of solutions w satisfying that $w_i \neq 0,\ i = 1, 2, \cdots, m$, which implies the desired result.

Let $\bar{\boldsymbol{b}} = (\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_m)^T = (w_1^2, w_2^2, \cdots, w_m^2)^T$. Since $w_i \neq 0,\ i = 1, 2, \cdots, m$, from (23) we have

$$\mathsf{E}\Big[\frac{X_i}{\sum\limits_{j=1}^{m} X_j \bar{b}_j}\Big] - 1 = 0, \qquad i = 1, 2, \cdots, m,$$

and $\sum\limits_{i=1}^{m} \bar{b}_i = 1$.

Thus, $\bar{\boldsymbol{b}}$ is an interior point of the simplex $\boldsymbol{B}$, and $\frac{\partial}{\partial b_i} W(\boldsymbol{b})|_{\boldsymbol{b}=\bar{\boldsymbol{b}}} = 0,\ i = 1, 2, \cdots, m$. Due to the strict concavity of $W(\boldsymbol{b})$, $\bar{\boldsymbol{b}}$ is the local maximum point of $W(\boldsymbol{b})$ on $\boldsymbol{B}$. If $\bar{\boldsymbol{b}} \neq \boldsymbol{b}^*$, the unique log-optimal portfolio. Then, from the strict concavity of $W(\boldsymbol{b})$ on $\boldsymbol{B}$, we easily know that for any $\lambda \in (0, 1)$,

$$W((1-\lambda)\boldsymbol{b} + \lambda \boldsymbol{b}^*) > W(\boldsymbol{b}).$$

We can choose λ sufficiently small such that $(1-\lambda)\boldsymbol{b} + \lambda \boldsymbol{b}^*$ belongs to any neighborhood domain of $\boldsymbol{b}$ in $\boldsymbol{B}$. This leads to a contradiction.

Hence, if $\boldsymbol{b}^*$ belongs to the interior of $\boldsymbol{B}$, $\bar{\boldsymbol{b}} = \boldsymbol{b}^*$, i.e., $w_i = \pm\sqrt{b_i^*},\ i = 1, 2, \cdots, m$; if $\boldsymbol{b}^*$ belongs to the boundary of $\boldsymbol{B}$, there are no solutions. We then have the above conclusion.

Furthermore, since we are only interested in the convergence of the **Algorithm P** and $\lim \eta(t) = 0$, without loss of generality, we assume the learning rate sequence $\{\eta(t)\}$ take values in $\left(0, \frac{1}{2}\right)$. Then, under the assumption **A1** Theorem 1 holds. Hence, the conditions of Proposition 2.1 in [1] are satisfied, and then we have from Theorem 1.2 and Corollary 3.3 in [1] that $\{\boldsymbol{w}^{(t)}\}$ converges toward an equilibrium in the equilibria set J a.s.. □

In the final part of this section, we proceed to consider the stability analysis of the dynamical system (24) corresponding to the iteration sequence (10).

Theorem 3 The dynamical system (24) with the initial value vector $\boldsymbol{w}(0) \in S^{m-1}$ is a smooth dynamical system on the compact manifold S^{m-1}. A equilibrium $\boldsymbol{w} \in J$ is asymptotically stable if and only if $bw^2 = (w_1^2, w_2^2, \cdots, w_m^2)^T = b^*$, the unique log-optimal portfolio. And for almost all initial value vector $\boldsymbol{w}(0) \in S^{m-1}$ (with conventional surface measure), $\{\boldsymbol{w}(t)\}$ converges toward one of such asymptotically stable equilibrium.

Proof We consider the following auxiliary dynamical system

$$\begin{cases} \boldsymbol{w}(t)|_{t=0} = \boldsymbol{w}(0) \in S^{m-1}; \\ \dfrac{dw_i(t)}{dt} = \mathsf{E}\Big[\dfrac{X_i w_i(t)}{\sum\limits_{j=1}^{m} X_j w_j^2(t)}\Big] - \dfrac{w_i(t)}{\sum\limits_{j=1}^{m} w_j^2(t)}. \end{cases} \tag{25}$$

Obviously,

$$\frac{d}{dt}\|\boldsymbol{w}(t)\|^2 = 2\sum_{i=1}^{m} w_i(t) \cdot \Big\{ \mathsf{E}\Big[\frac{X_i w_i(t)}{\sum\limits_{j=1}^{m} X_j w_j^2(t)}\Big] - \frac{w_i(t)}{\sum\limits_{j=1}^{m} w_j^2(t)} \Big\} = 0.$$

Therefore, $\|w(t)\|^2 = \|w(0)\|^2 = 1$, i.e., (25) is a smooth dynamical system on the compact manifold S^{m-1}, and thus this sequence $\{w(t)\}$ also satisfies the dynamical system (24). In other words, the dynamical system (24) with the initial value vector $w(0) \in S^{m-1}$ is also a smooth dynamical system on S^{m-1}.

Due to the strict concavity of $W(\boldsymbol{b})$, it is easy to see that the equilibrium $\boldsymbol{w}$ in J is asymptotically stable if and only if $w_i^2 = b_i^*$, $i = 1, 2, \cdots, m$. Hence, from the basic theories for dynamical system on compact manifold ([8] and [9]) that, for almost all $w(0) \in S^{m-1}$ (with the measure on S^{m-1}), $\{w(t)\}$ converges toward one of such asymptotically stable equilibria. The desired result then follows in th last. □

Remark 1 According to the deduction employed in the derivation of Theorem 2.3.1 in [11] and the theory in [1]. The limit of the iteration sequence $\{\boldsymbol{w}^{(t)}\}$ is determined by the asymptotic behavior of the dynamical system (24) with some specified initial vector $w(0) \in S^{m-1}$ which is induced by some function subsequence. Thus, intuitively, for almost all initial guess vector $\boldsymbol{w}^{(0)} \in R^m \setminus \{0\}$, $\{\boldsymbol{w}^{(t)}\}$ should converges asymptotically to one stable equilibrium $\boldsymbol{w}$ with $\boldsymbol{w}^2 = b^*$. Numerical tests in our experiments illustrate this. But we can not give a rigorous proof on this conclusion up to now.

§4. Applications

We apply the **Algorithm P** to search log-optimal portfolio with real data from the Exchange Institute of Shanghai Security. In the year of 1997, there are totally 235 working exchange days. We choose 20 stocks operated on this exchange institute to invest. The stock gain vector at instant t is the ratios of closing prices at this day over those of the previous day. We denote by $\{\boldsymbol{x}(t)\}^235_{t=1}$ the sequence of stock gain vectors in this year.

Since we have only finite data, at each iteration step of implementing the **Algorithm P** the related gain vector in use is chosen randomly from the sequence $\{\boldsymbol{x}(t)\}_{t=1}^{235}$ with equal probability. To make the iteration sufficiently stationary, the iteration number of times in the halt rule of **Algorithm P** is chosen to be 10^6, while the related control precision is taken 10^{-7}. And the learning rate sequence $\{\eta(t)\}$ is chosen to be $\eta(t) = \dfrac{1000}{1000+t}$. The concrete algorithm reads as follows.

Step 0 Randomly generate a nonzero vector $\boldsymbol{u} = (u_1, u_2, \cdots, u_{20})^T$ and set the initial guess vector to be $\boldsymbol{w}^{(0)} = \dfrac{1}{\sqrt{\sum_{i=1}^{20} u_i^2}}\boldsymbol{u}$, $t = 0$;

Step 1 generate a positive integer $n \in [1, 235]$ randomly with equal probability, then compute the modified results $\boldsymbol{w}^{(t+1)} \equiv (w_1^{(t+1)}, w_2^{(t+1)}, \cdots, w_{20}^{(t+1)})^T$ by

$$w_i^{(t+1)} = w_i^{(t)} + \frac{1000}{1000+t}\Big[\frac{x_i(n)w_i^{(t)}}{\sum_{j=1}^{20}(w_j^{(t)})^2 x_j(n)} - w_i^{(t)}\Big], \qquad i = 1, 2, \cdots, 20;$$

Step 2 halt if the iteration number of times t reaches 10^6 or if

$$\max\{|w_1^{(t+1)} - w_1^{(t)}|, \cdots, |w_{20}^{(t+1)} - w_{20}^{(t)}|\} < 10^{-7},$$

set the portfolio $\boldsymbol{b}^* = (\boldsymbol{w}^{(t+1)})^2$, i.e., $b_i^* = (w_i^{(t+1)})^2$, $i = 1, 2, \cdots, 20$; otherwise $t \leftarrow (t+1)$, go to Step 1.

We use the Matlab 5.0 to write a program of this algorithm. Numerical experiments show that for any randomly chosen initial vector $\boldsymbol{w}^{(0)}$ as above, this algorithm is convergent and the computed portfolio $\boldsymbol{b}^*$ satisfies the restriction condition $\sum_{i=1}^{20} b_i^* = 1$ with desired accuracy.

As usual, for any computed portfolio $\boldsymbol{b}^*$, the total gain of that year according to this portfolio is given by $\prod_{t=1}^{235}(\sum_{i=1}^{20} b_i x_i(t))$, while the total average gain of that year is given by $(\prod_{t=1}^{235}(\sum_{i=1}^{20} b_i x_i(t)))^{\frac{1}{235}}$. We find that for any computed portfolio b^* by this algorithm, the total gain is almost near the value 1.7000. For example, we implement this algorithm 10 times, and obtain 10 portfolios with the corresponding total gains 1.6944, 1.6902, 1.6992, 1.7012, 1.7021, 1.7004, 1.7021, 1.6820, 1.7034, 1.6969, respectively. Since there are only finite stock gain vectors for this problem, we can also use the constr function given in Matlab 5.0 to compute the optimal total gain of this problem. The related result is 1.7036. This shows that the results obtained from **Algorithm P** are satisfactory and acceptable.

For our algorithm, the computed portfolio corresponding to the total gain 1.7034 given above is 0.5425, 0, 0.002, 0(8 times), 0.3623, 0(6 times), 0.0950, 0, i.e., the 1th and 12th stocks are invested mainly. The related total average gain for this investment is 1.0023, while the total average gains for each stock are 1.0021, 0.9993, 1.0015, 0.9996, 0.9984, 0.9993, 0.9991, 1.0000, 1.0000, 0.9984, 1.0003, 1.0019, 0.9985, 1.0012, 1.0001, 0.9998, 1.0001, 0.9994, 1.0019, 0.9989, respectively. Thus, this portfolio investment is superior to the single stock investment.

Acknowledgements The authors are grateful to Mr. Cao Yijian for his beneficial help on the numerical simulation of this paper. Dr. Huang Jianguo is also very grateful to Prof. Xu Lei who gave this author a lot of help in scientific research when he visited the Department of Computer Science and Engineering, The Chinese University of Hong Kong from September, 1997 to March, 1998.

References

[1] M. Benaim, A dynamical system approach to stochastic approximations, *SIAM J. Control and Optimization*, **34**(1996), 437–372.

[2] A. Benveniste, M. Metivier and P. Priouret, *Stochastic Approximations and Adaptive Algorithms*, Springer-Verlag, New York, 1990.

[3] R.W. Brockett, Least squares matching problems, *LAA*, **122–124**(1989), 761-777.

[4] R.W. Brockett, Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems, *LAA*,**146**(1991), 79–91.

[5] T.M. Cover, Universal portfolios, *Mathematical Finance*, **1**(1)(1991), 1–29.

[6] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.

[7] J.L. Doob, *Stochastic Process*, John Wiley, New York, 1953.

[8] M.W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, New York, 1974.

[9] Uwe Helmke and J.B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, New York, 1994.

[10] M. Loeve, *Probability Theory (I,II)*, Springer-Verlag, New York, 1977–1978.

[11] H.J. Kushner and D.S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, New York, 1978.

[12] E. Oja and J. Karhunen, On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random mateix, *J. Math. Anal. Appli.*, **106**(1985), 69–84.

[13] M. Spivak, *A Comprehensive Introduction to Differential Geometry (1–5)*, Boston Publish or perish, 1970, 1975.

[14] Ye Zhongxing and Zhang Yijun, Improved genetic algorithm to optimal portfolio with risk control, *J. Shanghai Jiao Tong Univer.*, **E-1**(**2**)(1996), 9–16.

[15] Ye Zhongxing, Ren Qingsheng and Lin Weiran, Random segment searching algorithm and application in optimal portfolio with risk control, *Proc. of Int. Symp. on Intelligent Data Engineering and Atomatic Learning*, 121–126, Hong Kong, 1998.10.14–17.

求解log-最优组合投资问题的一个自适应算法及其理论分析

黄建国　　叶中行

(上海交通大学应用数学系, 上海, 200240)

本文给出了一个求解log-最优组合投资问题的自适应算法, 它是一个变型的随机逼近方法. 该问题是一个约束优化问题, 因此, 采用基于约束流形的梯度上升方向替代常规梯度上升方向. 在一些合理的假设下证明了算法的收敛性并进行了渐近稳定性分析. 最后, 本文将该算法应用于上海证券交易所提供的实际数据的log-最优组合投资问题求解, 获得了理想的数值模拟结果.

关键词: log-最优组合投资, 随机逼近, 自适应算法.

学科分类号: O212.4.