Master's Thesis

Universal portfolios with trading cost and downside risk

Guidance

Associate Professor Nobuo Yamashita

Rei UMEDA

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2014

Abstract

A universal portfolio is a portfolio sequence which theoretically achieves the almost same wealth by the best constant rebalanced portfolio in the long term. It is constructed from historical data of asset prices. However, since the existing models for the universal portfolio consider the wealth only, they are not necessarily practical models.

In this paper, in order to construct a more practical universal portfolio, we focus on the relationship between the universal property for the universal portfolio sequence and the no internal regret for the algorithm of the online optimization. We can construct more flexible portfolio sequence by using general framework of the online optimization. As the results, we propose universal portfolios taking account of trading cost and downside risk. Moreover, we conduct some numerical experiments to show the effectiveness of the proposed models.

Contents

1 Introduction **1**

2 Preliminary **2**

2.1 A best constant rebalanced portfolio and a universal portfolio 2

2.2 Online optimization and regret 4

3 Universal portfolios with trading cost and downside risk **5**

3.1 The universal portfolio with trading cost 6

3.2 The universal portfolio with downside risk 7

4 Numerical Experiments **10**

4.1 Data 10

4.1.1 Data 1 10

4.1.2 Data 2 10

4.2 The experiment with data 1 10

4.2.1 The model with trading cost 10

4.2.2 the model with downside risk 11

4.3 The experiment with data 2 12

4.3.1 The model with trading cost 12

4.3.2 The model with downside risk 13

5 Conclusion **13**

6 Appendix **15**

1 Introduction

We consider to invest in various assets to achieve a large amount of wealth. Since the prices of assets are uncertain, we first define what is the optimal investment of assets. The mathematical problem to find an optimal portfolio is called portfolio selection problem.

We may divide portfolio selection problems into two types in terms of the asset composition, the buy-and-hold strategy and the rebalance strategy. The buy-and-hold strategy is to buy stocks and hold them for a specific period. On the other hand, the rebalance strategy allows to buy and sell in the period to get more optimal portfolio. Therefore, under-weighted securities can be purchased with newly saved money; alternatively, over-weighted securities can be sold to purchase under-weighted securities.

We farther divide the rebalance strategy into two types, the constant rebalance strategy and the inconstant rebalance strategy. The constant rebalance strategy is to rebalance assets regularly so that a weight of each asset is fixed. On the other hand, the inconstant rebalance strategy is to allow to change portfolio weights and rebalance regularly.

The Best Constant Rebalanced Portfolio (BCRP) is the constant rebalanced portfolio that achieves the maximum wealth. However, we cannot construct BCRP in advance since the prices of assets are uncertain. Thus, we cannot invest in the BCRP. Therefore, there have been studied the inconstant rebalanced portfolio selection algorithm which theoretically achieves the almost same wealth by the BCRP in the long term. Cover [1] defined a portfolio given by such algorithms as a universal portfolio.

A universal portfolio is usually constructed from historical data of the asset prices. Cover [1] proposed an algorithm that constructs a universal portfolio. The algorithm exploits integrals of historical data of the asset prices. Thus, it is not easy to implement the algorithm because the numerical integral takes much time. On the other hand, Helmbold, Schapire, Singer and Warmuth [2] proposed the algorithm using exponentiated gradient method. They proved that the portfolio sequence generated by the algorithm was universal. We can implement it easily because we can define a portfolio sequence from the prices of assets only in each trading day. However, the models in [1, 2] are not necessarily practical ones since trading cost and downside risk are not considered.

In this paper, in order to construct a more practical universal portfolio, we focus on the relationship between the universal property for the universal portfolio sequence and the no internal regret for the algorithm of the online optimization. We can construct more flexible portfolio sequence by using general framework of the online optimization. Moreover, we conduct some numerical experiments to show the effectiveness of the proposed models.

This paper is organized as follows. Section 2 formally establish definitions. The proposed models are established in section 3. Section 4 presents numerical results. Finally, definitions of convex functions are written in the appendix.

Throughout this paper, we use the following notations. For a vector $\boldsymbol{w} \in \mathbb{R}^n$, $\boldsymbol{w}^{\mathrm{T}}$ denote the transposition of $\boldsymbol{w}$. For vectors $\boldsymbol{w},\ \boldsymbol{v} \in \mathbb{R}^n$, $\langle \boldsymbol{w}, \boldsymbol{v} \rangle$ denote inner product of $\boldsymbol{w}$ and $\boldsymbol{v}$. For a vector $\boldsymbol{w} \in \mathbb{R}^n$, $\|\boldsymbol{w}\|$ denote the Euclidean norm defined by $\|\boldsymbol{w}\| \overset{\text{def}}{=} \sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}$.

For a vector $\boldsymbol{w} \in \mathbb{R}^n$, $\|\boldsymbol{w}\|_1$ denote the l_1 norm defined by $\|\boldsymbol{w}\|_1 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n |w_i|}$. A vector $\boldsymbol{e} = (1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^n$ denote the vector whose components are all 1.

2 Preliminary

In this section, we introduce the best constant rebalanced portfolio and a universal portfolio and discuss their relation. Then, we explain the online convex programming problem and the "regret" which is one of the criterions for evaluating the online optimization. Finally, we introduce the online mirror decent algorithm that solves the online convex programming problem.

2.1 A best constant rebalanced portfolio and a universal portfolio

We consider a portfolio consisting of N assets. Let w_i be a weight of the asset i in the investment, and let $\boldsymbol{w} = (w_1, \cdots, w_N)^{\mathrm{T}}$ be a portfolio. In this paper, we suppose that the short selling is forbidden. Then, the portfolio satisfies $w_i \geq 0$ and $\sum_{i=1}^N w_i = 1$. Let P_i^t denote the price of the ith asset on day t. We use a rate of return defined as follows.

Definition 2.1. *A rate r_i^t of the return of the ith asset between $[t, t+1]$ is given by*

$$r_i^t = \frac{P_i^{t+1}}{P_i^t}.$$

We now simply call the rate of return a return. We denote a return vector on days $[t, t+1]$ as $\boldsymbol{R}^t = (r_1^t, \cdots, r_N^t)^{\mathrm{T}}$.

When the portfolio on t is $\boldsymbol{w}^t$, the rate of return of the portfolio in $[t, t+1]$ is given by

$$\langle \boldsymbol{w}^t, \boldsymbol{R}^t \rangle = \sum_{i=1}^N w_i^t r_i^t. \tag{1}$$

Suppose that we observed a return sequence $\{\boldsymbol{R}^t\}$, and that we select a portfolio sequence $\{\boldsymbol{w}^t\}$. Then, the wealth increases from day 1 to day $T+1$ by a factor of

$$S_T(\{\boldsymbol{w}^t\}, \{\boldsymbol{R}^t\}) \stackrel{\text{def}}{=} \prod_{t=1}^T \langle \boldsymbol{w}^t, \boldsymbol{R}^t \rangle. \tag{2}$$

When we adopt a constant rebalanced portfolio strategy, that is, $\boldsymbol{w}^t = \boldsymbol{w}$ for all t, the resulting wealth from day 1 to day $T+1$ increase by a factor of

$$\prod_{t=1}^T \langle \boldsymbol{w}, \boldsymbol{R}^t \rangle.$$

We define the best constant rebalanced portfolio (BCRP). When a return sequence $\boldsymbol{R}^1, \cdots, \boldsymbol{R}^T$ was observed, the BCRP was the constant rebalanced portfolio whose wealth was the most increased. Let $\boldsymbol{w}^*$ be the BCRP. Then, $\boldsymbol{w}^*$ is the optimal solution of the following problem.

$$\begin{array}{ll} \max & S_T(\{\boldsymbol{w}\},\{\boldsymbol{R}^t\}) \\ \text{s.t.} & \boldsymbol{w} \in K, \end{array} \tag{3}$$

where K denotes a simplex set defined as $K = \{\boldsymbol{w} \in \mathbb{R}^N | \sum_{i=1}^N w_i = 1,\ w_i \geq 0,\ i = 1, \cdots, N\}$.

Note that the problem (3) is not practical from the following two reasons.

- The objective function is not concave.
- We cannot deal with problem (3) when $\{\boldsymbol{R}^t\}$ is unknown.

We discuss how to remove these difficulties. First of all, we convert the non-concave problem (3) into a concave problem by using the logarithm function. Let ls_t and LS_T be defined

$$ls_t(\boldsymbol{w}) \stackrel{\text{def}}{=} \log(\langle \boldsymbol{w}, \boldsymbol{R}^t \rangle)$$

$$LS_T(\{\boldsymbol{w}^t\},\{\boldsymbol{R}^t\}) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T} ls_t(\boldsymbol{w}^t), \tag{4}$$

respectively. Note that, $LS_T(\{\boldsymbol{w}\},\{\boldsymbol{R}^t\})$ can be written as $\frac{1}{T}\sum_{t=1}^{T} ls_t(\boldsymbol{w})$. Thus, $LS_T(\{\boldsymbol{w}\},\{\boldsymbol{R}^t\})$ is concave with respect to $\boldsymbol{w}$. Since the logarithm function is monotonically increasing, the following convex programming problem is equivalent to (3).

$$\begin{array}{ll} \min & -LS_T(\{\boldsymbol{w}^t\},\{\boldsymbol{R}^t\}) \\ \text{s.t.} & \boldsymbol{w} \in K. \end{array} \tag{5}$$

Next we consider how to construct a portfolio similar to the BCRP if $\{\boldsymbol{R}^t\}$ is unknown. To this end, Cover [1] introduced a universal portfolio defined as follows.

Definition 2.2. *A portfolio sequence $\{\boldsymbol{w}^t\}$ is said to be universal for the problem (5) if*

$$\lim_{T\to\infty} \max_{\{\boldsymbol{R}^t\}} [LS_T(\{\boldsymbol{w}^*\},\{\boldsymbol{R}^t\}) - LS_T(\{\boldsymbol{w}^t\},\{\boldsymbol{R}^t\})] \leq 0 \tag{6}$$

We also call an algorithm that generates $\{\boldsymbol{w}^t\}$ universal. Cover proposed an investment strategy using an averaging method to pick their portfolio vectors. The portfolio vector given on day $t+1$ is the weighted average over all feasible portfolio vectors, where the weight of each possible portfolio vector is determined by its performance in the past. That is,

$$\hat{\boldsymbol{w}}^{t+1} = \frac{\int \hat{\boldsymbol{w}} S_t(\hat{\boldsymbol{w}}) d\hat{\boldsymbol{w}}}{\int S_t(\hat{\boldsymbol{w}}) d\hat{\boldsymbol{w}}},\ \hat{\boldsymbol{w}}^1 = (1/N, \cdots, 1/N).$$

Cover showed that $\{\boldsymbol{w}^t\}$ is universal for problem (5). Note that, we must calculate the above integrals in each t.

Helmbold, Schapire, Singer and Warmuth [2] proposed investment strategies without calculating integrals. When we construct a portfolio vector on day $t+1$ only, we may formulate the portfolio selection problem as follows.

$$
\begin{array}{ll}
\min_{\boldsymbol{w}} & f^t(\boldsymbol{w}) \\
\text{s.t.} & \boldsymbol{w} \in K,
\end{array}
\tag{7}
$$

where $f^t(\boldsymbol{w}) = -ls_t(\boldsymbol{w})$. When $\boldsymbol{w}^{t+1}$ is the optimal solution of (7), $\boldsymbol{w}^{t+1}$ depend on $\boldsymbol{R}^t$ only and it is independent of $\boldsymbol{R}^{t-1}, \boldsymbol{R}^{t-2}, \cdots$. Since $\boldsymbol{w}^t$ include the information on $\boldsymbol{R}^{t-1}$, we choose $\boldsymbol{w}^{t+1}$ near $\boldsymbol{w}^t$. In fact, Kivinen and Warmuth [3] show that good performance can be achieved by choosing $\boldsymbol{w}^{t+1}$ close to $\boldsymbol{w}^t$. Then, we add a Bregman divergence between $\boldsymbol{w}$ and $\boldsymbol{w}^t$ to (7).

$$
\begin{array}{ll}
\min_{\boldsymbol{w}} & f^t(\boldsymbol{w}) + \frac{1}{\eta_t} B_\psi(\boldsymbol{w}, \boldsymbol{w}^t) \\
\text{s.t.} & \boldsymbol{w} \in K,
\end{array}
\tag{8}
$$

when the Bregman divergence is defined as

$$
B_\psi(\boldsymbol{w}, \boldsymbol{v}) = \psi(\boldsymbol{w}) - \psi(\boldsymbol{v}) - \langle \nabla\psi(\boldsymbol{v}), \boldsymbol{w} - \boldsymbol{v} \rangle,
$$

with a strictly convex function ψ, and η_t is a positive parameter to balance the amount of variation. We cannot solve (8) easily because (8) is a nonlinear programming problem with constraints. We linearlized f^t adopt a Bregman divergence with $\psi_1(\boldsymbol{w}) = \sum_{i=1}^N w_i \log(w_i)$. Then, the problem is

$$
\begin{array}{ll}
\min & \langle \nabla f^t(\boldsymbol{w}), \boldsymbol{w} - \boldsymbol{w}^t \rangle + f^t(\boldsymbol{w}^t) + \frac{1}{\eta_t} B_{\psi_1}(\boldsymbol{w}, \boldsymbol{w}^t) \\
\text{s.t.} & \boldsymbol{w} \in K.
\end{array}
$$

The solution of this problem is written as

$$
w_i^{t+1} = \frac{w_i^t \exp(\frac{\eta_t x_i^t}{\langle \boldsymbol{w}^t, \boldsymbol{R}^t \rangle})}{\sum_{j=1}^N w_j^t \exp(\frac{\eta_t x_j^t}{\langle \boldsymbol{w}^t, \boldsymbol{R}^t \rangle})} \quad i = 1, \cdots, N.
\tag{9}
$$

The update (9) is called the exponentiated gradient update method (EG-update method). Moreover, it is proved that the EG-update method is universal for problem (5) when $\eta_t = \frac{1}{\sqrt{t}}$ in [2].

2.2 Online optimization and regret

An online optimization problem is an optimization problem that a decision maker must make a decision at each point t in time. An objective function value at $t+1$ in a period depends on the decision at t as well as the state of the world at $t+1$. The difficulty is that decision must be made in advance of any knowledge of the future state even probabilistic. In this paper, we consider that we decide $\boldsymbol{w}^t$ on day t when we do not know an objective function f^t. After we know f^t, we decide $\boldsymbol{w}^{t+1}$. The online optimization problem appears in the machine learning for example.

We introduce a evaluation criteria for the online optimazation. Since the objective function f^t is unknown at the decision made, it is difficult to minimize $\sum_{t=1}^T f^t(\boldsymbol{w}^t)$.

Then we compare with the optimal value of $\sum_{t=1}^{T} f^t(\boldsymbol{w})$. We call their cost difference as the regret. The regret is used to evaluate an online algorithm. The regret of the algorithm until the iteration T is written as

$$Regret(T) = \sum_{t=1}^{T} f^t(\boldsymbol{w}^t) - \sum_{t=1}^{T} f^t(\boldsymbol{w}^*),$$

where $\boldsymbol{w}^* \in \arg\min_{\boldsymbol{w}} \sum_{t=1}^{T} f^t(\boldsymbol{w})$. We usually care about the average regret $\frac{Regret(T)}{T}$. Foster and Vohra[4] called the algorithm as no internal regret if

$$\lim_{T\to\infty} \frac{Regret(T)}{T} = 0. \tag{10}$$

We introduce the online mirror descent method (OMD method), which is one of the popular algorithms to solve the online convex optimization problem. Let the online convex optimization problem be written as

$$\begin{array}{ll} \min & \sum_{t=1}^{n} f^t(\boldsymbol{w}) \\ \text{s.t.} & \boldsymbol{w} \in S, \end{array}$$

where each $f^t : \mathbb{R}^n \to \mathbb{R}$ is convex and S is a closed convex set. The OMD method is described as follows.

OMD method

$t = 1$ to T

$f_t' \in \partial f^t(\boldsymbol{w}^t)$

$\boldsymbol{w}^{t+1} = \arg\min_{\boldsymbol{w}\in S}[B_\psi(\boldsymbol{w}, \boldsymbol{w}^t) + \eta_t \langle f_t', \boldsymbol{w} - \boldsymbol{w}^t\rangle]$

Here, $\partial f^t(\boldsymbol{w}^t)$ is a subdifferential of f^t at $\boldsymbol{w}^t$. The subdifferential of the convex function f at $\boldsymbol{w}$ is defined as follows.

$$\partial f(\boldsymbol{w}) = \{\boldsymbol{\tau} \in \mathbb{R}^n | f(\boldsymbol{u}) - f(\boldsymbol{w}) \geq \langle \boldsymbol{\tau}, \boldsymbol{u} - \boldsymbol{w}\rangle, \text{ for all } \boldsymbol{u} \in \mathbb{R}^n\}.$$

We now discuss the relation between the OMD method and the EG-update method introduced in section 2.1. Let $\psi(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}} \log(\boldsymbol{w})$ and let S be the simplex set K, the OMD method is equivalent to the EG-update method. Then, the EG-update method is a special case of the OMD method.

We denote the regret of the OMD method. The regret of the OMD method is $O(\sqrt{T})$ if $\eta_t = \frac{1}{\sqrt{t}}$ in [5]. From (6) and (10), the universal property for the universal portfolio sequence is equivalent to the no internal regret for the algorithm of the online optimization. Then, we can construct a universal portfolio by using the OMD method when each objective function f^t is convex in the portfolio selection problem.

3 Universal portfolios with trading cost and downside risk

In this section, we propose universal portfolios taking account of trading cost and downside risk as more practical models.

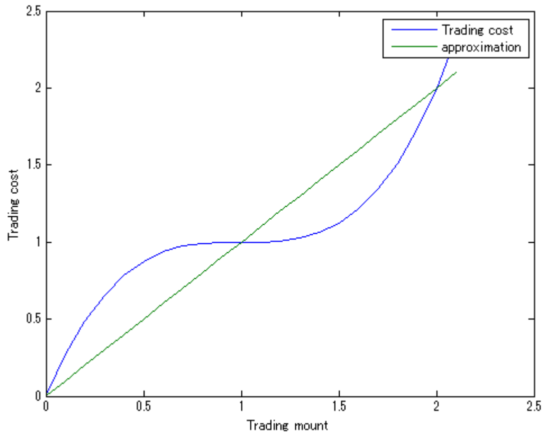

Figure 1: The linear approximation of trading cost

3.1 The universal portfolio with trading cost

First, we define trading cost. The trading cost is additional spending when we buy/sell an asset. In this paper, we consider a trading fee and a market impact as trading cost. The trading fee is required to pay to an asset company according to the trading amount when we buy/sell an asset. The market impact is a change of a price when a large amount of assets are traded. For example, if an investor attempt to buy a large amount of assets, then the price becomes higher than as the investor expected. The relation between the trading cost and the trading amount can be drawn as the figure 1. We approximate the trading cost to a straight line with a slope of σ.

Then, we must define a return with the trading cost. We consider the constant rebalance. When an asset price fluctuates on days $[t, t+1]$, a portfolio weight vector also fluctuates. Since we need to pay trading cost when we rebalance a portfolio to keep a portfolio weight vector constantly, We consider the constant rebalance. Let $v_i(\boldsymbol{w})$ deonote trading cost of the ith asset on day t. Then $v_i(\boldsymbol{w})$ can be written as follows.

$$v_i(\boldsymbol{w}) = \sigma \left| w_i - \frac{w_i r_i^t}{\sum_{j=1}^N w_j r_j^t} \right|.$$

Let $\hat{r}_i^t$ denote the return with the trading cost of the ith asset on days $[t, t+1]$. Then, $\hat{r}_i^t$ can be written as

$$\hat{r}_i^t(\boldsymbol{w}) = r_i^t - v_i^t(\boldsymbol{w}),$$

$$\hat{\boldsymbol{R}}^t(\boldsymbol{w}) = (\hat{r}_1^t(\boldsymbol{w}), \cdots \bar{r}_N^t(\boldsymbol{w}))^{\mathrm{T}}.$$

A function $\hat{f}^t(\boldsymbol{w}) = -\log(\langle \boldsymbol{w}, \hat{\boldsymbol{R}}^t(\boldsymbol{w}) \rangle)$ is not convex, and hence it is difficult to construct an algorithm with universal property. To ensure the convexity, we modify $\hat{f}^t$ as follows.

We may consider that r_i^t has a larger volatirity than $\sum_{j=1}^N w_j r_j^t$ because $\sum_{j=1}^N w_j r_j^t$ is similar to a return of the market in $v_i(\boldsymbol{w})$. Thus, we may replace $\sum_{j=1}^N w_j r_j^t$ with $\sum_{j=1}^N \bar{w}_j r_j^t$ where $\bar{\boldsymbol{w}} \in K$ is a constant vector. By using $\bar{\boldsymbol{w}}$ the trading cost can be written as follows.

$$\bar{v}_i(w_i) \approx \sigma \left| 1 - \frac{r_i^t}{\sum_{j=1}^N \bar{w}_j r_j^t} \right| w_i.$$

The return with the trading cost $\bar{v}_i(w_i)$ on days $[t, t+1]$ and the return vector on days $[t, t+1]$ can be written as follows.

$$\bar{r}_i^t(w_i) = r_i^t - \bar{v}_i(w_i),$$

$$\bar{\boldsymbol{R}}^t(\boldsymbol{w}) = (\bar{r}_1^t(w_1), \cdots, \bar{r}_N^t(w_N))^{\mathrm{T}}.$$

When we adopt this return model, the BCRP is the optimal solution of the following problem.

$$\begin{array}{ll} \min & \sum_{t=1}^T \bar{f}^t(\boldsymbol{w}) \\ \text{s.t.} & \boldsymbol{w} \in K, \end{array} \tag{11}$$

where

$$\bar{f}^t(\boldsymbol{w}) = \langle \boldsymbol{w}, \bar{\boldsymbol{R}}^t(\boldsymbol{w}) \rangle.$$

A function $\bar{f}^t$ is a differentiable function. Moreover, $\bar{f}^t$ is convex from Theorem 6.2 because $\bar{ls}_t(\boldsymbol{w})$ is a concave function. We can apply the following EG-update algorithm. The EG-update algorithm with trading cost

Step0. $\eta_t > 0$, $\bar{f}^t(\boldsymbol{w}) = -\bar{ls}_t(\boldsymbol{w})$, $\boldsymbol{w}^1 = (1/N, \cdots, 1/N)$

Step1. $t = 1$ to T
$\boldsymbol{w}^{t+1} = \arg\min_{\boldsymbol{w} \in K} [B_{\psi_1}(\boldsymbol{w}, \boldsymbol{w}^t) + \eta_t \langle \nabla \hat{f}^t(\boldsymbol{w}), \boldsymbol{w} - \boldsymbol{w}^t \rangle]$

Moreover, the regret of the OMD method is $O(\sqrt{T})$ when $\eta_t = \frac{1}{\sqrt{t}}$. Then, a portfolio vector sequence $\{\boldsymbol{w}^t\}$ constructed by this algorithm is universal for (11).

3.2 The universal portfolio with downside risk

We consider the Conditional Value at Risk (CVaR) as downside risk. The CVaR is defined by a loss. Let $g(\boldsymbol{w}, \boldsymbol{R})$ be a loss function of the portfolio. In this paper, we assume that $\boldsymbol{R}$ depend on the continuous probability density function $p(\boldsymbol{R})$. We often use $g(\boldsymbol{w}, \boldsymbol{R}) = -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R} - \boldsymbol{e})$. The probability of the loss under α is written as follows.

$$\Psi(\boldsymbol{w}, \alpha) = \int_{g(\boldsymbol{w}, \boldsymbol{R}) \leq \alpha} p(\boldsymbol{R}) d\boldsymbol{R}. \tag{12}$$

If $\boldsymbol{w}$ is fixed, $\Psi(\boldsymbol{w}, \alpha)$ is a nondecreasing and right-continuous function with respect to α. In this paper, we assume that $\Psi(\mathbf{x}, \alpha)$ is continuous with respect to α.

The Value at Risk (VaR) is defined as the minimum α such that the probability that the loss on the portfolio over the given time horizon exceeds the confidence level β, that is,

$$\mathrm{VaR}_\beta(\boldsymbol{w}) = \min\{\alpha \mid \Psi(\boldsymbol{w}, \alpha) \geq \beta\}. \tag{13}$$

$\mathrm{VaR}_\beta(\boldsymbol{w})$ is called VaR in the confidence level β. Since $\Psi(\mathbf{x}, \alpha)$ is continuous with respect to α, $\mathrm{VaR}_\beta(\boldsymbol{w})$ is the minimum α satisfying $\Psi(\boldsymbol{w}, \alpha) = \beta$.

The CVaR is a expected value which a portfolio exceeds VaR, that is

$$\mathrm{CVaR}_\beta(\boldsymbol{w}) = \frac{\int_{g(\boldsymbol{w},\boldsymbol{R}) \geq \mathrm{VaR}_\beta(\boldsymbol{w})} g(\boldsymbol{w}, \boldsymbol{R}) p(\boldsymbol{R}) d\boldsymbol{R}}{\int_{g(\boldsymbol{w},\boldsymbol{R}) \geq \mathrm{VaR}_\beta(\boldsymbol{w})} p(\boldsymbol{R}) d\boldsymbol{R}}.$$

$\mathrm{CVaR}_\beta(\boldsymbol{w})$ is called CVaR in the confidence level β.

We now introduce some properties of the CVaR and explain how to estimate CVaR. Since CVaR is a expected value over VaR, we have

$$\mathrm{VaR}_\beta(\boldsymbol{w}) \leq \mathrm{CVaR}_\beta(\boldsymbol{w}).$$

Moreover, since CVaR is a coherent risk of measure, CVaR has the following properties. Let $\boldsymbol{R}$, $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be probability variables,

- monotonicity: If $\boldsymbol{R}_1 \leq \boldsymbol{R}_2$, then $\mathrm{CVaR}_\beta(\boldsymbol{R}_1) \leq \mathrm{CVaR}_\beta(\boldsymbol{R}_2)$.
- subadditivity: $\mathrm{CVaR}_\beta(\boldsymbol{R}_1 + \boldsymbol{R}_2) \leq \mathrm{CVaR}_\beta(\boldsymbol{R}_1) + \mathrm{CVaR}_\beta(\boldsymbol{R}_2)$.
- positive homogeneity: $\mathrm{CVaR}_\beta(\lambda \boldsymbol{R}) = \lambda \mathrm{CVaR}_\beta(\boldsymbol{R})$ for any $\lambda > 0$.
- translation invariance: $\mathrm{CVaR}_\beta(\boldsymbol{R} + c) = \mathrm{CVaR}_\beta(\boldsymbol{R}) + c$ for any c.

Next we explain how to evaluate CVaR. Let a function $F_\beta(\boldsymbol{w}, \alpha)$ be defined by

$$F_\beta(\boldsymbol{w}, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\boldsymbol{R} \in \mathbb{R}^N} [g(\boldsymbol{w}, \boldsymbol{R}) - \alpha]^+ p(\boldsymbol{R}) d\boldsymbol{R}, \tag{14}$$

where $[t]^+ = \max\{t, 0\}$. $\mathrm{CVaR}_\beta(\boldsymbol{w})$ can be calculated with F_β [6].

Theorem 3.1. *As a function of α, $F_\beta(\boldsymbol{w}, \alpha)$ is convex and continuously differentiable. The* CVaR_β *of the loss associated with any $\boldsymbol{w} \in K$ can be determined from the formula*

$$\mathrm{CVaR}_\beta(\boldsymbol{w}) = \min_{\alpha \in \mathbb{R}} F_\beta(\boldsymbol{w}, \alpha) \tag{15}$$

The right side of (15) is a convex programming problem when the loss function g is convex. When $\{\boldsymbol{R}^t\}$ is known, the BCRP is the optimal solution of the following problem.

$$\begin{array}{ll} \min & -\xi LS_T(\boldsymbol{w}, \{\boldsymbol{R}^t\}) + \mathrm{CVaR}_\beta(\boldsymbol{w}) \\ \text{s.t.} & \boldsymbol{w} \in K, \end{array}$$

where ξ is a parameter to balance $\mathrm{CVaR}_\beta(\boldsymbol{w})$ and the return $LS_T(\boldsymbol{w},\{\boldsymbol{R}^t\})$. From (15), we may write the above problem as

$$\begin{array}{ll} \min & \sum_{t=1}^{T}\{-\xi ls_t(\boldsymbol{w},\boldsymbol{R}^t)+\kappa^t(\boldsymbol{w},\alpha)\} \\ \text{s.t.} & \boldsymbol{w}\in K,\ \alpha\in\mathbb{R}, \end{array} \tag{16}$$

where

$$\kappa^t(\boldsymbol{w},\alpha) \overset{\text{def}}{=} \{\alpha+\frac{1}{(1-\beta)}[g(\boldsymbol{w},\boldsymbol{R}^t)-\alpha]^+\}. \tag{17}$$

Let $\hat{f}^t(\boldsymbol{w},\alpha)$ be defined by

$$\hat{f}^t(\boldsymbol{w},\alpha) = -\xi ls_t(\boldsymbol{w},\boldsymbol{R}^t)+\kappa^t(\boldsymbol{w},\alpha).$$

We cannot directly apply the EG-update method to (16) because decision variables in (16) are not only $\boldsymbol{w}\in K$ but also $\alpha\in\mathbb{R}$. Thus, we apply the general OMD method to (16). Let $\psi_2:\mathbb{R}^{N+1}\to\mathbb{R}$ be given by

$$\psi_2(\boldsymbol{R},\alpha)=\psi_1(\boldsymbol{R})+\frac{1}{2}\alpha^2.$$

Then, the Bregman divergence with ψ_2 can be written as follows.

$$B_{\psi_2}((\boldsymbol{w},\alpha),(\boldsymbol{v},\alpha')) = \sum_{i=1}^{N}(v_i-w_i)+\sum_{i=1}^{N}w_i\log(\frac{w_i}{v_i})+\frac{1}{2}(\alpha-\alpha')^2$$

Moreover, $\partial\hat{f}^t(\boldsymbol{w},\alpha)=(\partial_{\boldsymbol{w}}\hat{f}^t(\boldsymbol{w},\alpha),\partial_\alpha\hat{f}^t(\boldsymbol{w},\alpha))^{\mathrm{T}}$ can be written as follows.

$$\partial_{\boldsymbol{w}}\hat{f}^t(\boldsymbol{w},\alpha)=\begin{cases} \{-\frac{\xi\boldsymbol{R}^t}{\langle\boldsymbol{w}^t,\boldsymbol{R}^t\rangle}-\frac{\boldsymbol{R}^t-\boldsymbol{e}}{1-\beta}\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha>0, \\ \{-\frac{\xi\boldsymbol{R}^t}{\langle\boldsymbol{w}^t,\boldsymbol{R}^t\rangle}-\frac{\gamma(\boldsymbol{R}^t-\boldsymbol{e})}{1-\beta}|\gamma\in[0,1]\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha=0, \\ \{-\frac{\xi\boldsymbol{R}^t}{\langle\boldsymbol{w}^t,\boldsymbol{R}^t\rangle}\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha<0, \end{cases}$$

$$\partial_\alpha\hat{f}^t(\boldsymbol{w},\alpha)=\begin{cases} \{1-\frac{1}{1-\beta}\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha>0, \\ \{1-\frac{\gamma}{1-\beta}|\gamma\in[0,1]\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha=0, \\ \{1\} & \text{If } -\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R}^t-\boldsymbol{e})-\alpha<0, \end{cases}$$

where $\partial_{\boldsymbol{w}}$ and ∂_α denote subdifferentials with respect to $\boldsymbol{w}$ and α respectively. Note that

$$\partial\hat{f}(\boldsymbol{w},\alpha)=\mathrm{co}\{(\boldsymbol{b},c)^{\mathrm{T}}\in\mathbb{R}^{N+1}|\boldsymbol{b}\in\partial_{\boldsymbol{w}}\hat{f}(\boldsymbol{w},\alpha),\ c\in\partial_\alpha\hat{f}(\boldsymbol{w},\alpha)\}.$$

The OMD method for problem (16) can be written as follows.
The OMD method with downside risk

Step0. $\xi>0$, $\boldsymbol{w}^1=(1/N,\cdots,1/N)$, $\alpha^1=\langle\boldsymbol{w}^1,(\boldsymbol{R}^1-\boldsymbol{e})\rangle$,

Step1. $t=1$ to T
$\boldsymbol{b}^t\in\partial_{\boldsymbol{w}}\hat{f}^t(\boldsymbol{w}^t,\alpha^t),\ c^t\in\partial_\alpha\hat{f}^t(\boldsymbol{w}^t,\alpha^t)$
$\boldsymbol{w}^{t+1}=\arg\min_{\boldsymbol{w}\in K}[B_{\psi_2}((\boldsymbol{w},\alpha),(\boldsymbol{w}^t,\alpha^t))+\eta_t\langle\boldsymbol{b}^t,\boldsymbol{w}-\boldsymbol{w}^t\rangle]$
$\alpha^{t+1}=\arg\min_{\alpha\in\mathbb{R}}[\frac{1}{2}(\alpha-\alpha^t)^2+\eta_t\langle c^t,\alpha\rangle]$

Moreover, the regret of the OMD method is $O(\sqrt{T})$ when $\eta_t=\frac{1}{\sqrt{t}}$. Then, a portfolio vector sequence $\{\boldsymbol{w}^t\}$ constructed by this algorithm is universal for problem (16).

4 Numerical Experiments

In this section, we test universal portfolios taking account of trading cost and downside risk by using two data set. One data set is random data set according to normal distribution. The other data set is historical market data set.

4.1 Data

In this experiment, we test on two type data set.

4.1.1 Data 1

We use random data set according to normal distribution to confirm the effectiveness of the models. We consider a portfolio consisting of 4 assets. We assume $T = 10000$. Let a mean value vector μ and a variance-covariance matrix Σ be as follows.

$$\mu_1 = (1, 1, 1, 1),^{\mathrm{T}}$$

$$\Sigma_1 = \begin{pmatrix} 0.1 & 0.05 & 0 & 0 \\ 0.05 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}.$$

We use this data as a return sequence $\{\boldsymbol{R}^t\}$.

4.1.2 Data 2

We tested our models on historical market data from Tokyo Stock Exchange accumulated 13-year period. For each experiment, we restricted our attention to a subset of the 4 stocks and money for which we have data, and compared with the EG-update method as well as the portfolio which is invested equally for the subset.

4.2 The experiment with data 1

4.2.1 The model with trading cost

In this section, we confirm the effectiveness the model taking account of trading cost by changing a slope of σ to construct the market in which there are large and small cap stocks in data 1. Because a small cap stock is sharper than a large cap stock for the market impact, we assume that σ of a small cap stock is larger than σ of a large cap stock. Let σ of a small cap stock as 0.04 and σ of a large cap stock as 0.01. We test on three large cap stocks and one small cap stock. Let a portfolio of three large cap stocks be w_1, w_2 and w_3 respectively and a portfolio of one small stock be w_4. The fluctuation of a portfolio vector sequence of the EG-update model is drawn in the figure 2 and the fluctuation of a portfolio vector sequence of the model taking account of trading cost is drawn in the figure 3. After T days, a portfolio of a small stock w_4 in the figure 3

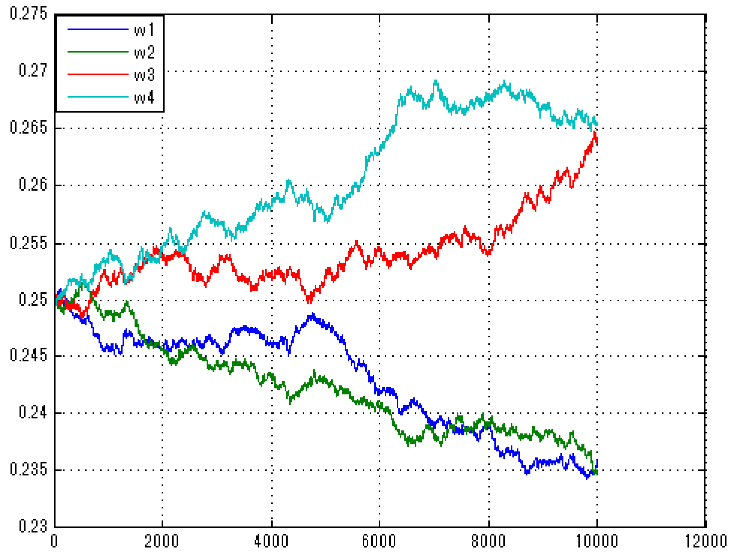

Figure 2: The fluctuation of $\boldsymbol{w}$ in the EG-update method

is smaller than w_4 in the figure 2. It may be because trading cost is taken account of. Compared to w_3, w_1 and w_2 make a similar fluctuation. This may be because there is a correlation between w_1 and w_2.

Moreover, we draw the amount of trading cost in the table 1. Compared to w_4 of the EG-update method, the amount of trading cost of w_4 in the model taking account of trading cost decreases a little. This may be because trading cost is taken account of.

Table 1: The amount of trading cost

	w_1	w_2	w_3	w_4
the EG-update method	23.8992	23.9159	30.0298	47.6858
the model with trading cost	23.9807	23.9963	30.0840	47.2135

4.2.2 the model with downside risk

In this section, we confirm the effectiveness of the model taking account of downside on data 1. We compare the VaR and the CVaR of three portfolios, the best constant rebalanced portfolio calculated by (15), a portfolio that we invest in assets equally and a portfolio of the problem of (16) when we use $\xi = 0$ and $\beta = 0.95$. Let the VaR and the CVaR of the BCRP be VaR and CVaR. Let the VaR and the CVaR of a portfolio that we invest in assets equally be $\mathrm{VaR_N}$ and $\mathrm{CVaR_N}$. Let the VaR and the CVaR of a portfolio of the problem of (16) with $\xi = 0$ be $\mathrm{VaR_{OMD}}$ and $\mathrm{CVaR_{OMD}}$. The result

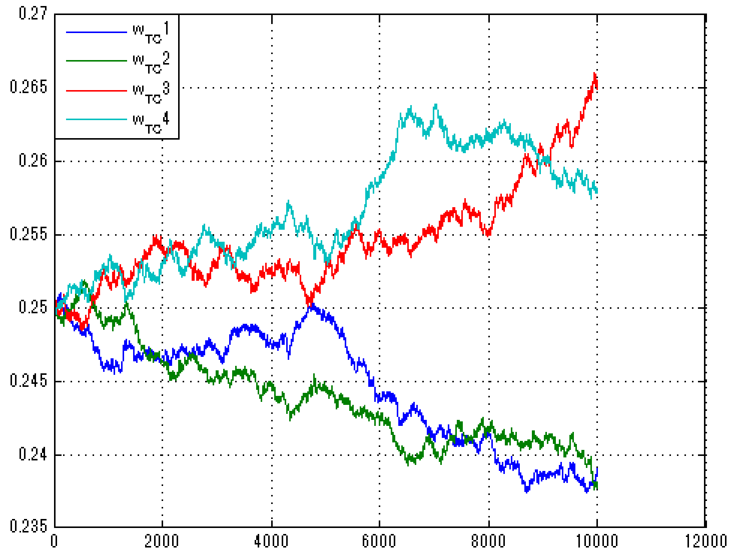

Figure 3: The fluctuation of $\boldsymbol{w}$ in the model with trading cost

is drawn in the table 2. Compared to $\mathrm{VaR_N}$, $\mathrm{CVaR_N}$, $\mathrm{VaR_{OMD}}$, $\mathrm{CVaR_{OMD}}$ is close to VaR, CVaR. This may be because downside risk is taken account of.

Table 2: Comparing VaR and CVaR

	VaR	CVaR
The best constant rebalance	0.3177	0.3990
The OMD method with downside risk	0.3218	0.4048
$\boldsymbol{w} = (1/N, \cdots, 1/N)^{\mathrm{T}}$	0.3425	0.4268

4.3 The experiment with data 2

4.3.1 The model with trading cost

In this experiment, we compare the model taking account of trading cost to the EG-update method to confirm the effectiveness of the model taking account of trading cost. We assume that both models need trading cost when we sell or buy assets and approximate the trading cost to a straight line with a slope of σ. We assume that σ depend on a turnover. We assume that trading cost is not necessary for money. Let a slope of the linear trading cost of four stocks be $\sigma_i (i = 1, 2, 3, 4)$. From the amount of a turn over, let $\sigma_i (i = 1, 2, 3, 4)$ be

$$(\sigma_1, \sigma_2, \sigma_3, \sigma_4)^{\mathrm{T}} = (0.0238, 0.0493, 0.0055, 0.0370)^{\mathrm{T}}.$$

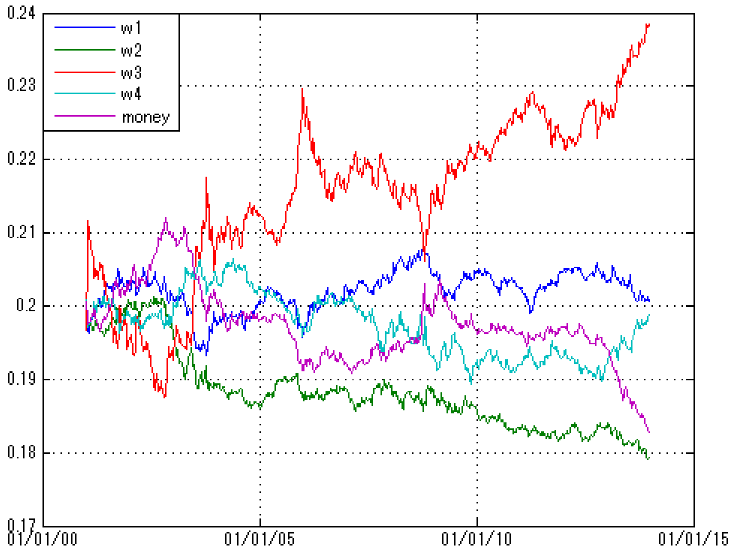

Figure 4: The fluctuation of $\boldsymbol{w}$ of the EG-update method

In the figure 4, the fluctuation of $\boldsymbol{w}$ of the EG-update method is drawn. In the figure 5, the fluctuation of $\boldsymbol{w}$ of the model taking account of trading cost is drawn. Compared to the fluctuation of $\boldsymbol{w}$ in the figure 4, the fluctuation of $\boldsymbol{w}$ in the figure 5 is a little small. This may be because trading cost is taken into account.

4.3.2 The model with downside risk

In this experiment, we compare the model taking account of downside risk when we use $\beta = 0.95$ and $\xi = 1, 5$ to the EG-update method and the model in which we invest in assets equally. The result is drawn in the figure 6 and 7.

In the figure 6, we may confirm that the model taking account of downside risk prevents the asset price from decreasing. In the figure 7, we may confirm that the rate of assets of the model taking account of downside risk is close to the rate of assets of EG-update method by being ξ larger.

5 Conclusion

We focused on the relation between the universal property for the universal portfolio sequence and the no internal regret for the algorithm of the online optimization. We could construct more flexible portfolio sequence by using general framework of the online optimization. As the results, we proposed universal portfolios taking account of trading

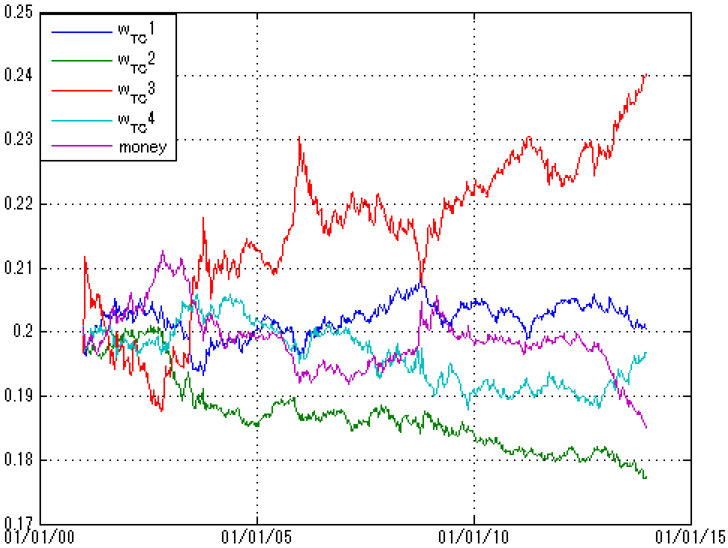

Figure 5: The fluctuation of $\boldsymbol{w}$ of the model with trading cost

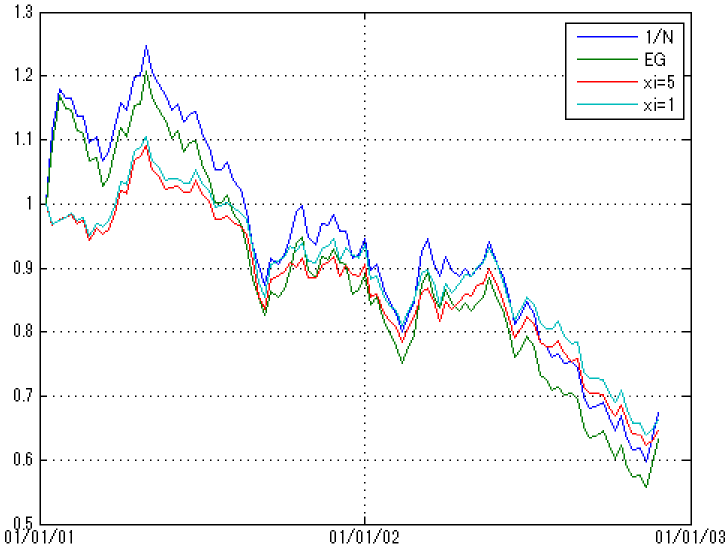

Figure 6: The rate of asset increase from 2001 to 2002

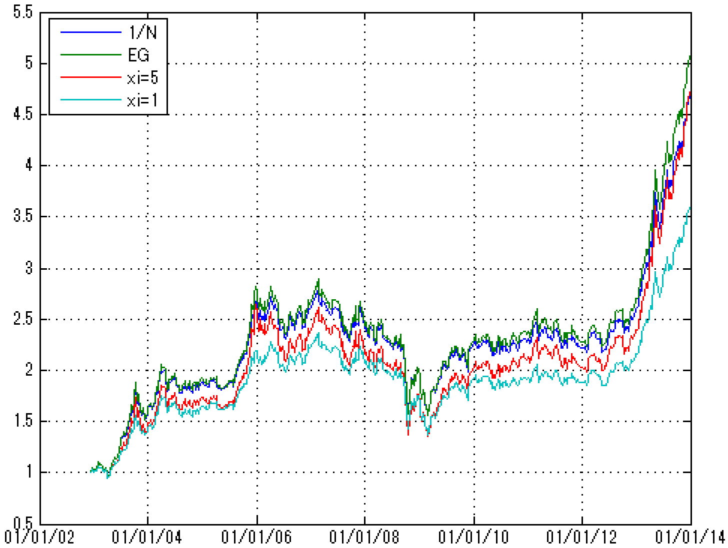

Figure 7: The rate of asset increase from 2003 to 2013

cost and downside risk. Moreover, we conducted some numerical experiments to show the effectiveness of the proposed models.

Acknowledgment

First of all, I would like to express sincere appreciation to Associate Professor Nobuo Yamashita. Although I sometimes troubled him due to my greenness, he always kindly looked after me and gave me plenty of precise advice. It is an honor to have studied under him. I am grateful to Assistant Professor Ellen Hidemi Fukuda. She was so kind that she gave me plenty of precise advice. Finally, I would like to thank all members of Yamashita Laboratory, my friends and my family for their encouraging words.

6 Appendix

In this section, we introduce theorems, definitions and lemmas about a convex function used in this paper. The proof is written in Shalev-Shwartz[7] and Rockafellar[8].

Definition 6.1. *Let f be a function from S to $(-\infty, +\infty]$, where S is a convex set (for example $S = \mathbb{R}^n$). Then f is convex on S if and only if*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y),\ 0 < \lambda < 1,$$

for every x and y in S.

Theorem 6.1. *Let S be a convex set. A function $f : S \to \mathbb{R}$ is convex set iff for all $\boldsymbol{w} \in S$ there exists $\boldsymbol{z}$ such that*

$$\forall \boldsymbol{z} \in \partial f(\boldsymbol{w}) \forall \boldsymbol{u} \in S,\ f(\boldsymbol{u}) \geq f(\boldsymbol{w}) + \langle \boldsymbol{z}, \boldsymbol{u} - \boldsymbol{w} \rangle. \tag{18}$$

Definition 6.2. *A function $f : S \to \mathbb{R}$ is σ-strongly-convex over S with respect to a norm $\|\cdot\|$ if for any $\boldsymbol{w} \in S$ we have*

$$\forall \boldsymbol{z} \in \partial f(\boldsymbol{w}), \forall \boldsymbol{u} \in S,\ f(\boldsymbol{u}) \geq f(\boldsymbol{w}) + \langle \boldsymbol{z}, \boldsymbol{u} - \boldsymbol{w} \rangle + \frac{\sigma}{2}\|\boldsymbol{u} - \boldsymbol{w}\|^2$$

Lemma 6.1. *Let S be a nonempty convex set. Let $f : S \to \mathbb{R}$ be a σ-strongly-convex function over S with respect to a norm $\|\cdot\|$. Let $\boldsymbol{w} = \arg\min_{\boldsymbol{v} \in S} f(\boldsymbol{v})$. Then, for all $\boldsymbol{u} \in S$*

$$f(\boldsymbol{u}) - f(\boldsymbol{w}) \geq \frac{\sigma}{2}\|\boldsymbol{u} - \boldsymbol{w}\|^2.$$

Lemma 6.2. *If f is twice differentiable, then it is easy to verify that a sufficient condition for strong convexity of f is that for all $\boldsymbol{w},\ \boldsymbol{v}, \langle \nabla^2 f(\boldsymbol{w})\boldsymbol{v}, \boldsymbol{v} \rangle \geq \sigma\|\boldsymbol{v}\|$, where $\nabla^2 f(\boldsymbol{w})$ is the Hessian matrix of f at $\boldsymbol{w}$, namely, the matrix of second-order partial derivatives of f at $\boldsymbol{w}$.*

Lemma 6.3. *The function $R(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}} \log(\boldsymbol{w})$ is 1-strongly-convex with respect to the l_1-norm over the simplex set $K = \{\boldsymbol{w} \in \mathbb{R}^N : 0 \leq w_i,\ \sum_{i=1}^N w_i = 1\}$.*

Theorem 6.2. *Let f be a convex function from $\mathbb{R}^n$ to $(-\infty, \infty]$, and let φ be a convex function from $\mathbb{R}$ to $(-\infty, \infty]$ which is nondecreasing. Then $h(x) = \varphi(f(x))$ is convex on $\mathbb{R}^n$ (where one sets $\varphi(+\infty) = +\infty$).*

References

[1] T. M. Cover, *Universal Portfolios*, Mathematical Finance, Vol. 1, No. 1, 1-29, 1991.

[2] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K.Warmuth, *On-line Portfolio Selection Using Multiplicative Updates*, Mathematical Finance, Vol. 8, No. 4, 325-347, 1998.

[3] J. Kivinen and M. K. Warmuth, *Exponentiated Gradient versus Gradient Descent for Linear Predictors*, Information and Computation, Vol. 132, No. 1, 1-63, 1997.

[4] D. P. Foster and R. Vohra, *Regret in the On-Line Decision Problem*, Games and Economic Behavior, Vol.29, No. 1, 7-35, 1999.

[5] J. C. Duchi, S. Shalev-Shwartz, Y. Singer and A. Tewari, *Composite Objective Mirror Decent*, 2010.

[6] R. T. Rockafellar and S. Uryasev, *Optimization of Conditional Value-at-Risk*, Journal of risk, Vol.2, 21-42, 2000.

[7] S. Shalev-Shwartz, *Online Learning and Online Convex Optimization*, Foundations and Trends in Machine Learning, Vol. 4, No. 2, 107-194, 2011.

[8] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Vol. 28, 1997.

Master's Thesis

Universal portfolios with trading cost and downside risk

Guidance

Associate Professor Nobuo Yamashita

Rei UMEDA

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2014

Universal portfolios
with trading cost and downside risk

Rei UMEDA

February 2014

Universal portfolios
with trading cost and downside risk

Rei UMEDA

Abstract

A universal portfolio is a portfolio sequence which theoretically achieves the almost same wealth by the best constant rebalanced portfolio in the long term. It is constructed from historical data of asset prices. However, since the existing models for the universal portfolio consider the wealth only, they are not necessarily practical models.

In this paper, in order to construct a more practical universal portfolio, we focus on the relationship between the universal property for the universal portfolio sequence and the no internal regret for the algorithm of the online optimization. We can construct more flexible portfolio sequence by using general framework of the online optimization. As the results, we propose universal portfolios taking account of trading cost and downside risk. Moreover, we conduct some numerical experiments to show the effectiveness of the proposed models.