UNIVERSAL PORTFOLIOS VIA CONTEXT TREES

Suleyman S. Kozat

Koc University
Istanbul, Turkey
Email: skozat@ku.edu.tr

Andrew C. Singer and Andrew J. Bean

University of Illinois at Urbana Champaign
Urbana, IL 61801, USA
Email: acsinger@uiuc.edu
Email: ajbean@uiuc.edu

ABSTRACT

In this paper, we consider the sequential portfolio investment problem considered by Cover [3] and extend the results of [3] to the class of piecewise constant rebalanced portfolios that are tuned to the underlying sequence of price relatives. Here, the piecewise constant models are used to partition the space of past price relative vectors where we assign a different constant rebalanced portfolio to each region independently. We then extend these results where we compete against a doubly exponential number of piecewise constant portfolios that are represented by a context tree. We use the context tree to achieve the wealth of a portfolio selection algorithm that can choose both its partitioning of the space of the past price relatives and its constant rebalanced portfolio within each region of the partition, based on observing the entire sequence of price relatives in advance, uniformly, for every bounded deterministic sequence of price relative vectors. This performance is achieved with a portfolio algorithm whose complexity is only linear in the depth of the context tree per investment period. We demonstrate that the resulting portfolio algorithm achieves significant gains on historical stock pairs over the algorithm of [3] and the best constant rebalanced portfolio.

Index Terms— universal, portfolio, investment, context tree, piecewise models.

I. INTRODUCTION

In sequential portfolio selection [1] [2] [3], the objective is to select sequential portfolios for a market with a finite number of stocks to maximize wealth with respect to a candidate class of investment strategies. The market is modeled by a sequence of price relative vectors $\boldsymbol{x}^n = \boldsymbol{x}[1], \ldots, \boldsymbol{x}[n]$, $\boldsymbol{x}[t] \in \mathbb{R}_+^m$. The jth entry $x_j[t]$ of a price relative vector $\boldsymbol{x}[t]$ represents the ratio of closing to opening price of the jth stock for the tth trading day. An investment at day t is represented by the portfolio vector $\boldsymbol{b}[t]$, $\boldsymbol{b}[t] \in \mathbb{R}_+^m$ and $\sum_{j=1}^m b_j[t] = 1$ for all t. Each entry $b_j[t]$ corresponds to the portion of the wealth invested in stock j at day t. The achieved wealth after n investment periods is given by $\prod_{t=1}^n \boldsymbol{b}^T[t]\boldsymbol{x}[t]$.

In [3], Cover presented an algorithm that asymptotically achieves the wealth of the best constant rebalanced portfolio from the class of all constant rebalanced portfolios for any sequence of price relative vectors, i.e., an algorithm that achieves $\max_{\boldsymbol{b}} \prod_{t=1}^n \boldsymbol{b}^T \boldsymbol{x}[t]$, where the maximizing $\boldsymbol{b}^*$ can only be chosen in hindsight. Here, the competition class is the class of all constant rebalanced portfolios. We first extend Cover's algorithm and construct sequential portfolios that compete against the best piecewise constant rebalanced portfolios. In our framework, the space of past price relatives is partitioned into a union of disjoint regions over each of which, a constant rebalanced portfolio is fitted independently. As an example, suppose at trading period t, we divide the space of past price relatives, $\boldsymbol{x}[t-1] \in \mathbb{R}_+^m$ as in Figure 1 into J disjoint regions R_j where $\bigcup_{j=1}^J R_j = \mathbb{R}_+^m$ (e.g., $J = 4$ for Figure 1). Here, if $\boldsymbol{x}[t-1] \in R_1$, then stock 1 performed better than stock 2 at trading day $t-1$ (however, both stocks lost money, i.e., $x_1[t-1] < 1, x_2[t-1] < 1$). If $\boldsymbol{x}[t-1] \in R_1 \bigcup R_2$ where the gain of stock 1 was greater than the gain of stock 2 at trading day $t-1$, then investing in stock 1 more than stock 2 in the next trading period t may be a good idea. This strategy may work if there are no drastic shifts in stock trends. Hence, we assign each region a different portfolio and invest at each trading day depending on the relative performance of each stock on the previous day. For piecewise constant rebalanced portfolios, the portfolio used is $\boldsymbol{b}_j$ such that if $\boldsymbol{x}[t-1] \in R_j$ then we invest with $\boldsymbol{b}_j$ at trading period t.

We first present results for the piecewise constant rebalanced portfolios when the regions R_j are fixed and known. Given such a partition $\bigcup_{j=1}^J R_j = \mathbb{R}_+^m$ and the past values of price relative vectors $\boldsymbol{x}[t]$, $t = 1, \ldots, n-1$, we define a competing algorithm from the class of all piecewise constant rebalanced portfolios as $\hat{\boldsymbol{b}}[n] = \boldsymbol{b}_{s[n-1]}$, where $s[n-1] = j$ when $\boldsymbol{x}[n-1] \in R_j$ is an indicator variable and $j \in \{1, \ldots, J\}$. For each region, the constant rebalanced portfolio vector $\boldsymbol{b}_j$ can be selected independently. Here we try to achieve $\sup_{\boldsymbol{b}_j \in \mathcal{B}} \prod_{t=1}^n \boldsymbol{b}_{s[t-1]}^T \boldsymbol{x}[t]$, $j \in \{1, \ldots, J\}$, where $\mathcal{B} = \{\boldsymbol{b} : \sum_{i=1}^m b_i = 1, b_i \in [0, 1]\}$ is the simplex, i.e., we try to achieve the performance of the best piecewise constant rebalanced portfolios tuned to the underlying sequence of price relatives $\boldsymbol{x}^n$. We first demonstrate an algorithm $\hat{\boldsymbol{b}}[t]$ whose achieved wealth, over that of the best piecewise constant rebalanced portfolios is upper bounded by $O(\frac{J(m-1)}{2}\ln(n/J))$. Our algorithm pays a "parameter regret" of $O(\frac{m-1}{2}\ln(n/J))$ per region to effectively learn (or compete against) the best parameters for that region.

However, in this basic form, this problem was solved in [3], [4]. Since the partitions are fixed, we already know the side-information generating mechanism, i.e., $s[t-1]$. Independently applying Cover's algorithm for each region will yield the required sequential portfolio with the corresponding result. However, we extend these results to the case when the boundaries of each region are also selected by the class, i.e., the side-information generating mechanism is also a design parameter. Hence, we compete against the best side information generating mechanism from the class of all such sequences represented by a context tree. Here, we try to achieve the performance of the best sequential piecewise constant rebalanced portfolio when the partitioning of the past price relatives is taken from a doubly exponentially large class of possible partitions. These partitions will be compactly represented using a "context tree" [5]. Here, we have neither a priori knowledge of the selected partition nor the best model parameters, i.e., constant rebalanced portfolios given that partition.

We demonstrate an algorithm that asymptotically achieves the performance of the best sequential portfolio (corresponding to a particular partition) from the doubly exponentially large class of such partitioned portfolios. To this end, we define a depth-K context tree for a partition with up to 2^K regions, as shown in Figure 2, where, for this tree, $K = 2$. For a depth-K context tree, the 2^K finest partition bins correspond to leaves of the tree. On this tree, each of the bins are assigned to regions: $\{1 \geq x_2 > x_1 \geq 0\}$, $\{x_2 > x_1 \geq 1\}$, $\{1 \geq x_1 > x_2 \geq 0\}$ and $\{x_1 > x_2 \geq 1\}$. Of course, more general partitioning schemes could be represented by such a context tree.

For a tree of depth-K, there exist $2^{K+1} - 1$ nodes, including leaf nodes and internal nodes. Each node η on this tree represents a portion of positive orthant $\mathbb{R}_+^m$, $R\eta$. The region corresponding to each node η, $R\eta$, (if it is not a leaf) is constructed by the union of regions represented by the nodes of its children; the upper node $R\eta_u$ and the lower node $R\eta_l$, $R\eta = R\eta_u \cup R\eta_l$. By this definition, any inner node is the root of a subtree and represents the union of its corresponding leaves (or bins).

We define a "partition" of $\mathbb{R}_+^m$ as a specific partitioning $\mathcal{P}_i = \{R_{i,1}, \ldots, R_{i,J_i}\}$ with $\bigcup_{j=1}^{J_i} R_{i,j} = \mathbb{R}_+^m$, where each $R_{i,j}$ is represented by a node on the tree in Figure 1 and $R_{i,j}$ are disjoint. There exist a doubly-exponential number, $N_K \approx (1.5)^{2^K}$ such partitions, $\mathcal{P}_i$, $i = 1, \ldots, N_K$, embedded within a depth-K full tree. This is equivalent to the number of "proper binary trees" of depth at most K, and is given by Sloane's sequence $A003095$ [6], [7]. For each such partition, there exists a corresponding sequential algorithm that achieves the performance of the best piecewise constant rebalanced portfolio for that partition. We can then construct an algorithm that will achieve the performance of the best sequential algorithm from this doubly exponential class.

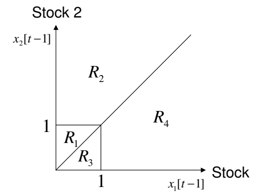

Fig. 1. A partition of the $\mathbb{R}_+^m$.

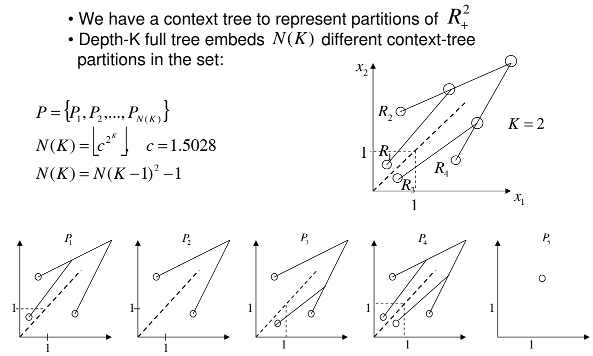

Fig. 2. A full tree of depth 2 that represents all context-tree partitions of $\mathbb{R}_+^m$ into at most four possible regions.

To achieve the performance of the best sequential algorithm (i.e., the best partition), we try to achieve $\sup_{\mathcal{P}_i} \prod_{t=1}^{n} \hat{\boldsymbol{b}}_{\mathcal{P}_i}^T[t]\boldsymbol{x}[t]$ where $\hat{\boldsymbol{b}}_{\mathcal{P}_i}[t]$ is the corresponding sequential piecewise constant rebalanced portfolio for partition $\mathcal{P}_i$. We will then demonstrate a sequential portfolio, $\hat{\boldsymbol{b}}[t]$, such that the "structural regret" is at most $O(2C(\mathcal{P}_i))$ with respect to the best $\hat{\boldsymbol{b}}_{\mathcal{P}_i}[t]$, where $C(\mathcal{P}_i)$ is a constant which depends only on the partition $\mathcal{P}_i$. This yields, upon combining the parameter and structural regret, an algorithm achieving $O(\frac{J_i(m-1)}{2}\ln(n/J_i)) + O(2C(\mathcal{P}_i))$ uniformly for any $\boldsymbol{x}^n$.

Hence, the algorithms introduced here are "twice-universal" in that they asymptotically achieve wealth of the best portfolio in which the parameters of the piecewise constant rebalanced portfolio and also the partitioning structure of the model itself can be selected, based on observing the whole sequence in advance.

We begin our discussion of piecewise constant rebalanced portfolios with the case when the partition is fixed and known in the initial part of Section II. We then extend these results using context trees. We present the corresponding theorems and outline their proofs due to space limitations.

Full proofs and complete implementations (with corresponding MATLAB code) for the universal portfolio selection algorithms with complexity linear in $O(K)$ per combination are given in [8]. We conclude the paper by demonstrating the substantial gain achieved by our algorithms on benchmark historical stock pairs.

II. PIECEWISE CONSTANT REBALANCED PORTFOLIOS

In this section, we first construct a sequential portfolio that performs over every sequence of price relatives $\boldsymbol{x}^n$ as well as the best fixed piecewise constant rebalanced portfolio for that sequence with a partition of the price relative space given by $\bigcup_{j=1}^{J} R_j = \mathbb{R}_+^m$. One such portfolio from the class against which this algorithm will compete can be represented by $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_J]$ and would achieve a wealth $\prod_{t=1}^{n} \boldsymbol{b}_{s[t-1]}^T \boldsymbol{x}[t]$. Since the number and boundaries of the regions are known, we have J independent constant rebalanced portfolio selection problems. Applying Cover's algorithm in each region yields the following theorem.

Theorem 1: *Let $\boldsymbol{x}^n = \boldsymbol{x}[1], \ldots, \boldsymbol{x}[n]$ be an arbitrary sequence of price relative vectors such that $\boldsymbol{x}[t] \in \mathbb{R}_+^m$ for all t and some components of $\boldsymbol{x}[t]$ can be zero. Then we can construct sequential portfolios $\tilde{\boldsymbol{b}}[n]$ with complexity linear in n^m such that*

$$\ln \prod_{t=1}^{n} \tilde{\boldsymbol{b}}^T[t]\boldsymbol{x}[t] \geq$$
$$\ln \prod_{t=1}^{n} \boldsymbol{b}_{s[t-1]}^T \boldsymbol{x}[t] - \sum_{j=1}^{J} \frac{(m-1)}{2} \ln(n_j + 1) - O(1)$$

where n_j is the number of elements of region j and $s[t-1]$ is the state indicator variable, i.e., $s[t-1] = j$ when $\boldsymbol{x}[t-1] \in R_j$.

The proof of Theorem 1 directly follows application of [3] into J separate regions. We next consider the portfolio selection problem where the class against which the algorithm must compete includes not only the best constant rebalanced portfolios for a given partition, but also the best partition of the space of past price relatives as well. As such, we are interested in the following wealth $\prod_{t=1}^{n} \hat{\boldsymbol{b}}_{\mathcal{P}_i}^T[t]\boldsymbol{x}[t]$, where $\mathcal{P}_i$ is a partition of the $\mathbb{R}_+^m$ with the state indicator variable $s_i[t-1] = j$ if $\boldsymbol{x}[t-1] \in R_{i,j}$, and $\mathcal{P}_i = \{R_{i,1}, \ldots, R_{i,J_i}\}$ with $\bigcup_{j=1}^{J_i} R_{i,j} = \mathbb{R}_+^m$ for a J_i, and $\hat{\boldsymbol{b}}_{\mathcal{P}_i}[t]$ is the corresponding sequential portfolio for the partition $\mathcal{P}_i$. The partition $\mathcal{P}_i$ can be viewed as in Figure 2 as a subtree or "context tree" of a depth K full tree with the $R_{i,j}$ corresponding to the nodes of the tree. Each $R_{i,j}$ is represented by a node on the full tree and $R_{i,j}$ are disjoint. Given the full tree, there exist N_K such partitions, i.e., $\mathcal{P}_i$, $i = 1, \ldots, N_K$, where $N_K = N_{K-1}^2 + 1$.

Similar to [5], we define $C(\mathcal{P}_i)$ as the number of bits that would have been required to represent each partition $\mathcal{P}_i$ on the tree using a universal code: $C(\mathcal{P}_i) = J_i + n_{\mathcal{P}_i} - 1$, where $n_{\mathcal{P}_i}$ is the total number of leaves in $\mathcal{P}_i$ that have depth less than K, i.e., leaves of $\mathcal{P}_i$ that are inner nodes of the tree. Since $n_{\mathcal{P}_i} \leq J_i$, $C(\mathcal{P}_i) \leq 2J_i - 1$. Given the tree, we can construct a sequential algorithm with linear complexity (to combine the exponential number of partions) in the depth of the context tree per price relative vector that asymptotically achieves both the performance of the best sequential portfolio and also the performance of the best constant rebalanced portfolio for any partition as follows.

Theorem 2: *Let $\boldsymbol{x}^n = \boldsymbol{x}[1], \ldots, \boldsymbol{x}[n]$ be an arbitrary sequence of price relative vectors such that $\boldsymbol{x}[t] \in R_+^m$ for all t and some components of $\boldsymbol{x}[t]$ can be zero. Then we can construct sequential portfolios $\tilde{b}_u[t]$ with complexity linear in n^m such that*

$$\ln \prod_{t=1}^{n} \tilde{\boldsymbol{b}}_u^T[t]\boldsymbol{x}[t] \geq \sup_{\mathcal{P}_i} \Bigg(\sup_{\boldsymbol{b}_{i,j} \in \mathcal{B}} \ln \prod_{t=1}^{n} \boldsymbol{b}_{i,s_i[t-1]}^T \boldsymbol{x}[t]$$
$$-2C(\mathcal{P}_i)\ln(2) - \frac{J_i(m-1)}{2} \ln(\frac{n}{J_i} + 1) \Bigg) + O(1),$$

$\mathcal{P}_i$ is any partition on the context tree, $C(\mathcal{P}_i)$ is a constant that is less than or equal to $2J_i - 1$, $s_i[t-1]$ is the state indicator variable for partition $\mathcal{P}_i$, i.e., $s_i[t-1] = j$ if $\boldsymbol{x}[t-1] \in R_{i,j}^m$.

The complexity $O(n^m)$ is due to calculation of Cover's portfolios. The actual complexity of the combination algorithm is $O(K)$, i.e., linear in the depth of the context-tree. As an example, if one replaces Cover's algorithm with [2], the complexity of the overall algorithm would be $O(K+m)$, albeit with different performance bounds. The construction of the universal portfolio $\tilde{\boldsymbol{b}}_u[t]$ is given [8]. Note that the inequality in Theorem 2 holds for *any* partition of the data, including that achieving $\sup_{\mathcal{P}_i}$ over the right hand side. This implies that, without prior knowledge of any complexity constraint on the algorithm, such as prior knowledge of the depth of the context tree against which it is competing, the universal prediction algorithm can compete well with each and every subpartition (context-tree) within the depth-K full tree used in its construction.

II-A. Outline of Proof of Theorem 2

Given a partition $\mathcal{P}_i = \bigcup_{j=1}^{J_i} R_{i,j}$ of the $\mathbb{R}_+^m$, $\mathcal{P}_i \in \mathcal{P}$ (the competing class), we consider a family of portfolios, each with its own set of portfolio vectors $\boldsymbol{B}_i = [\boldsymbol{b}_{i,1}, \ldots, \boldsymbol{b}_{i,J_i}]$. Here, each $\boldsymbol{b}_{i,j}$ represents a constant portfolio vector for the jth region of partition $\mathcal{P}_i$, i.e., when $\boldsymbol{x}[t-1] \in R_{i,j}$, we use $\boldsymbol{b}_{i,j}$. For each pairing of $\mathcal{P}_i$ and $\boldsymbol{B}_i$, we consider the sequential wealth achieved by the corresponding algorithm $W(\boldsymbol{x}^n \mid \boldsymbol{B}_i, \mathcal{P}_i) = \prod_{t=1}^{n} \boldsymbol{b}_{i,s_i[t-1]}^T \boldsymbol{x}[t]$, where $s_i[t-1]$ is the state indicator variable for partition $\mathcal{P}_i$, i.e., $s_i[t-1] = j$ if $\boldsymbol{x}[t-1] \in R_{i,j}$. Applying Cover's algorithm for each segment will yield a sequential algorithm $\hat{\boldsymbol{b}}_{\mathcal{P}_i}[t]$ which will asymptotically achieve $\sup_{\boldsymbol{B}_i} W(\boldsymbol{x}^n \mid \boldsymbol{B}_i, \mathcal{P}_i)$. For each partition $\mathcal{P}_i$, we define a similar sequential predictor, each achieving the wealth of the best constant rebalanced portfolio

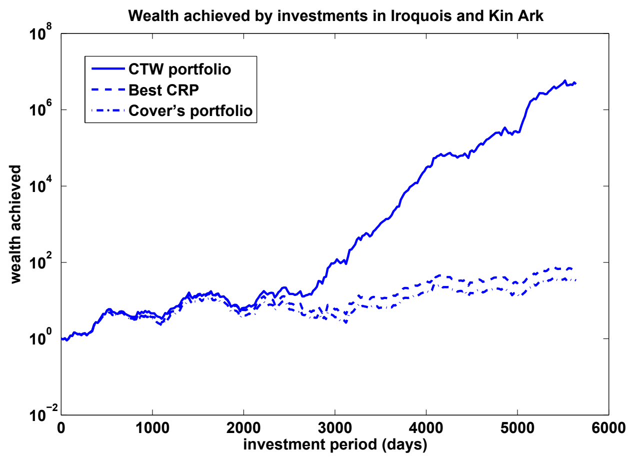

Fig. 3. Wealth achieved on stock pairs Iroquois-KinArk.

for that partition. We next show that a weighted combination of all these portfolios achieves a wealth asymptotically as large as the best sequential algorithm, i.e., $\hat{\boldsymbol{b}}_{\mathcal{P}_i}[t]$ with the largest wealth. This weighted combination can also be used to derive the corresponding portfolio that achieves this weighted wealth. However, this weighted combination can only be calculated by combining all the sequential algorithms corresponding to all N_K possible partitions which is naturally infeasible for large K. Nevertheless, we can demonstrate that, if we assign each node on the context tree a certain sequential algorithm (which is derived from Cover's algorithm that only runs over the sub-sequence of price relatives belonging to that node), we can construct the combination with complexity just linear in the depth of the context-tree. Hence, the final portfolio can be constructed on the context-tree with complexity $O(K)$ by combining N_K possible partitions.

III. SIMULATIONS

In this section, we apply our algorithm to historical stock prices (that are now classical) collected from the New York Stock Exchange over a 22-year period until 1985. These stock pairs include Iroquois-Kinark, IBM-Coca Cola and Meico-Fisch. We use a depth-4 context tree algorithm, with the partion given in Figure 1. We first try our algorithms on the Kinark-Iroquois pair as shown in Figure 3 which are chosen because of their volatility. In Figure 3, we plot the wealth achieved by our algorithm (CTW), Cover's algorithm (CRP) and the best constant portfolio tuned to the underlying stock prices (BCRP). We also repeat this experiment for other stock pairs and present final wealths in Figure 4. In each pair, our algorithm achieves significant gains over both algorithms.

Stocks	CRP	BCRP	CTW
Ir&Kn	37.14	72.59	5,503,998
Coke&Ibm	14.10	15.02	19.44
Fi&Me	27.09	35.28	50.66

Fig. 4. Performance on historical stock pairs.

IV. CONCLUSION

In this paper, we consider the problem of investing using piecewise constant rebalanced portfolios from a competitive algorithm perspective. Using context trees and methods based on sequential probability assignment, we have shown a portfolio selection algorithm whose achieved wealth is within $O(\ln(n))$ in the exponent of that of the best piecewise constant rebalanced portfolio that could only have been selected using all of the data in hindsight. We use a method similar to context tree weighting to compete well against a doubly exponential class of possible partitionings of the space of price relatives. We also demonstrated the significant gains achieved by this algorithm on historical data.

V. REFERENCES

[1] S. S. Kozat and A. C. Singer, "Universal constant rebalanced portfolios with switching," in *ICASSP*, 2007, pp. 1129–1132.

[2] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth, "Online portfolio selection using multiplicative updates," *Mathematical Finance*, vol. 8, no. 4, pp. 325–347, 1998.

[3] T. Cover and E. Ordentlich, "Universal portfolios with side-information," *IEEE Trans. on Info. Theory*, vol. 42, no. 2, pp. 348–363, 1996.

[4] J. E. Cross and A. R. Barron, "Efficient universal portfolios for past dependent target classes," *Mathematical Finance*, vol. 13, no. 2, pp. 245–276, 2003.

[5] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The context-tree weighting method: basic properties," *IEEE Trans. on Info. Theory*, vol. 41, pp. 653–664, 1995.

[6] N. J. Sloane, *A. Sequence A003095 (formerly M1544) in the On-Line Encyclopedia of Integer Sequences*.

[7] A. V. Aho and N. J. A. Sloane, "Some doubly exponential sequences," *Fibonacci Quarterly*, vol. 11, pp. 429–437, 1970.

[8] S. S. Kozat, A. C. Singer, and A. J. Bean, "Universal portfolios with context-trees," *to be submitted to IEEE Trans. on Sig. Proc.*, 2007.